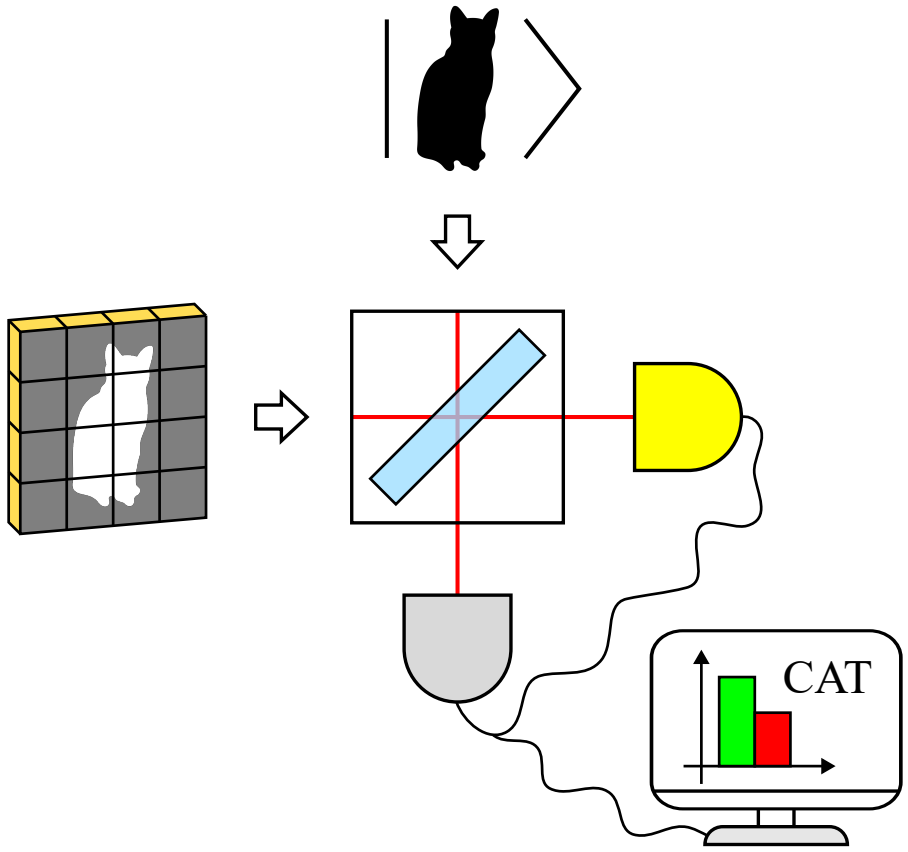
UNIVERSITÀ DEGLI STUDI DI PAVIA
DOTTORATO DI RICERCA IN FISICA – XXXVII CICLO

From imaging to timekeeping.
Position, phase and time in quantum technologies

Simone Roncallo



Università degli Studi di Pavia
Dipartimento di Fisica

DOTTORATO DI RICERCA IN FISICA – XXXVII CICLO

From imaging to timekeeping.
Position, phase and time in quantum technologies

Simone Roncallo

Submitted to the Graduate School of Physics of the University of Pavia in partial fulfilment of the requirements for the degree of

DOTTORE DI RICERCA IN FISICA

DOCTOR OF PHILOSOPHY IN PHYSICS

Supervisors
Prof. Lorenzo Maccone
Prof. Chiara Macchiavello

Cover: Image classification through the Hong-Ou-Mandel effect.

Title: From imaging to timekeeping. Position, phase and time in quantum technologies

Author: Simone Roncallo
PhD Thesis, Graduate School of the University of Pavia
Pavia, Italy, December 2024

To my family and cherished ones, for the unconditional love and care.

SUMMARY

In this thesis, we discuss applications to imaging, positioning, phase and time estimations in quantum mechanics, which encompass technological implications, practical scenarios, and foundational problems.

In the first part, we focus on the manipulation of information in terms of images. We discuss imaging in quantum-mechanical terms, by designing a quantum encoder that consists of a multiatom lattice sensor that electromagnetically interacts with the photons diffracted by the object, encoding its two-dimensional shape in the probabilities of a multiqubit quantum state. We introduce a quantum algorithm that can compress such a state, by reducing the number of encoding qubits without altering the overall visual pattern, i.e. a lossy compression. This allows to capture, compress, and send images even with limited quantum resources for storage and communication, providing an exponential speedup, in terms of gates only, over classical downsampling algorithms. Finally, we discuss image classification from a completely different perspective: by processing the information encoded in the spectrum of a single-photon state, without direct image reconstruction. Using the Hong-Ou-Mandel effect, we introduce an interferometric pattern recognition method for binary classification tasks, which shows a constant complexity in the number of computational operations and photons required by each classification instance. Mathematically, the response of the apparatus resembles that of a single artificial neuron, which, however, can classify images bypassing their reconstruction.

In the second part, we take a theoretical perspective on the challenges provided by characterizing real multiqubit devices. We summarize the notion of quantum channels as a quantum-mechanical framework for describing noisy processes. In this context, we introduce a tomographic protocol for the reconstruction of multiqubit noise models, which, using a limited number of experimental resources, can partially characterize channels with an exponential speedup over quantum process tomography. We discuss a deconvolution method for mitigating the expectation values of noisy measurements at the output of such channels.

In the last part, we consider the estimation of phase and the measurement of time in quantum mechanics. First, we present a novel interferometric protocol, which translates the estimation of an unknown phase into a binary search task. Such a method can work at arbitrary degrees of accuracy, while undergoing the resource scaling set by the Heisenberg limit. Second, we analyse the task of predicting the time at which a particle arrives at a screen, commonly known as the arrival time problem. We compare the theoretical proposals that attempt to solve this issue, also beyond the standard quantum theory. Exploring multiple regimes, we characterize under which conditions such proposals provide different, and experimentally discriminable, predictions, paving the way for future arrival time experiments.

ACKNOWLEDGEMENTS

When I first arrived in Pavia, I was both excited and scared about the path ahead. I am deeply grateful to my supervisors, Lorenzo Maccone and Chiara Macchiavello, for their invaluable guidance and mentorship throughout my PhD journey. Their support and encouragement to explore multiple topics have made my PhD the enriching experience I looked forward to.

I would like to extend my thanks to all the members of the QuIT group of the University of Pavia, as well as to everyone I have collaborated with. In particular, Marco Genovese, Fabrizio Piacentini, Krzysztof Sacha, Fabio Sciarrino, and Nicolò Spagnolo. I am grateful to Pérola Milman and Stefano Olivares for their suggestions on improving this manuscript. A special thanks goes to Stefano Mangini for his insightful advice and support, which guided me as I entered this fascinating research field. I am very grateful to my fellow PhD colleagues, Angela Rosy Morgillo and Emanuele Tumbiolo, with whom I shared countless challenging and enjoyable research efforts.

My deepest gratitude goes to my family: my father, Andrea, my mother, Rosella, and my partner, Martina. Fostered by their love and support, no obstacle seemed overwhelming and no challenge unattainable. An honourable mention to my cat, Rollo, who graciously and patiently posed as a model for several figures used in this thesis.

LIST OF PUBLICATIONS

Publications included in this manuscript:

[P1] **Simone Roncallo**, Lorenzo Maccone and Chiara Macchiavello, Quantum JPEG, AVS Quantum Sci. 5, 043803 (2023) [1]

[P2] **Simone Roncallo**, Angela Rosy Morgillo, Chiara Macchiavello, Lorenzo Maccone and Seth Lloyd, Quantum optical classifier with superexponential speedup (Preprint), arXiv:2404.15266 (2024) [2]

[P3] **Simone Roncallo**, Lorenzo Maccone and Chiara Macchiavello, Pauli transfer matrix direct reconstruction: channel characterization without process tomography, Quantum Sci. Technol. 9 015010 (2024) [3]

[P4] **Simone Roncallo**, Lorenzo Maccone and Chiara Macchiavello, Multiqubit noise deconvolution and characterization, Phys. Rev. A 107, 022419 (2023) [4]

[P5] **Simone Roncallo**, Xi Lu and Lorenzo Maccone, Interferometric binary phase estimations (Preprint), arXiv:2407.10966 (2024) [5]

[P6] **Simone Roncallo**, Krzysztof Sacha and Lorenzo Maccone, When does a particle arrive?, Quantum 7, 968 (2023) [6]

Publications not included in this manuscript:

[P7] Emanuele Tumbiolo, **Simone Roncallo**, Chiara Macchiavello, Lorenzo Maccone, Quantum frequency resampling, arXiv.2501.16283 (2025) [7]

[P8] Lorenzo Maccone, **Simone Roncallo** and Krzysztof Sacha, Quantum Measurements of Time: A reply to criticisms, arXiv.2501.10416 (2025) [8]

[P9] Paolo Meda, Nicola Pinamonti, **Simone Roncallo** and Nino Zanghì, Evaporation of four-dimensional dynamical black holes sourced by the quantum trace anomaly, Class. Quantum Grav. 38 195022 (2021) [Corrigendum: Class. Quantum Grav. 39 059501 (2022)] [9]

CONTENTS

LIST OF FIGURES

INTRODUCTION

We can only see a short distance ahead, but we can see plenty there that needs to be done.

Alan M. Turing [10]

HISTORICALLY, understanding the quantum mechanical behaviour of matter and light has contributed to several technological developments, including semiconductor devices and transistors (without which this draft would have required a typewriter).

Quantum mechanics continues fostering innovative ideas [11], giving rise to multiple research fields, which leverage quantum mechanical (or quantum-inspired) effects to improve or overcome the limitations of our current technological paradigm. The development of quantum optics has introduced innovative methodologies and techniques for imaging, e.g. by exploiting classical and quantum correlations to reconstruct the spatial shape of an object without directly looking at it [12, 13], to enable refocusing capabilities after detection occurred [14–16], or by leveraging interferometric [17–19] and metrological [20–23] strategies to enhance the resolution capabilities and low-light performances of imaging apparata. Whether they are captured in a classical or quantum way, images are an example of high-dimensional data, which may require a great computational effort to be further processed and manipulated. Quantum computing offers a promising perspective on this matter, allowing to encode and manipulate information with quantum mechanical systems, such as superconducting devices [24, 25], trapped-ions [26], neutral atoms [27] and photonic circuits [28]. Quantum algorithms can perform some computational tasks using fewer resources, outperforming their classical counterparts. From phase estimation [29], database search [30] and integer factorization [31], to linear systems solution [32] and machine learning [33]: the regimes of applicability of quantum computers span multiple fields, including image processing [34]. However, noise and error sensitivity remain one of the main bottlenecks that prevent quantum devices to achieve such desiderata. In recent years, increasing research efforts are being spent in developing error mitigation, protection, and correction techniques, allow-

ing more robust and faithful computations with the available devices [35]. Although fault-tolerant computing still represents the ultimate challenge, the most recent hardware and error-correcting developments suggest that we are gradually departing the noisy intermediate-scale quantum era (NISQ) [36], in favour of a logical intermediate scale (LISQ) [37, 38], in which noise still affects computation outputs but in a less prominent way, due to the combination of mitigation and correction strategies [39–43]. Despite the countless number of successes of quantum mechanics, there are several open problems, which remain unsolved to date. There exists no self-adjoint operator associated to time, due to the Pauli's objection [44], i.e. there are no standard prescriptions for time measurements. Straightforward questions, like asking when a particle hits a waiting screen [45], cannot just be answered without semi-classical approximations. This is called the arrival time problem. In this thesis, we explore the multiple facets of the quantum theory: from technological-oriented and informational approaches to theoretical and foundational aspects.

Contents This manuscript is organized in three parts, each tackling a distinct aspect of quantum theory. We start with the applications of quantum mechanics to imaging, including image capturing, processing and classification, subsequently. Adopting a closer look at practical issues, we consider the mathematical description and characterization of noisy processes, as well as their mitigation at the experimental output. Finally, we discuss phase measurements as well as one of the open problems that lies at the foundations of the theory: time and the description of arrival time measurements.

In Part I, we discuss the problem of capturing, processing, and classifying images by quantum optical means. In Chapter 1, we briefly review the electromagnetic field quantization in the Coulomb gauge, introducing the framework of quantum states and optical operations. We specialize to a conventional imaging setup, consisting of four parts: a photon source, an object, a linear optical system, and a spatial-resolving detector. We explicitly discuss such apparatus in quantum optical terms, for both single-photon and multimode coherent sources. Then, we fully quantize the above setup, by substituting the conventional photodetector with a multiatom lattice sensor, modelled as a multiqubit quantum register [1]. At each acquisition, the register electromagnetically interacts with the photons reflected by the object, by means of the multimode and multiatom Jaynes-Cummings Hamiltonian. This encodes the visual information contained in the spatial modes of light, i.e. the intensity of light reflected by the object, into the probabilities of a multiqubit quantum state, to be retrieved or stored for further quantum computational processing. In Chapter 2, we analyse the problem of processing an image encoded in the probabilities of such a state, either prepared on the quantum devices or encoded according to the above procedure. We specialize to image downsampling, namely the task of reducing the number of encoding pixels by quantum computational means, while preserving the original visual pattern. We present an algorithm that uses the quantum Fourier transform to discard the most significant qubits of the register, thus filtering the high spatial-frequency components of the image [1]. Inverting the Fourier transform compresses the original pattern in a register of fewer qubits. In terms of operations, this

algorithm achieves an exponential speedup over classical interpolation-based counterparts. In practice, measurement and image retrieval can reduce this speedup, due to the amount of statistics required to reconstruct the pattern. After analysing the trade-off between the number of encoding and discarded resources, we show that a residual advantage is still present, which exponentially increases with the size of the input register, i.e. when downsampling high-resolution data. In Chapter 3, we analyse imaging from an unfamiliar perspective: by removing images from the overall paradigm. We discuss a two-mode interferometric setup to classify objects without reconstructing their patterns [2], using, instead, only two photons to assess the spectral distinguishability between the input and a reference state, which contains the necessary features for discriminability. Classification is achieved through the Hong-Ou-Mandel effect, by looking at the probability that the photons exit the interferometer in different modes, i.e. that they produce a coincidence event. Mathematically, this scheme corresponds to (the analog version of) a classical artificial neuron, i.e. it is a quantum optical neuron. By analytically comparing the resources needed, we show that our method requires constant computational operations and injected photons, whereas the classical methods depend linearly on the image resolution, which is a superexponential speedup.

In Part II, we discuss multiqubit operations and measurements taken on systems that interact with their surrounding environment. In Chapter 4, we briefly summarize the framework of quantum channels, which generalizes the dynamical description provided by unitary operations to open quantum systems. We consider the characterization of such processes, by specializing to their Pauli transfer matrix representation. Channels are usually identified through quantum process tomography, by correlating a complete set of input states to a complete set of measurements at the output. Such correlations require multiple independent experiments to extract even a single Pauli transfer matrix element. We introduce a tomographic protocol that uses a specific class of input states, for which the set of Pauli measurements, performed at the output of the channel, directly relate to its Pauli transfer matrix components [3]. Although equivalent to standard quantum process tomography for complete channel reconstructions, we show that our method requires exponentially fewer experiments than that for a single matrix element characterization, under the same number of experimental repetitions. In Chapter 5, we consider the task of obtaining noiseless expectation values of noisy observables at the output of multiqubit quantum channels. We introduce a noise deconvolution technique, namely a tomographic reconstruction that acts like a postprocessing operation on the noisy measurement [4]. For Pauli channels, our method shows a quadratic speedup, providing the deconvolution from a set of Pauli measurements rescaled by a few components of the Pauli transfer matrix of the channel, bypassing the full process tomography and inversion of the noise map.

In Part III, we discuss the problem of estimating an unknown phase and the description of arrival time measurements. In Chapter 6, we present a global phase estimation protocol designed as a sequence of interferometric iterations, each providing one bit of the binary expansion of an unknown phase [5]. The protocol is optically implemented and specifically designed to reproduce square wave response

functions with different periods. This translates the estimation problem into a binary search task, allowing the reconstruction of the unknown phase at arbitrary accuracy. In contrast to conventional parallel and sequential strategies, our method reaches the Heisenberg bound independently of statistics, and not only asymptotically. Moreover, it does not rely on feedforward propagation: the protocol can be started, stopped, and restored at any level of accuracy, e.g. to improve the resolution of a previous estimation, or to leverage prior information about the first digits of the desired expansion. In Chapter 7, we discuss the problem of time measurements in quantum mechanics. Unlike position and momentum, time is not associated with any self-adjoint operator: it is not an observable. For this reason, simple questions, like predicting at what time a particle is most likely to reach a detector, cannot be answered using textbook quantum theory. This is called the arrival time problem. We review and compare the theoretical proposals that try to solve this issue. We identify multiple quantum regimes where different proposals give inequivalent and experimentally discriminable predictions [6], by considering different superpositions of Gaussian wave packets that travel towards the detector in different kinematic configurations. Our results show that multiple discrepancies appear in strongly quantum regimes, namely when the packets interfere at the detector in the time domain. Finally, we consider an explicit example, where a superposition of packets is obtained by applying Bragg diffraction to a Bose-Einstein condensate trapped in an accelerator ring. We discuss the feasibility of a hypothetical experiment involving this scenario.

This thesis presents the results of my PhD, during which I explored multiple interdisciplinary lines of research. Future developments are discussed in the conclusions section.

PART I

QUANTUM IMAGING AND IMAGE PROCESSING

CHAPTER 1

QUANTUM IMAGING

The transistor worked with the sense of sound
while the CCD worked with the sense of sight.

James M. Early by Willard S. Boyle [46]

DIGITAL IMAGES represent the visual information of one, or more, objects. Mathematically, they are described by matrices of elements called pixels, obtained at the output of an imaging apparatus. Usually, this consists of a source of light that emits a stream of photons towards the object. A sequence of lenses and optical devices focus the diffracted light on a two-dimensional plane, called image plane. A semiconductor detector, such as a charge-coupled device (CCD) [47], collects the photons and assigns a set of numerical values to a grid, depending on the light intensity measured at each point. Electronic processing provides the image as a collection of bitstrings, ready for storage or further manipulation by standard computational means.

In this chapter, we review the basic notions of wave optics and imaging, from a quantum-mechanical perspective. We consider a conventional imaging scenario, which differs from the classical one only in the choice of radiation states. In this framework, the shape of the object is encoded in the spatial modes of light, fed through an arbitrary lens system, and subsequently detected by a CCD. We first consider images obtained with single-photon states. Then, we generalize our treatment to multimode coherent states, i.e. the case of imaging with a classical source of light. After introducing the above formalism, we fully quantize the process of image capturing, by designing a multiatom lattice sensor that electromagnetically interacts with the photons, transferring the spatial pattern of the object from the optical image plane to the probabilities of a multiqubit quantum register. This method paves the way to the next chapter, which discusses the quantum compression of the spatial information captured here (although independently of this hardware implementation).

1.1 Preliminaries of quantum optics

In this section, we briefly summarize the quantization of the electromagnetic field, setting up the notation and the formalism. Readers familiar with basic notions of quantum optics and field theory, can skip to Section 1.2. Classical electrodynamics is described by the Maxwell's equations, which, in absence of sources and media, read

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 0 \,, \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \,, \\ \nabla \cdot \mathbf{B} &= 0 \,, \\ \nabla \times \mathbf{B} &= \partial_t \mathbf{E} \,, \end{aligned} \tag{1.1}$$

where we adopted units in which $c = 1$ [48]. Here, $\mathbf{E}$ and $\mathbf{B}$ denote the electric and magnetic fields, respectively. We introduce the potentials $\mathbf{A}$ and V, such that

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \,, \\ \mathbf{E} &= -\nabla V - \partial_t \mathbf{A} \,. \end{aligned} \tag{1.2}$$

By exploiting the freedom in the choice of the potentials, we impose the Coulomb's gauge-fixing condition, namely $\nabla \cdot \mathbf{A} = 0$. In this case, the Maxwell's equations reduce to the wave equations, which read

$$\Box \mathbf{A}(\mathbf{r}, t) = 0 \,, \tag{1.3}$$

where $\Box := \nabla^2 - \partial_t^2$ denotes the d'Alembert operator, with Cartesian coordinates $\mathbf{r} = (x, y, z)$ and $\nabla = (\partial_x, \partial_y, \partial_z)$. Consider the plane wave expansion of the potential, that is

$$\mathbf{A}(\mathbf{r}, t) = \int \mathrm{d}\omega \int \mathrm{d}^3 k \sum_s \exp(-i\omega t) \exp(i\mathbf{k} \cdot \mathbf{r}) \, U_{s\omega}(\mathbf{k}) \, \mathbf{e}_s(\mathbf{k}) \,, \tag{1.4}$$

with wave vector $\mathbf{k} = (k_x, k_y, k_z)$ and ω denoting the frequency associated with the temporal degree of freedom. Here, $s \in \{0, 1\}$ labels the two orthogonal independent polarizations of light, for which the gauge-fixing condition reads $\mathbf{k} \cdot \mathbf{e}_s(\mathbf{k}) = 0$. We temporarily neglect the polarizations degrees of freedom, which do not play any role in the computations below. In other words, we restrict to scalar waves $A(\mathbf{r}, t)$, i.e. to the case of the real Klein-Gordon field. By substitution of Eq. (1.4) in the (scalar) wave equation, we obtain the dispersion relation

$$\omega^2 = |\mathbf{k}|^2 \,, \tag{1.5}$$

which identifies the solutions of the wave equation in the energy-momentum space. We impose this constraint directly on the plane wave expansion, i.e. restricting to dynamical configurations on the lightcone hypersurface $\mathcal{C}$. In practice, we do this by restricting the integration to

$$\int \mathrm{d}\omega \int \mathrm{d}^3 k \; f_\omega(\mathbf{k}) \delta(\omega^2 - |\mathbf{k}|^2) = \int_{\mathcal{C}} \mathrm{d}^3 k \; f_\omega(\mathbf{k}) \,, \tag{1.6}$$

with $f_\omega(\mathbf{k})$ a generic solution and $\omega = \omega(\mathbf{k})$. Then, Eq. (1.4) reads

$$A(\mathbf{r},t) = \int_{\mathcal{C}} \mathrm{d}^3k \ \exp(-i\omega t)(i\mathbf{k}\cdot\mathbf{r})\, U_\omega(\mathbf{k}) \,. \tag{1.7}$$

This expression contains two contributions, which correspond to the positive (upper) and negative (lower) energy half-lightcones, denoted $\mathcal{C}^+$ and $\mathcal{C}^-$, i.e. the positive and negative frequencies solutions of the dispersion relation, for which $\omega = |\mathbf{k}| > 0$ and $\omega = -|\mathbf{k}| < 0$, respectively. Hence

$$\begin{aligned} A(\mathbf{r},t) = &\int_{\mathcal{C}^+} \mathrm{d}^3k \ \exp(-i\omega t)\exp(i\mathbf{k}\cdot\mathbf{r})\, U_\omega(\mathbf{k}) \\ &+ \int_{\mathcal{C}^-} \mathrm{d}^3k \ \exp(-i\omega t)\exp(i\mathbf{k}\cdot\mathbf{r})\, U_\omega(\mathbf{k}) \,. \end{aligned} \tag{1.8}$$

Mapping the second integral from the lower to the upper lightcone, i.e. through the substitution $\omega \to -\omega$, gives

$$\begin{gathered} A(\mathbf{r},t) = \int_{\mathcal{C}^+} \mathrm{d}^3k \ \left\{ a_\omega(\mathbf{k}) \exp[-i(\omega t - \mathbf{k}\cdot\mathbf{r})] + b_\omega(\mathbf{k}) \exp[i(\omega t + \mathbf{k}\cdot\mathbf{r})] \right\} , \\ \text{with } a_\omega(\mathbf{k}) = U_{|\mathbf{k}|}(\mathbf{k}) \text{ and } b_\omega(\mathbf{k}) = U_{-|\mathbf{k}|}(\mathbf{k}) \,. \end{gathered} \tag{1.9}$$

The conjugated field reads

$$A^*(\mathbf{r},t) = \int_{\mathcal{C}^+} \mathrm{d}^3k \ \left\{ a^*_\omega(\mathbf{k}) \exp[i(\omega t - \mathbf{k}\cdot\mathbf{r})] + b^*_\omega(\mathbf{k}) \exp[-i(\omega t + \mathbf{k}\cdot\mathbf{r})] \right\} . \tag{1.10}$$

For a real field $A = A^*$. Combining this condition with the substitution $\mathbf{k} \to -\mathbf{k}$ yields

$$b_\omega(\mathbf{k}) = a^*_\omega(-\mathbf{k}) \,. \tag{1.11}$$

By substituting this condition in Eq. (1.9), followed by another change of variable $\mathbf{k} \to -\mathbf{k}$, we finally get

$$\mathbf{A}(\mathbf{r},t) = \int_{\mathcal{C}^+} \mathrm{d}^3k \sum_s a_{s\omega}(\mathbf{k})\mathbf{e}_s(\mathbf{k}) \exp[i(-\omega t + \mathbf{k}\cdot\mathbf{r})] + \text{H.c.} \,, \tag{1.12}$$

where we have restored the polarization subscript. From Eq. (1.2), the mode expansion of the electric and magnetic fields reads

$$\mathbf{E}(\mathbf{r},t) = i \int_{\mathcal{C}^+} \mathrm{d}^3k \ |\mathbf{k}| \sum_s a_{s\omega}(\mathbf{k})\mathbf{e}_s(\mathbf{k}) \exp[i(-\omega t + \mathbf{k}\cdot\mathbf{r})] + \text{H.c.} \,, \tag{1.13}$$

$$\mathbf{B}(\mathbf{r},t) = i \int_{\mathcal{C}^+} \mathrm{d}^3k \sum_s a_{s\omega}(\mathbf{k})[\mathbf{k}\times\mathbf{e}_s(\mathbf{k})] \exp[i(-\omega t + \mathbf{k}\cdot\mathbf{r})] + \text{H.c.} \,. \tag{1.14}$$

Consider a cavity of finite volume, in which the electromagnetic field is confined. Under the above expansion, the total energy of the field is given by the sum of the

energies of infinitely many independent harmonic oscillators, each vibrating with frequency $\omega(\mathbf{k})$ $\forall \mathbf{k}$. By imposing the commutation relations on the canonical variables associated with each mode, the quantization of the field is obtained by promoting the coefficients of the mode expansion to creation and annihilation operators, which obey

$$[a_s(\mathbf{k}), a_{s'}(\mathbf{k'})] = 0 \ , \quad [a_s^\dagger(\mathbf{k}), a_{s'}^\dagger(\mathbf{k'})] = 0 \ , \tag{1.15}$$

$$[a_s(\mathbf{k}), a_{s'}^\dagger(\mathbf{k'})] = \delta_{ss'}\delta(\mathbf{k}-\mathbf{k'}) \ , \tag{1.16}$$

where we adopted units in which $\hbar = 1$ [49, 50]. Indeed, $a_s(\mathbf{k})$ and $a_s^\dagger(\mathbf{k})$ act as operators on the Fock space

$$\mathcal{F} = \bigoplus_{p=1}^{+\infty} \mathcal{H}^{\otimes p} \ , \tag{1.17}$$

such that

$$a_s(\mathbf{k}) : \mathcal{H}^{\otimes p} \to \mathcal{H}^{\otimes p-1} \ , \tag{1.18}$$

$$a_s^\dagger(\mathbf{k}) : \mathcal{H}^{\otimes p} \to \mathcal{H}^{\otimes p+1} \ . \tag{1.19}$$

Here, $\mathcal{H}$ denotes the Hilbert space generated by the single-photon states, and $p \in \{0, 1, 2, \ldots\}$ the number of photons in each subspace. From now on, we consider only modes described by the operators $a(\mathbf{k})$ and $a^\dagger(\mathbf{k})$, i.e. after implicitly summing over the polarization degrees of freedom. There exist several other strategies to quantize the electromagnetic field, besides the harmonic oscillator decomposition. For example, one can use the optical transfer function [51]. In the Lagrangian formalism, the quantization is achieved by imposing the commutation relation directly on the field and its conjugate momentum [52, 53]. Alternatively, the quantization can be performed by purely algebraic means, bypassing the mode expansion of the field, and imposing the canonical commutation relations on the causal propagator generated by the fundamental solutions of the wave equation [54], thus encoding the symplectic structure of the classical theory [55]. This framework is consistent with the Gelfand-Naimark-Segal representation theorem , and it can also be generalized to curved spacetimes [56, 57]. Departing from the Coulomb gauge brings further difficulties, which can be treated by removing the contribution of timelike and longitudinal photons, e.g. using the Gupta-Bleuler condition [58] or through the Fadeev-Popov trick [59].

1.2 States of light

In the Fock space, pure states of light are described in terms of the eigenstates of the continuous number operator $|p\rangle_{\mathbf{k}}$, also called number states. Namely

$$\begin{aligned} |\Psi\rangle &= \sum_p c_p \left[A^\dagger(\Psi)\right]^{\otimes p} |0\rangle \ , \\ \text{with } A^\dagger(\Psi) &= \int \mathrm{d}^3k \ \hat{\Psi}(\mathbf{k}) a_\omega^\dagger(\mathbf{k}) \ , \end{aligned} \tag{1.20}$$

where $a^\dagger(\mathbf{k})\,|p\rangle_\mathbf{k} = \sqrt{p+1}\,|p+1\rangle_\mathbf{k}$ and $a(\mathbf{k})\,|p\rangle_\mathbf{k} = \sqrt{p}\,|p-1\rangle_\mathbf{k}$. Here, Ψ describes the spatial spectrum of the multimode field and $A^\dagger(\Psi)$ the multimode creation operator. From now on, $|0\rangle$ identifies the vacuum state, with the hat operator denoting the Fourier transform. Mathematically, Eq. (1.20) represents a weighted expansion in the (multimode) number states $\left[A^\dagger(\Psi)\right]^{\otimes p}|0\rangle$, with amplitudes c_p that fulfil the normalization condition $\sum_p |c_p|^2 = 1$. Equivalently, such expansion can be rewritten in terms of a generic map Λ, such that

$$|\Psi\rangle = \Lambda\left(A^\dagger(\Psi)\right)|0\rangle \ . \tag{1.21}$$

When Λ is the identity, this expression identifies a single-photon state. When Λ is the multimode displacement operator

$$\Lambda\left(A^\dagger(\Psi)\right) = \exp\left(\alpha A^\dagger(\Psi) - \alpha^* A(\Psi)\right) , \tag{1.22}$$

Eq. (1.21) describes a multimode coherent state with amplitude α. If $\hat{\Psi}(\mathbf{k}) = \delta(\mathbf{k} - \mathbf{k'})$, i.e. when the coherent spectrum contains only one mode, such expression reduces to the Glauber prescription

$$|\alpha\rangle = \exp\left(-\frac{1}{2}|\alpha|^2\right)\sum_p \frac{\alpha^p}{\sqrt{p!}}\,|p\rangle_{\mathbf{k'}} \ , \tag{1.23}$$

with $|\alpha|^2$ the intensity of the field [60]. Equivalently, the same expression can be directly obtained from the Fock representation, assuming a single-mode spectrum, i.e. $A \to a$, and

$$c_p = \exp\left(-\frac{1}{2}|\alpha|^2\right)\frac{\alpha^p}{\sqrt{p!}} \ . \tag{1.24}$$

In this case, the probability of measuring a state with p photons is described by the Poisson distribution. Eq. (1.21) does not include thermal states, which can be represented as a statistical mixture of number states, with the thermal distribution as normal-ordered two-point function [61].

1.3 Transfer functions

We introduce transfer functions, a mathematical description of linear optical operations that preserve the photon count on the Fock space. Examples include free-space propagation, beam splitters, lenses and phase transformations. In the Schrödinger picture, the evolution of a Fock state, against a unitary operator U, reads

$$|\Psi\rangle \to U\Lambda\left(A^\dagger(\Psi)\right)|0\rangle \ . \tag{1.25}$$

Consider operations that preserve the number of photons, namely that $U\,|p\rangle_\mathbf{k} = |p\rangle_\mathbf{k}$, up to a phase factor. Under such evolution, the vacuum state is invariant, i.e. $U\,|0\rangle = |0\rangle$. By substitution in Eq. (1.20), this yields

$$U\,|\Psi\rangle = \Lambda\left(UA^\dagger(\Psi)U^\dagger\right)|0\rangle \ . \tag{1.26}$$

In particular, the multimode creation operator undergoes

$$\begin{aligned} UA^\dagger(\Psi)U^\dagger &= \int \mathrm{d}^3k\ \hat{\Psi}(\mathbf{k})\hat{\mathcal{L}}(\mathbf{k})a^\dagger_\omega(\mathbf{k}) \\ \text{with } Ua^\dagger_\omega(\mathbf{k})U^\dagger &= \hat{\mathcal{L}}(\mathbf{k})a^\dagger_\omega(\mathbf{k})\ . \end{aligned} \tag{1.27}$$

We call $\mathcal{L}$ transfer function. For example, consider a single-photon state with spectrum Φ. Its evolution against a linear optical system, with transfer function $\mathcal{L}$, reads

$$|\Phi\rangle \to U|\Phi\rangle = \int \mathrm{d}^3k\ \hat{\Phi}(\mathbf{k})\hat{\mathcal{L}}(\mathbf{k})a^\dagger_\omega(\mathbf{k})|0\rangle\ , \tag{1.28}$$

This state can be written in the spatial domain, by applying the convolution theorem, which yields

$$\begin{aligned} U|\Phi\rangle &= \int \mathrm{d}^3r\ [\Phi * \mathcal{L}](\mathbf{r})a^\dagger_\omega(\mathbf{r})|0\rangle\ , \\ \text{with } a^\dagger_\omega(\mathbf{r}) &= \int \mathrm{d}^3k\ a^\dagger_\omega(\mathbf{k})\exp(-i\mathbf{k}\cdot\mathbf{r})\ , \end{aligned} \tag{1.29}$$

where $*$ denotes the convolution operation. Below, we adopt this framework to represent the spatial transformation of light that passes through a sequence of linear optical transformations, such as an imaging apparatus.

1.4 Quantum imaging

Consider a generic single-photon state, generated by a monochromatic source with longitudinal position z

$$|\Phi\rangle = \int \mathrm{d}^3k\ \hat{\Phi}(\mathbf{k})a^\dagger_\omega(\mathbf{k})|0\rangle\ , \tag{1.30}$$

where Φ is spatial spectrum and $\mathbf{k} = (k_x, k_y, k_z)$ the wave vector. We neglect the polarization of the photon and consider the single-frequency-mode assumption [60], i.e. we assume that the wavefront propagates along positive z-directions only. Then, $k = (k_x, k_y)$ represents the sole independent degrees of freedom of the single-photon state, which reads

$$|\Phi\rangle = \int \mathrm{d}^2k\ \hat{\phi}_\omega(k)a^\dagger_\omega(k)|0\rangle\ . \tag{1.31}$$

In the spatial domain this equivalently yields

$$|\Phi\rangle = \int_S \mathrm{d}^2r\ \phi_\omega(r)a^\dagger_\omega(r)|0\rangle\ , \tag{1.32}$$

where

$$\hat{\phi}_\omega(k) = \hat{\Phi}\left(k_x, k_y, \sqrt{\omega^2 - k_x^2 - k_y^2}\right)\ , \tag{1.33}$$

and $r = (r_x, r_y)$ labels the transverse coordinates on the source plane S. Suppose that the source is placed at the longitudinal origin $z = 0$. Consider an object with two-dimensional shape $\mathcal{O}$, placed at longitudinal position z_o. After free-space propagation

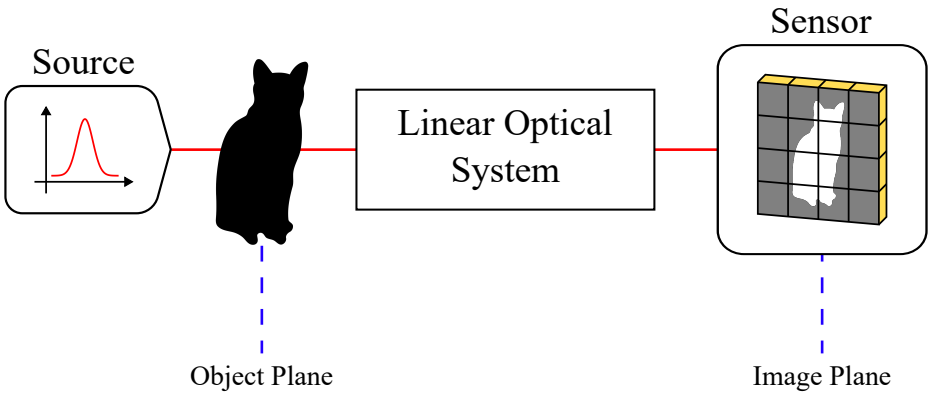


Figure 1.1: Representation of a digital imaging system. A source emits a shower of photons towards the object, which acts as an amplitude-modulating mask. A linear optical system focuses the incoming light, forming a real image on a two-dimensional plane, i.e. the image plane. Here, the photons are collected by a sensor, which produces the image as an ensemble of pixels, whose values depend on the brightness, i.e. the light intensity, observed at each point. In this chapter, we replace the conventional sensor with a multiatom lattice device. At its output, the fingerprint of the object is encoded in the probabilities of a multiqubit register.

occurs, the single-photon spectrum undergoes spatial amplitude modulation [62], that is

$$\Psi_{\text{in}}(r) \to \mathcal{O}(r)\Psi_{\text{in}}(r)\,, \tag{1.34}$$

where $\Psi_{\text{in}}(r)$ is the input wavefront emitted by the source and propagate to the object plane O. Explicitly, the output reads

$$|\mathcal{O}\rangle = \int_O \mathrm{d}^2 r\, \left[\phi_\omega * \mathfrak{H}_{z_o}\right](r)\mathcal{O}(r)a^\dagger_\omega(r)\,|0\rangle\,, \tag{1.35}$$

where $\mathfrak{H}_{z_o}$ denotes the free-space transfer function between the S and O planes, and $*$ the convolution operation. Using twice the convolution theorem, it follows that

$$|\mathcal{O}\rangle = \int \mathrm{d}^2 k\, \left[\left(\hat{\phi}_\omega \mathfrak{H}_{z_o}\right) * \hat{\mathcal{O}}\right](k)a^\dagger_\omega(k)\,|0\rangle\,. \tag{1.36}$$

Consider an arbitrary, linear optical system with transfer function $\mathcal{L}$, with image plane at longitudinal position z_i. Using Eq. (1.28), which reads

$$\begin{gathered}\hat{\Psi}_{\text{in}}(k) \to \hat{\mathcal{L}}_d(k)\hat{\Psi}_{\text{in}}(k)\,,\\ \text{with } d = z_i - z_o\,,\end{gathered} \tag{1.37}$$

the state on the image plane becomes

$$\begin{gathered}|\Psi\rangle = \int \mathrm{d}^2 k\, \hat{\Psi}(k)a^\dagger_\omega(k)\,|0\rangle\,,\\ \text{with } \hat{\Psi} = \left[\left(\hat{\phi}_\omega \mathfrak{H}_{z_o}\right) * \hat{\mathcal{O}}\right]\hat{\mathcal{L}}_d\,.\end{gathered} \tag{1.38}$$

This state encodes the image as it would be observed by placing a screen at distance d from the object, after gathering enough statistics. In the spatial domain, it reads

$$\Psi = \left[\left(\phi_\omega * \mathfrak{H}_{z_o}\right)\mathcal{O}\right] * \mathcal{L}_d \ , \tag{1.39}$$

Similarly, the single-photon state can be substituted with a multimode coherent state, with intensity $|\alpha|^2$. Namely

$$|\Psi_\alpha\rangle = D(\alpha,\Psi)\,|0\rangle \ . \tag{1.40}$$

In this case

$$D(\alpha,\Psi) = \exp\!\left(\alpha A^\dagger(\Psi) - \alpha^* A(\Psi)\right) , \tag{1.41}$$

$$A^\dagger(\Psi) = \int \mathrm{d}^2k\ \hat{\Psi}(k) a^\dagger_\omega(k) \ , \tag{1.42}$$

are the multimode displacement operator and multimode creation operator on the transverse plain, respectively, and Ψ is the multimode spatial spectrum obtained at the image plane, given by Eq. (1.39). We summarize this setup in Fig. 1.1. The above discussion holds in the monochromatic case only, i.e. for grayscale images. Colours can be included by either dropping the single-frequency-mode assumption, namely through a more complex description of diffraction as well as a color-sensitive detection, or by repeating the procedure of Fig. 1.1 multiple times, each with a different red, green, and blue (RGB) filter placed after the source.

1.5 Quantum camera in the Tavis-Cummings model

We show how to capture the image into a multiqubit register, by designing a multiatom lattice sensor, i.e. a collection of atoms that interact with the multimode coherent input state, whose spatial arrangement mimics that of a CCD. Consider a lattice made of N identical atoms, with evenly-spaced energy levels and position labelled by (m,n). Each atom is modelled as a two-level system, i.e. a qubit with computational basis $\{|0\rangle_{mn}, |1\rangle_{mn}\}$. The lattice is initialized in the vacuum state

$$|b\rangle = \bigotimes_{mn} |0\rangle_{mn} \ , \tag{1.43}$$

which represents a black image, with each qubit in the ground state. We model the interaction between the lattice and the electromagnetic field using a multimode and multiatom Jaynes-Cummings Hamiltonian [63–65], also known as multimode Tavis-Cummings model [66]. In the resonant rotating wave approximation [49], the free Hamiltonian reads

$$H_F = \frac{1}{2}\sum_{mn} \omega_0 \sigma^{(3)}_{mn} + \int \mathrm{d}^2k\ \omega(k) a^+_\omega(k) a_\omega(k) \ , \tag{1.44}$$

while the interaction Hamiltonian yields

$$H_I = \int \mathrm{d}^2k \sum_{mn} \gamma^*_{mn}(k) a_\omega(k) \sigma^+_{mn} + \text{ H.c. } , \tag{1.45}$$

with

$$\sigma^{+}_{mn} = |1\rangle_{mn}\langle 0| \ , \ \ \sigma^{-}_{mn} = |0\rangle_{mn}\langle 1| \ , \tag{1.46}$$

$$\sigma^{(3)}_{mn} = |1\rangle_{mn}\langle 1| - |0\rangle_{mn}\langle 0| \ . \tag{1.47}$$

Here, γ_{mn} is the Fourier transform of g_{mn}, which describes the coupling between the coherent photons and the qubit located at (m,n). We assume that g_{mn} is compactly supported $\forall m,n$ and that $\mathrm{supp}(g_{mn}) \cap \mathrm{supp}(g_{m'n'}) \simeq 0$. The total Hamiltonian reads $H = H_F + \lambda H_I$, where λ is an additional perturbative parameter. We describe the evolution of the system. Consider a shutter that initially prevents the photons from interacting with the qubits. The bipartite light-sensor initial state reads

$$|i\rangle = |\Psi_\alpha\rangle \otimes |b\rangle \ . \tag{1.48}$$

When the shutter opens, each photon releases energy in the sensor, exciting one of the qubits as $|0\rangle_{mn} \to |1\rangle_{mn}$. Since $[H_F, H_I] = 0$, the dynamics is completely driven by the unitary operators generated by H_I, so that

$$U_t = \mathbb{1} + it\lambda H_I + \mathcal{O}(\lambda^2) \ . \tag{1.49}$$

The output state reads

$$|f\rangle = U_t\,|i\rangle = |\Psi_\alpha\rangle \otimes |b\rangle + it\lambda H_I\,|i\rangle + \mathcal{O}(\lambda^2) \ . \tag{1.50}$$

Intuitively, a single interaction may give no excitation, a single excitation, or multiple qubits excitations, described by the following contributions to the final state:

$$\begin{gathered} c_0^0\,|b\rangle + c_0^1\,|0\ldots001\rangle + c_1^1\,|0\ldots010\rangle + \ldots \\ + c_{N-1}^1\,|10\ldots00\rangle + c_0^2\,|0\ldots011\rangle + \ldots \end{gathered} \tag{1.51}$$

We neglect the vacuum contribution in post-selection, by discarding all the completely-black images reconstructed at the output. Then, the leading order contribution becomes dominant and the evolution is solely controlled by the interaction terms

$$|f\rangle \simeq it\lambda \int \mathrm{d}^2k \ \sum_{mn} \gamma^*_{mn}(k) a_\omega(k) \sigma^{+}_{mn}\,|\Psi_\alpha\rangle \otimes |b\rangle \ , \tag{1.52}$$

where the Hermitian conjugate is annihilated by $|b\rangle$. The multimode coherent state satisfies

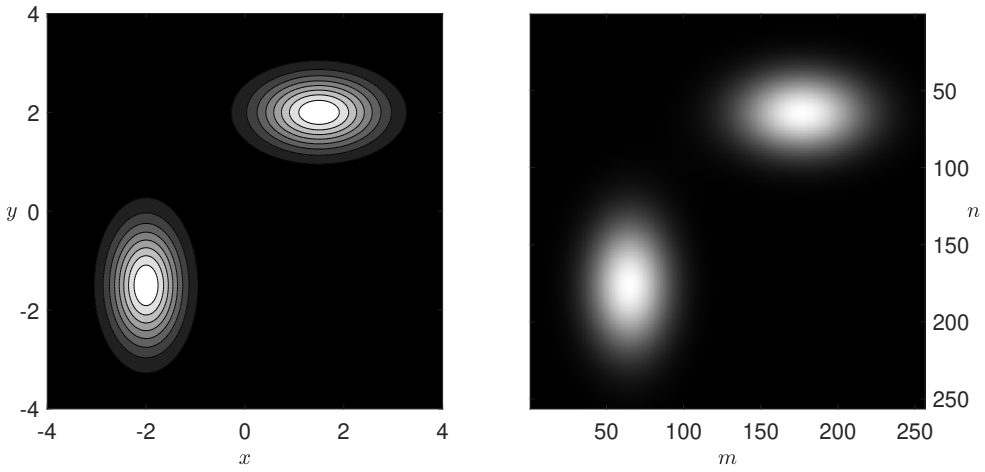
$$D(\alpha,\Psi)\,|0\rangle \simeq \prod_k \exp\!\big(\hat{\Psi}(k) a^\dagger_\omega(k) + \ \mathrm{H.c.}\big) \ , \tag{1.53}$$

which implies that

$$a_\omega(k)\,|\Psi_\alpha\rangle = \alpha\hat{\Psi}(k)\,|\Psi_\alpha\rangle \ . \tag{1.54}$$

A projection on the coherent state gives $|\Omega\rangle = \langle\Psi_\alpha\,|f\rangle$. Substituting the Fourier transform of Ψ and g^*_{mn} yields

$$|\Omega\rangle = it\lambda\alpha \sum_{mn} \int \mathrm{d}^2r \ \Psi(r) g^*_{mn}(r) \sigma^{+}_{mn}\,|b\rangle \ . \tag{1.55}$$



(a) (b)

Figure 1.2: Qualitative comparison between the spectral amplitude Ψ and the multiqubit quantum state $|\Omega\rangle$, encoded using the multimode and multiatom Jaynes-Cummings Hamiltonian. Quantitative compatibility is inferred from Eq. (1.56). (a) Contour plot of $|\Psi|^2$, with parameters $x_0 = -y_1 = 1.5$, $y_0 = -x_1 = 2$, $\alpha_0 = \beta_1 = 0.35$, $\alpha_1 = \beta_0 = 1$. (b) Probabilities of $|\Omega\rangle$, plotted as a grayscale image, with $m, n = 0, 1, \ldots, 255$. Figure adapted from [1].

Here, the interaction term $g_{mn}(r)$ acts as a spatial transfer function that discretizes the image in $\mathrm{supp}(g_{mn})$, encoding the light intensity in the probability amplitude of the corresponding qubit to occupy the higher energy level state. Each qubit identifies the spatial region $\mathrm{supp}(g_{mn})$, i.e. a pixel with position $r_{mn} = (x_m, y_n)$, over which the multimode coherent spectrum is approximated by the average

$$\Psi(r_{mn}) \to \int \mathrm{d}^2 r \; \Psi(r) g^*_{mn}(r) \; . \tag{1.56}$$

Such identification is exact for $g_{mn}(r) = \delta(r - r_{mn})$, ideally for a camera with perfect spatial resolution.

1.6 Example with Gaussian packets

In this section, we compare the probabilities encoded from a superposition of two Gaussian packets. We consider the multimode coherent spatial spectrum

$$\Psi(x, y) = K \sum_{i=0}^{1} \exp\left[-\alpha_i (x - x_i)^2 - \beta_i (y - y_i)^2\right] , \tag{1.57}$$

and normalization constant K. The sensor is an evenly-spaced lattice of identical two-level systems, with spatial coupling functions

$$g_{mn}(x, y) = \exp\left[-5(x - x_m)^2 - 5(y - y_n)^2\right] \tag{1.58}$$

and coordinates $(x_m, y_n) \in [-4,4] \times [-4,4]$. Fig. 1.2 compares the numerical evaluation of Eq. (1.55) with the (theoretical) probability density $|\Psi|^2$. Qualitatively, the data correctly reproduces the pattern of $|\Psi|^2$, but discretized over the chosen lattice.

1.7 Discussion

In this chapter, we introduced the basic notions and formalism of quantum optics and quantum imaging. In this framework, we showed how to encode the output of an imaging system into a multiqubit state, using a multiatom lattice sensor. Modelled by the Jaynes-Cummings Hamiltonian, we designed a hardware encoder that maps the light intensity at each lattice location to the probability amplitude of the corresponding qubit of being excited by a single-photon interaction. Its output represents the image using a sub-optimal encoding, with contributions of the form

$$\bigotimes_{\substack{m' \neq m \\ n' \neq n}} |0\rangle_{m'n'} \otimes |1\rangle_{mn} \quad \forall m, n \ . \tag{1.59}$$

We can further optimize this representation by appending a one-hot to binary converter [67, 68], which uses a combination of CNOTs operation to re-encode the image in $n_0 = \log_2(N)$ qubits. Notice that this is a first lossless compression, from one-hot (or unary) to binary encoding, which consists in reassigning the probabilities of $|\Omega\rangle$ to a smaller, but optimally occupied, register. After this operation, we get

$$|\Omega\rangle \to |\Psi\rangle_0 = \sum_j w_j \, |j\rangle_0 \ . \tag{1.60}$$

This conversion requires $\mathcal{O}(N)$ operations, which is the same readout cost of the CCD bidirectional shift registers.

CHAPTER 2

QUANTUM DOWNSAMPLING

If your pictures aren't good enough,
you aren't close enough.

Robert Capa

QUANTUM IMAGE PROCESSING seeks to encode, manipulate, and retrieve visual information in a quantum-mechanical way. Historically, the quantum representation of images has been introduced under a pure classical analogy, by means of a bijective identification between pixels and qubits [69]. Alternative strategies have been considered so far, which provide better encoding and processing capabilities [34, 70–72]. For example, images can be represented as bipartite states, where each pixel has value described by a single-qubit rotation, and position encoded in the bitstrings of a multiqubit register [73]. When the basis elements also encode the binary expansion of the pixel value, such representation provides a deterministic image retrieval [74]. Images can also be vectorized and encoded as multiqubit quantum states, by identifying each pixel value with the probability of observing one of the computational basis eigenstates. This approach is connected to the amplitude encoding technique, commonly employed in quantum machine learning [75], and requires fewer resources than the other ones, i.e. $\log_2(N)$ qubits for N pixels [76]. Importantly, it is fully compatible with the output of the hardware encoder introduced in the previous chapter.

Formerly, we discussed how to achieve a quantum image representation in an optical way, bypassing quantum state preparation techniques while occupying the entire configuration space of the system. In this chapter, we discuss how to compress such images, reducing the size of the encoding register by quantum computational means. In classical image processing, compression and downsampling algorithms can reduce the cost for data storage and transmission, while preserving the original visual pattern [77]. An example is the joint photographic expert group (JPEG) algorithm [78],

which operates in the spatial-frequency domain. This algorithm divides the input image into smaller subimages, taking the discrete cosine transform of each element. The high spatial frequencies are removed, reducing the amount of information stored at the cost of the quality of the output image, i.e. a lossy compression. Similarly, we introduce a lossy downsampling algorithm that uses the quantum Fourier transform (QFT) to discard the most significant qubits in the input register, thus filtering the high spatial-frequency (low probability) components of the image. The compressed register is mapped back to the computational basis, now encoding an image with reduced number of pixels, leaving the original pattern visually unaffected. This operation is schematically represented in Fig. 2.1. In terms of pure operations, our algorithm provides an exponential advantage over its classical counterpart. When statistics and image retrieval are considered, we show that an advantage remains, depending on the trade-off between the input and the output resolutions.

This chapter is built upon the results of [1]. The underlying code that generated the data for this study is openly available in GitHub [79].

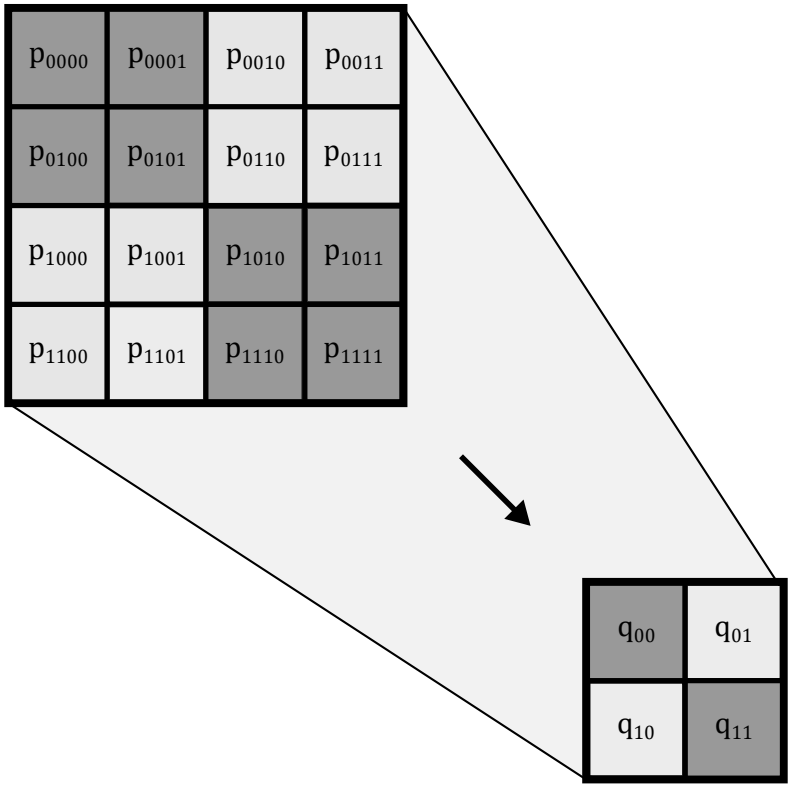


Figure 2.1: Downsampling of 4×4 image to a 2×2 resolution. The input data are encoded in the probabilities of a 4-qubit state. At the output, the compressed image is retrieved from at 2-qubit state.

2.1 Amplitude encoding

We mathematically discuss the encoding. Consider a digital grayscale image of an object, namely a matrix of N pixels. Each pixel is associated with a finite and discrete quantity called gray value, representing its brightness as a result of the object transmissivity. Classically, each value is encoded in a string of c bits, with a total number of $L = 2^c$ gray levels of the image. From a quantum-mechanical point of view, we encode the image in quantum state of n_0 qubits, that is

$$|\Psi\rangle_0 = \sum_j \sqrt{V(\mathscr{I})_j}\,|j\rangle_0 \ , \tag{2.1}$$

where $|j\rangle_0$ labels the computational basis elements on the Hilbert space, with $j \in \{0,\dots,2^{n_0}-1\}$ and $\mathscr{I}$ the (classical) digital image. We specialize to square images only, although the following discussion can be generalized to multidimensional arrays of arbitrary shape [7]. From now on, states and operations with subscript 0 are referred to the n_0-qubit register. In Eq. (2.1), V is a row-wise vectorization operations that maps matrices to column vectors, namely

$$V(\mathscr{I}) = [\mathscr{I}_{00}, \mathscr{I}_{01}, \dots]^T \ . \tag{2.2}$$

The choice of the vectorization path is conventional, as long as it is respected throughout the algorithm. This representation encodes the gray values of N pixels in the probabilities of $n_0 = \log_2(N)$ qubits, normalized by rescaling the whole image to its total brightness. This strategy yields a first lossless exponential compression of the image: from N pixels to n_0 qubits, which can be optically achieved by feeding the output of the Tavis–Cummings hardware to a one-hot to binary converter, as discussed in the previous chapter. Alternatively, the amplitude encoding can be computationally achieved by means of the Grover-Rudolph technique [80]. The algorithm works by bisecting the state preparation scheme. At each step, it appends a qubit to the register and applies a controlled rotation, doubling the number of available configurations and distributing the target amplitudes. Thus, any probability distribution that is efficiently integrable on N bins, can be prepared using $\log_2(N)$ controlled rotations [81].

2.2 Quantum downsampling

We introduce a quantum algorithm for image downsampling. After the encoding and a QFT, it traces out the qubits that correspond to the high spatial-frequency components of the image, reducing its resolution while preserving the original pattern. Back to the computational basis, the image is compressed to a register of fewer qubits. The QFT operates on the n_0-qubit register as [82]

$$U_{\mathrm{QFT}_0}\,|j\rangle_0 = \frac{1}{\sqrt{N}} \sum_k \exp(2\pi i j k/N)\,|k\rangle_0 \ . \tag{2.3}$$

Consider the n_1-qubit subregister made only of the first $n_1 = n_0 - \tilde{n}$ qubits of the initial one, and let $|m\rangle_1$ label the elements of its computational basis, with $m \in \{0,\dots,2^{n_1}-$

1}. Then $|j\rangle_0 = |l\rangle_{\tilde{n}} \otimes |m\rangle_1$, with $|l\rangle_{\tilde{n}}$ defined on the $\tilde{n}$-qubit register. On the n_1-qubit subregister, the QFT operates as

$$(\mathbb{1} \otimes U^{\dagger}_{\mathrm{QFT}_1}) |j\rangle_0 = |l\rangle_{\tilde{n}} \otimes U^{\dagger}_{\mathrm{QFT}_1} |m\rangle_1 \ . \tag{2.4}$$

After choosing $\tilde{n}$, the following algorithm downsamples the image $|\Psi\rangle_0$ from the n_0-qubit register to the n_2-qubit register, reducing the number of encoding qubits from n_0 to $n_2 = n_0 - 2\tilde{n}$. The number of pixels of the output image is $N/2^{2\tilde{n}}$. In the following, we adopt the little-endian ordering, with the least significant qubit placed on the top of the register and labelled by $q = 0$, with $0 \leq q \leq m-1$ and $m \in \{n_0, n_1, n_2\}$. For example, consider the 3-qubit state $|abc\rangle = |a\rangle \otimes |b\rangle \otimes |c\rangle$, with $a, b, c = 0, 1$. The little-endian ordering interprets a as the most significant qubit in the register, which, according to the above notation, is labelled by $q = 3$. We define

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} , \tag{2.5}$$

the single-qubit Hadamard gate.

Algorithm 2.1 Quantum downsampling

Input: Input image $|\Psi\rangle_0$, integer $\tilde{n} < n_0/2$

Output: Downsampled image

```
apply H^{⊗n_0}                                   ▷ n_0-qubit register
apply U_{QFT_0}
for q in n_0-qubit register do
    if n_0 − ñ ≤ q ≤ n_0 − 1 then
        discard the qth qubit                    ▷ Rule 1 (n_0 → n_1)
apply U†_{QFT_1}                                 ▷ n_1-qubit register
for q in n_1-qubit register do
    if n_0/2 − ñ ≤ q ≤ n_0/2 − 1 then
        discard the qth qubit                    ▷ Rule 2 (n_1 → n_2)
apply H^{⊗n_2}                                   ▷ n_2-qubit register
```

The algorithm operates as follows. After applying the Hadamard gates $H^{\otimes n_0}$, take a QFT on the register in which $|\Psi\rangle_0$ is initially encoded. Discard (or simply ignore) the first $\tilde{n}$ most significant qubits from the bottom of the register, which means averaging out the low-probability frequency components of the original image. We call this Rule 1. Take the QFT†. Up to this step, the image resolution is reduced by a factor $2^{\tilde{n}}$ along only one of its axes. After the QFT†, discard the last $\tilde{n}$ positions of the first half of the initial register, which contain redundant information, then take the Hadamard gates $H^{\otimes n_2}$. As we show below, a measurement on the output state ρ_2 yields an image of $N/2^{2\tilde{n}}$ pixels, conformally downscaled along both its axes. There are different implementations of this protocol, which can achieve the Fourier compression on both the image axes, simultaneously [7]. Qualitatively, these generalizations produce compatible results, noticeably optimizing the complexity of the whole algorithm. We

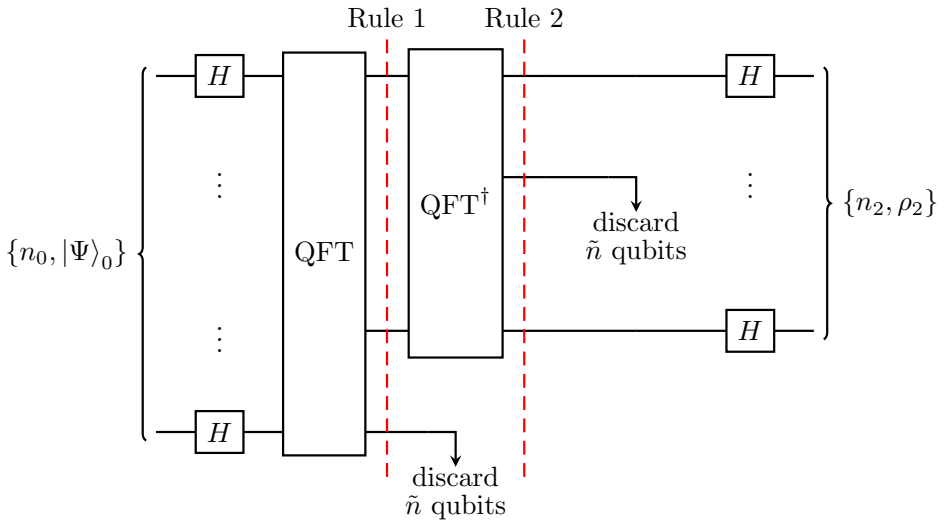


Figure 2.2: Quantum image downsampling to a register of fewer qubits. An image of N pixels is initially encoded in the probabilities of the n_0-qubit state $|\Psi\rangle_0$. The most significant qubit is placed at the bottom of the register. Take the QFT. Apply Rule 1, i.e. chosen an integer $\tilde{n} < n_0/2$, take a QFT† on the first $n_0 - \tilde{n}$ qubits, and discard the others. Rule 2 follows, i.e. discard the same number of qubits from the n_1-qubit register, whose positions correspond to the last $\tilde{n}$ qubits of the first half of the n_0-register. The number of pixels is reduced by a factor $2^{2\tilde{n}}$. Figure adapted from [1].

report the circuit implementation of Algorithm 2.1 in Fig. 2.2. Finally, we remark that the downsampling algorithm works independently of the presence of the Hadamard gates. As we also show in the next sections, their use improves the quality of the output, reducing the statistical fluctuations at each pixel while preserving the original image contrast.

2.2.1 Example of a discarding pattern

As an example of Rules 1 and 2, consider a 512×512 image encoded in an 18-qubit register, with qubits labelled by the list $[0, 1, \ldots, 17]$. Suppose that $\tilde{n} = 4$. The discarding sequence reads

$$\begin{aligned}[0,1,\ldots,17] &\xrightarrow[\text{Discard } [14,15,16,17]]{\text{Rule 1}} [0,1,\ldots,13] \\ &\xrightarrow[\text{Discard } [5,6,7,8]]{\text{Rule 2}} [0,1,2,3,4,9,10,11,12,13]\,.\end{aligned} \tag{2.6}$$

Combined with the QFT and QFT†, this yields a downsampled 32×32 image, encoded in a 10-qubit register of qubits $[0, 1, 2, 3, 4, 9, 10, 11, 12, 13]$.

2.2.2 Example with a lower triangular matrix

We provide a complete example of the downsampling algorithm for a simple 4×4 pattern, as well as a first assessment on the improvements brought by the Hadamard gates. To this extent, we first perform the calculation without the Hadamard gates and compare the result when these are added according to Algorithm 2.1. Consider a lower triangular 4×4 image, and the expected downsampled 2×2 matrix

$$\mathscr{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \rightarrow \widetilde{\mathscr{I}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} . \tag{2.7}$$

We vectorize and represent $\mathscr{I}$ as 4-qubit state

$$\mathscr{I} \rightarrow \frac{1}{\sqrt{8}} [1\,0\,0\,0\,1\,1\,0\,0\,1\,1\,1\,0\,1\,1\,1\,1]^T \rightarrow |\Psi\rangle_0 \ . \tag{2.8}$$

Choose $\tilde{n} = 1$, so that $n_2 = 2$. By postponing the discarding operations of Rule 1 and 2, the downsampling algorithm reads

$$|\Phi\rangle_0 = (\mathbb{1} \otimes U^{\dagger}_{\mathrm{QFT}_1})\, U_{\mathrm{QFT}_0} |\Psi\rangle_0 \ . \tag{2.9}$$

We denote the computational basis elements of the output state as $|abcd\rangle$, where a and d label the bottom and top qubits of the register, respectively, while $a, b, c, d = 0, 1$. We apply Rule 1 by tracing out the high spatial-frequency qubit a from $|\Phi\rangle_0$, i.e. by summing the probabilities of $|0bcd\rangle$ and $|1bcd\rangle$. Then, we apply Rule 2 by doing the same for the redundant qubit c, i.e. by summing the probabilities of $|b0d\rangle$ and $|b1d\rangle$. The output probabilities correspond to the matrix

$$\widetilde{\mathscr{I}} \simeq \begin{bmatrix} 0.35 & 0.06 \\ 0.32 & 0.27 \end{bmatrix} . \tag{2.10}$$

As expected, the output state closely resembles the input pattern, but downsampled and using fewer qubits than $|\Psi\rangle_0$. The result improves when adding the Hadamard gates. Following Algorithm 2.1, these gates right before the QFT_0 and after (Rule 2). Then, the output probabilities better reproduce the expected pattern, yielding

$$\widetilde{\mathscr{I}} \simeq \begin{bmatrix} 0.30 & 0.00 \\ 0.41 & 0.29 \end{bmatrix} . \tag{2.11}$$

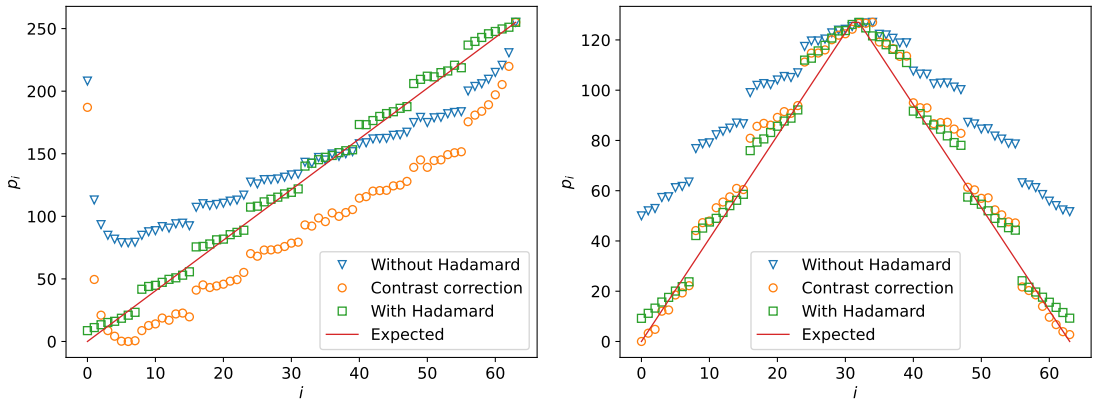
2.2.3 Example with discrete signals

Consider a monotonic linear pattern, represented by the array

$$[0, 1, \ldots, 255] , \tag{2.12}$$

normalized and encoded in an 8-qubit register, with computational basis

$$\{|0\rangle , |1\rangle , \ldots , |255\rangle\} . \tag{2.13}$$



(a) (b)

Figure 2.3: Downsampling of two one-dimensional patterns prepared as 8-qubit quantum states, for $\tilde{n} = 1$, with and without the Hadamard gates (with contrast correction in the latter case). Simulation performed with Qiskit Aer and 10^6 repetitions. The plot shows the probabilities of each computational basis element. (a) Monotonic linear pattern. (b) Triangular pattern. Figure adapted from [1].

A successful downsampling to a 6-qubit register would reproduce the same pattern, with less entries distributed across the computational basis $\{|0\rangle, \ldots, |63\rangle\}$. We downsample the pattern with and without using the Hadamard gates. Then, we perform the same task, but on a triangular pattern

$$[0, 1, \ldots, 126, 127, 127, 126, \ldots, 1, 0] , \tag{2.14}$$

normalized and encoded in an 8-qubit register, with computational basis

$$\{|0\rangle, \ldots, |255\rangle\} . \tag{2.15}$$

For a 6-qubit register, we expect the same pattern to distribute on the basis elements $\{|0\rangle, \ldots, |63\rangle\}$. The results are compared and reported in Fig. 2.3. Without the Hadamard gates, the output reproduces the original pattern, with lowered contrast. Moreover, asymmetrical boundary artifacts occur for the monotonic array, which increase with the number of discarded qubits, i.e. inversely with the size of $\mathrm{QFT}^{\dagger}{}_1$. For both patterns, the Hadamard improves the quality of the output: it regularizes the contrast, while removing the artifacts from the former. Both plots show discontinuities every 8 points, which can be removed by turning off Rule 2, which, However, remains essential to preserve the aspect ratio of a two-dimensional image.

2.2.4 Example with analytical results

We provide an analytical example, comparing the results with and without the Hadamard gates. Consider a 1×4 pattern, described by the one-dimensional array

$$\mathscr{I} = \begin{bmatrix} a & b & 0 & 0 \end{bmatrix} . \tag{2.16}$$

We encode $\mathscr{I}$ in the 2-qubit state $|\Psi\rangle_0 = a/K\,|00\rangle + b/K\,|01\rangle$, with $K^2 = a^2 + b^2$. Choose $\tilde{n} = 1$, so that $n_2 = 1$. In this case, $U^{\dagger}_{\mathrm{QFT}_1} = 1 \otimes H$. Without Hadamard gates, the downsampling algorithm reads

$$|\Phi\rangle_0 = (1 \otimes H) U_{\mathrm{QFT}_0} |\Psi\rangle_0 \ . \tag{2.17}$$

Explicitly

$$\begin{gathered} |\Phi\rangle_0 = \frac{1}{2K}\left[2a\,|{+}0\rangle + b(1+i)\,|{-}0\rangle + b(1-i)\,|{-}1\rangle\right] , \\ \text{with } |\pm\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \pm |1\rangle\right) \ . \end{gathered} \tag{2.18}$$

Discard the most significant qubit. The output probabilities correspond to the normalized 1×2 array

$$\widetilde{\mathscr{I}} \simeq \frac{1}{2K^2}\begin{bmatrix} 2a^2 + b^2 & b^2 \end{bmatrix} . \tag{2.19}$$

Such output closely resembles the input pattern, despite the artifact on the second entry (whose intensity depends on the contrast of the original array). We perform the same calculation, by applying the Hadamard gates before the QFT_0 and after (Rule 2). In this case, Eq. (2.17) becomes $|\Phi\rangle_0 = (H \otimes 1) U_{\mathrm{QFT}_0} (H \otimes H) |\Psi\rangle_0$. Namely

$$|\Phi\rangle_0 = \frac{1}{K\sqrt{2}}\left[(a+b)\,|00\rangle + (a-b)\,|10\rangle\right] . \tag{2.20}$$

Discard the most significant qubit. By substituting the normalization factor, the output probabilities yield

$$\widetilde{\mathscr{I}} \simeq \begin{bmatrix} 1 & 0 \end{bmatrix} . \tag{2.21}$$

This result perfectly matches the original pattern, now encoded in fewer pixels. A complete analytical discussion of Algorithm 2.1 can be found in [7].

2.3 Statistics and image retrieval

As anticipated in the previous example, the reconstruction of the output image requires the complete knowledge of the computational basis probabilities

$$p_j = \mathrm{Tr}\left(|j\rangle_2\langle j|\,\rho_2\right) , \tag{2.22}$$

which corresponds to the gray values $\tilde{g}_j$, with $j \in \{0, 1, \ldots, 2^{n_2} - 1\}$. The image is obtained through the devectorization operation θ^{-1}, which rearranges the gray values as the entries of a matrix of 2^{n_2} pixels. Characterizing the sample size for a good-quality output requires further statistical considerations. In the amplitude encoding, the gray values are probabilistically encoded in the output state. Hence, a single measurement would produce a single white pixel in the image plane, with position specified by the output bitstring: a complete reconstruction requires the full knowledge of the output probability distribution. Let n_p the number of photons subsequently collected at the output, i.e. the number of repetitions of the algorithm and the measurement, and L the number of gray levels at the output. For example,

$L = 256$ for 8-bit images. Let f_j be the frequency of each outcome, which approximates the theoretical values p_j. We identify the color *white* with the value of the pixel with highest probability

$$\begin{aligned} f_w &= \max_j f_j \in [0,1] \,, \\ &\text{with } f_j \in [0, f_w] \;\forall j \,. \end{aligned} \tag{2.23}$$

Using the output frequencies, we reconstruct the gray values as

$$\tilde{g}_j = f_j L / f_w \,, \tag{2.24}$$

where $\tilde{g}_j = 0, 1, \dots L-1$ and $\tilde{g}_w = L$. For each j, a measurement leads to two outcomes: the photon is assigned to the jth bin, i.e. to the jth pixel, with probability p_j, or it is not with probability $1-p_j$. Under the normal approximation and with the 95% confidence level [83], the output probabilities can be estimated as

$$p_j = f_j \pm 2\sqrt{f_j(1-f_j)/n_p} \,, \tag{2.25}$$

yielding

$$g_j = \tilde{g}_j \pm 2\sqrt{f_j(1-f_j)L^2/f_w^2 n_p} \,. \tag{2.26}$$

By requiring that the gray values fluctuate of at most one level, we get

$$n_p \geq 4 f_j (1-f_j) L^2 / f_w^2 \,. \tag{2.27}$$

Any realistic image with at least two non-black pixels has $f_w < 0.5$, that is $n_p \geq 4L^2/f_w$. We consider the worst-case scenario, i.e. a white image of N pixels. In this case $f_w = 1/N$ and each pixel uniformly collects an average of $4L^2$ shots, yielding

$$n_p \gtrsim 4L^2 N \,. \tag{2.28}$$

This bound provides a conservative estimation for any image. In practice, the optimal number of photons depends on the proportion between dark and bright pixels, which determines the value of f_w. An example of this is given by highly contrasted and mostly black images, for which photons will quickly localize in the most significant sectors of the pattern. Moreover, for a specific image, the reconstruction can be actively optimized by measuring the fluctuation value each time a new photon is collected, then stopping the experimental repetitions as soon as the desired standard deviation is reached by the chosen number of pixels.

2.4 Simulations

In this section, we evaluate the quantum downsampling algorithm against real digital images. Moreover, we investigate the effect of the Hadamard gates on the output reconstruction, from both a qualitative and a statistical point of view. In Fig. 2.4, we downsample a 512×512 image to different resolutions. We reconstruct the output using the procedure described in Section 2.3. The results are upscaled and compared

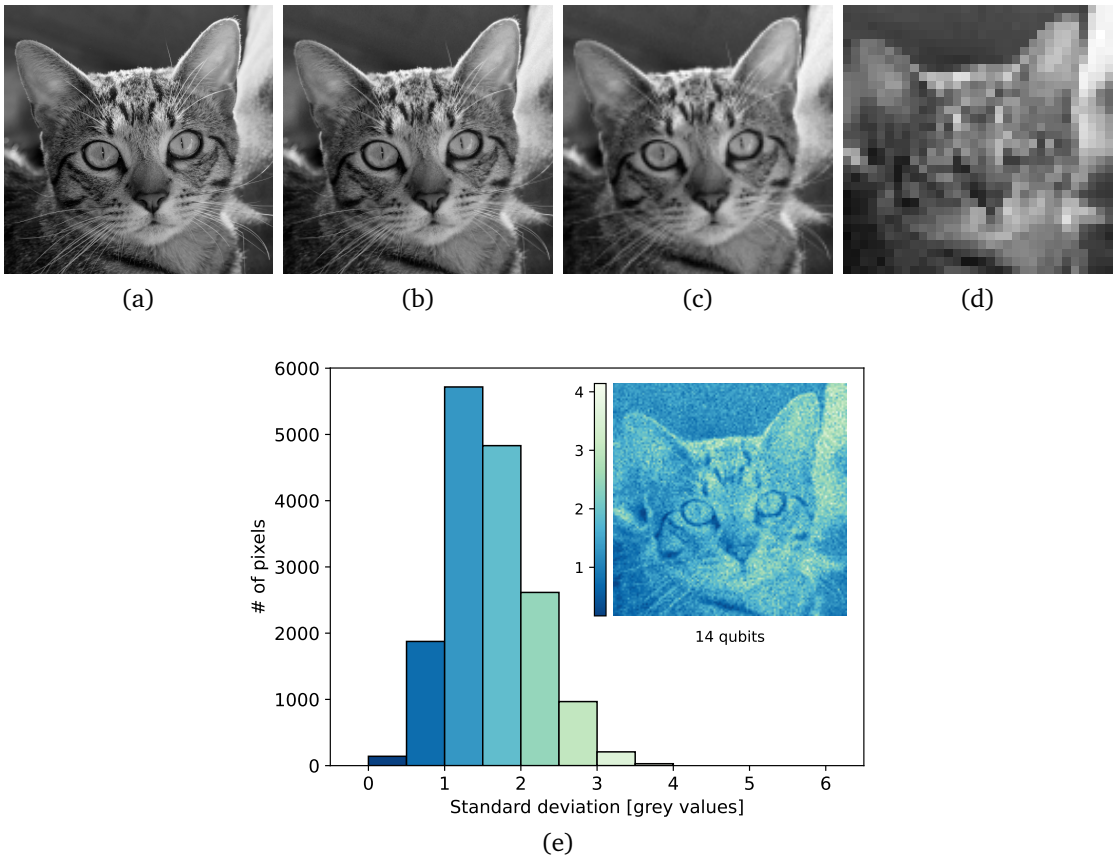


Figure 2.4: Downsampling of an 8-bit 512 × 512 digital image. (a) Input image encoded in an 18-qubit register. (b-d) Reconstructed images, for different number of n_2 qubits in the output register. Simulation performed with Qiskit Aer and $2^{n_2} \times 256^{3/2}$ repetitions. Aliasing starts to appear on cat's whiskers. (b) 16 qubits, 256 × 256 pixels. (c) 14 qubits, 128 × 128 pixels. (d) 10 qubits, 32 × 32 pixels. (e) Statistical fluctuations, plotted as the distribution of pixels with a certain standard deviation. The inset shows the standard deviation at each pixel location. Figure adapted from [1].

in Fig. 2.4. We consider the statistical fluctuations observed with $L^{3/2}N$ photons, estimated by repeating the reconstruction for 20 times. For a typical image, this order of magnitude provides the adequate statistics for a good-quality reconstruction. The Hadamard gates improve the reconstruction quality, as well as the overall contrast. This observation is further addressed in Fig. 2.5, where we show that, under the same number of photons, the Hadamard gates reduce the value of output fluctuations. Finally, in Fig. 2.8 we show several examples in which aliasing starts to appear, consistently with classical downsampling algorithms that operate below the Nyquist frequency.

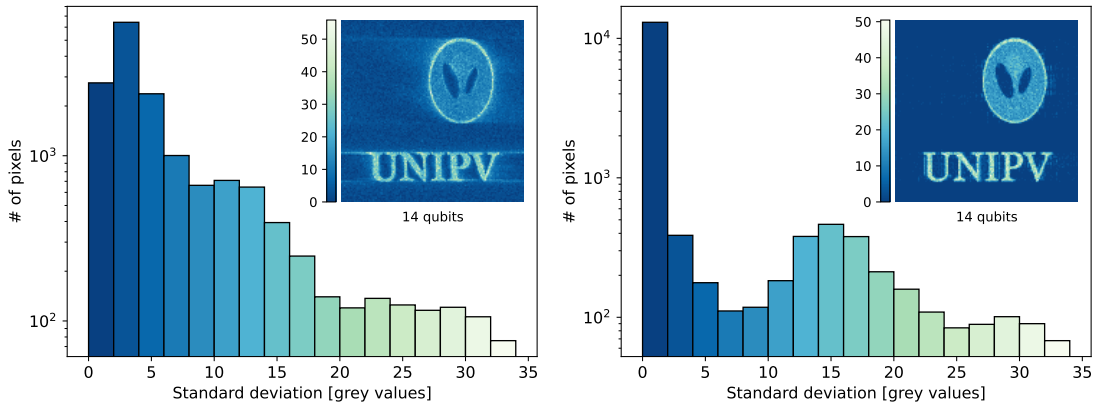


Figure 2.5: Characterization of the output statistics, without (a) and with (b) the Hadamard gates. The input image represents a Shepp-Logan phantom, commonly used in medical tomography [84], downsampled from 18 to 14 qubits. The simulation is performed with Qiskit Aer and 2^{14} repetitions. The choice of a sub-optimal sample size purposely highlights the effect of the statistical fluctuations. The histograms plot the number of pixels with respect to their standard deviation, obtained by repeating the reconstruction for 20 times. The vertical axis is in $\log_{10}$ scale. The inset displays the standard deviation at each pixel location. Figure adapted from [1]

2.5 Comparison with classical algorithms

We discuss the number of resources needed to downsample and retrieve a single image. In terms of number of gates, Algorithm 2.1 provides an exponential speedup over its classical counterpart. As discussed above, the output reconstruction requires multiple runs of it, increasing the overall cost of the protocol. We show that a quantum advantage remains, which increases when registers of different sizes are all downsampled to the same resolution. Consider an image of N pixels, encoded in n_0 qubits and downsampled to 2^{n_2} pixels, namely n_2 qubits. We have two situations. For $n_1 < n_0$, the complexity of QFT_0 dominates that of $\mathrm{QFT}^\dagger{}_1$. For $n_1 \simeq n_0$, both operations contribute up to the same order. In both cases, their composition can be upper bounded at $2n_0^2$ gates, or to $2n_0 \log_2(n_0)$ gates, by simply approximating the QFT [85]. As discussed in Section 2.3, a conservative retrieval requires $4L^2 2^{n_2}$ photons, yielding a total cost

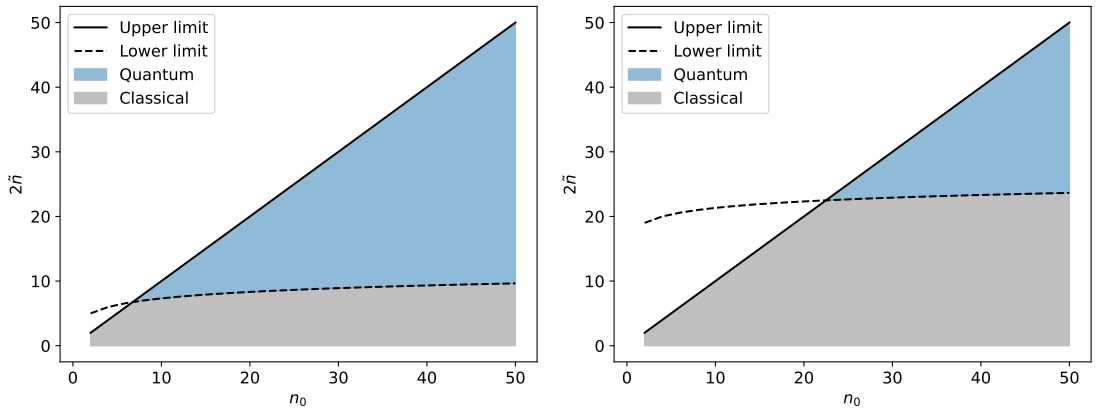
$$Q(n_0, n_2) = \mathcal{O}\left(8L^2 n_0^2 2^{n_2}\right) , \tag{2.29}$$

where $L = 2^c$ is the number of output gray levels. Notice the trade-off between the input and the output resolutions, which is essential to achieve the advantage. Classically, a single fast Fourier transform (FFT) requires at most $2N^2 \log_2 N$ operations [86], yielding a total cost

$$C(n_0) = \mathcal{O}\left(2n_0 2^{n_0}\right) . \tag{2.30}$$

Then, the relative cost in terms of operations and statistics is

$$Q(n_0, n_2)/C(n_0) = \mathcal{O}\left(n_0 2^{2+2c+n_2-n_0}\right) . \tag{2.31}$$



(a) (b)

Figure 2.6: Quantum advantage bounds, obtained by taking into account the image retrieval cost according to Eq. (2.32). The gray and coloured regions represent the classical or quantum advantage, respectively, with respect to the of number of input and discarded qubits. (a) Black-and-white image. (b) Grayscale 8-bit image.

An advantage holds whenever $Q(n_0, n_2) < C(n_0)$, namely

$$2(1+c) + \log_2(n_0) < 2\tilde{n} < n_0 \ . \tag{2.32}$$

For example, downsampling a 128×128 (14 qubits) black-and-white image to an 8×8 (6 qubits) resolution requires $\sim 12\%$ fewer operations than the FFT. In Fig. 2.6 we visualize the above inequality, for $c = 1$, i.e. a black-and-white image, and $c = 8$, i.e. an 8-bit image.

2.6 Quantum JPEG

Algorithm 2.1 can be adapted to limited-size NISQ devices [36]. Let b be the maximum number of logical qubits supported by a hypothetical hardware device and consider an image exceeding the encoding capability of the hardware, with $N^2 > 2^b$. The algorithm still works by first splitting the image in subimages of 2^b pixels, i.e. set of states $N/2^b$ states, to be processed sequentially or in parallel by Algorithm 2.1. A similar pre-encoding is also performed by the classical JPEG algorithm, which splits the original image in 8×8 blocks, before mapping them to the spatial frequencies domain [78]. Finally, we outline an alternative algorithm that compresses the image without downscaling its resolution, i.e. that gives $n_2 = n_0$. Consider $\mathrm{QFT}_0 |\Psi\rangle_0$. Instead of discarding the last $\tilde{n}$ qubits of the n_0-qubit register, we reinitialize them, e.g. through $\tilde{n}$ ancillas and SWAP operations. We take the $\mathrm{QFT}^\dagger{}_0$ and skip Rule 2. We report this algorithm in Fig. 2.7. In this case, output resolution remains unchanged, while the number of qubits that carries information is reduced from n_0 to $n_0 - \tilde{n}$. Although the Hadamard gates are not necessary, spatial artefacts appears for large $\tilde{n}$. This variant closely resembles the classical JPEG compression, in particular when combined with the previous subimages encoding.

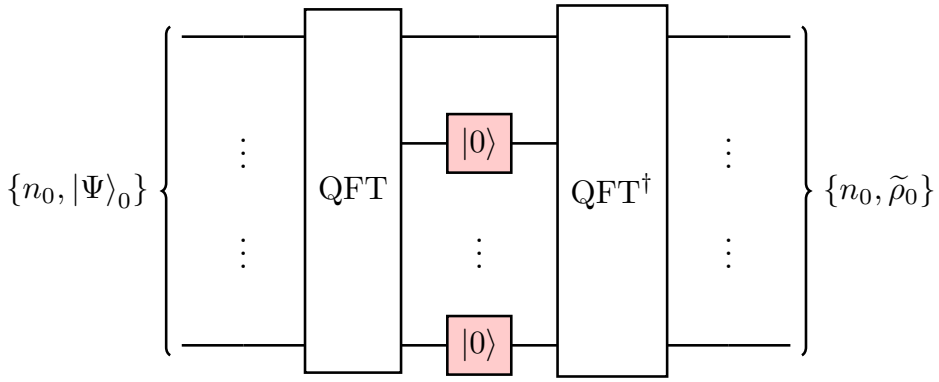
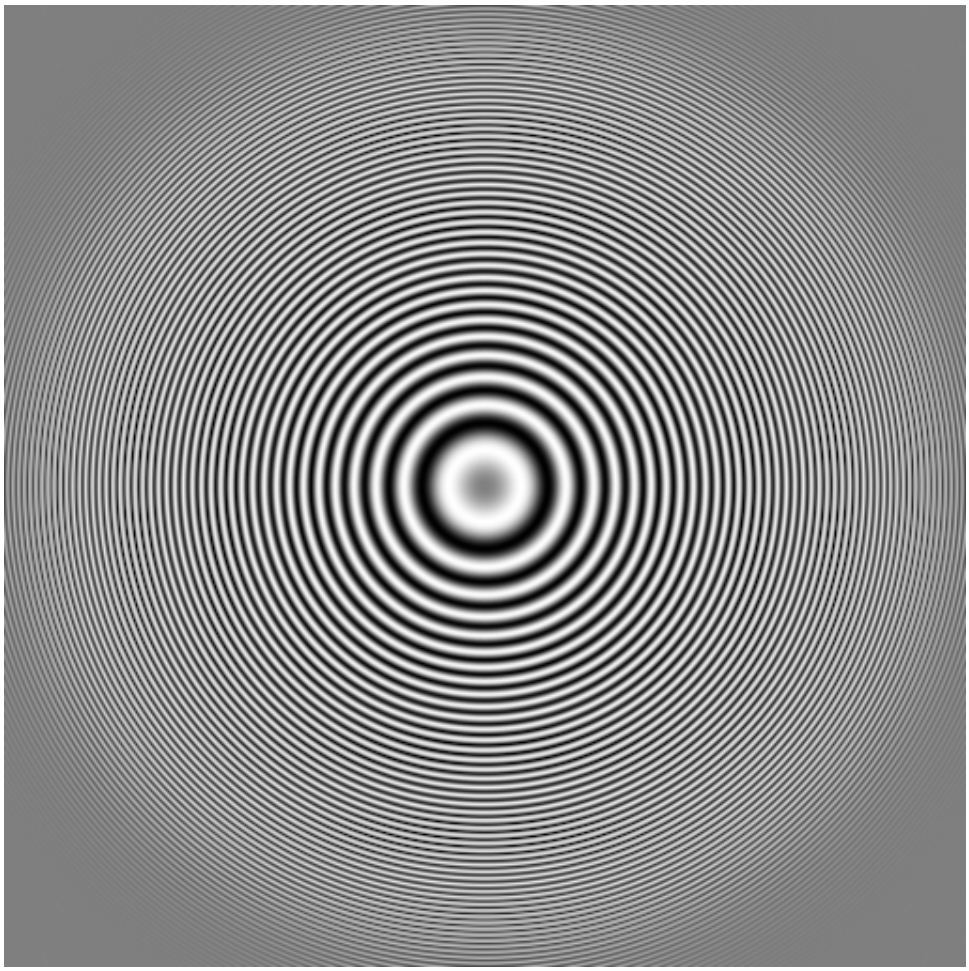


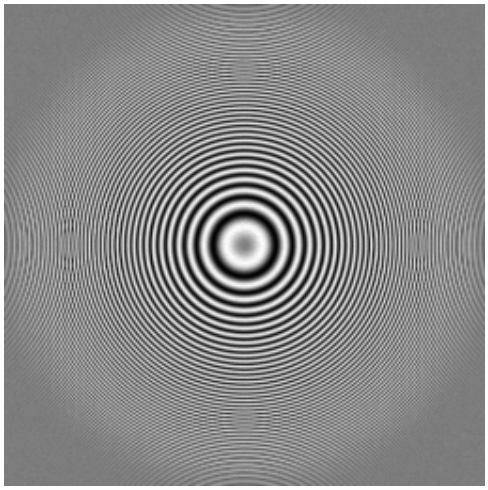
Figure 2.7: Quantum image compression by resetting the most significant $\tilde{n}$ qubits, instead of discarding them using Rule 1.

2.7 Discussion

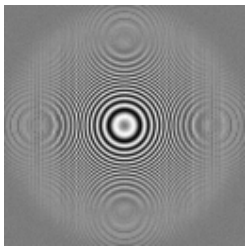
In this chapter, we discussed a quantum algorithm for image downsampling and compression. It leverages the QFT to downscale the original visual pattern, while reducing the number of encoding qubits. Our implementation differs from other examples in the literature [88–91], in both the encoding and the compression strategies. In particular, [88] leverages matrix product states truncation, while [89] and [90, 91] propose hybrid or interpolation-based algorithms, respectively. It also differs from [92], which focuses on image filtering rather than downsampling. In terms of number of gates only, it provides an exponential speedup over its classical counterpart and over interpolation-based quantum algorithms [90, 91]. We investigated the cost of the output retrieval, showing that the advantage increases consistently with the size of the input register. Whenever a full reconstruction is not needed, e.g. when used as subroutine rather than a standalone module, the theoretical advantage is completely recovered. For example, after capturing an image using the hardware encoder of the previous chapter, the downsampling algorithm can reduce the image storage requirements, e.g. the number of qubits of a quantum random access memory [93], as well as the dimension of the channel needed for communicating the image to other users [94]. Information theory, in particular rate-distortion theory, can be used to benchmark these tasks. We expect our results to satisfy Shannon's bounds. Indeed, Algorithm 2.1 is equivalent to a block-wise average [7], which is suboptimal compared to the Whittaker–Shannon formula and exhibits aliasing when sampled below the Nyquist frequency (see Fig. 2.8). Nevertheless, it incorporates unique features compared to classical interpolation algorithms, such as complementarity and probabilistic image retrieval. For this reason, such analysis represents an interesting challenge and a potential perspective of our results.



(a)



(b)



(c)

Figure 2.8: Aliasing test for the quantum downsampling algorithm, performed on 8-bit 512×512 digital image [87]. Output proportions are preserved. (a) Input encoded in 18 qubits. (b) 16 qubits, 256×256 pixels. (c) 14 qubits, 128×128 pixels. Aliasing is observed both in (b) and (c), where additional spatial frequencies mask the original one. For a faithful visualization use 100% PDF scale factor.

CHAPTER 3

QUANTUM OPTICAL CLASSIFIER

Ceci n'est pas une pipe.

René Magritte

IMAGE CLASSIFICATION has been significantly fostered by the development of deep learning algorithms, which provide several architectures to learn and extract image features [95–98]. However, such effectiveness comes with high computational costs, both for training and inference. The large number of parameters required by conventional neural networks is motivating a consistent effort in finding alternative implementations beyond the standard computational paradigm, e.g. by leveraging all-optical implementations that bypass hardware usage [99–105], or quantum mechanical effects that can provide a significant speedup in these computations [106–115]. Quantum optical neural networks harness the best of both worlds, i.e. deep learning capabilities from quantum optics [116–122].

Previously, we discussed an architecture to capture and compress a digital image into a quantum state, by leveraging a combined optical and computational quantum framework. In this chapter, we completely revisit our paradigm. We introduce a quantum optical setup to classify objects without reconstructing their images. Our approach uses the Hong-Ou-Mandel effect, for which the probability that two photons exit a beam splitter in different modes depends on their spectral distinguishability [123–126]. In our setup, an input object is targeted by a single-photon source, and eventually followed by an arbitrary lens system. The single-photon state interferes with another one, which encodes the target category in a set of trainable parameters. After the Hong-Ou-Mandel interferometer, the photons are collected by two bucket detectors with no spatial sensitivity, one for each output mode. The classification label is obtained by measuring the rate of two-photon coincidences at the output. Ideally, the classifier returns 1 when coincidences are not observed, i.e. when the object belongs to the same category of the reference, otherwise 0 when they are

orthogonal. After the measurement, the response function of our apparatus mathematically resembles that of a classical neuron. For this reason, we refer to our setup as a quantum optical neuron. By analytically comparing the resource cost of the classical and quantum neurons, we show that our method requires constant $\mathcal{O}(1)$ computational operations and injected photons, whereas the classical methods are at least linear in the image resolution: a superexponential advantage. Our discussion is completely general, and independent of the type of encoding. Additionally, we consider an explicit implementation in terms of amplitude-only and phase-only spatial light modulators.

This chapter is built upon the results of [2]. In collaboration with A. R. Morgillo, I have contributed to all the aspects of this research, including the theoretical analysis and the numerical simulations. The underlying code that generated the data for this study is openly available in GitHub [127].

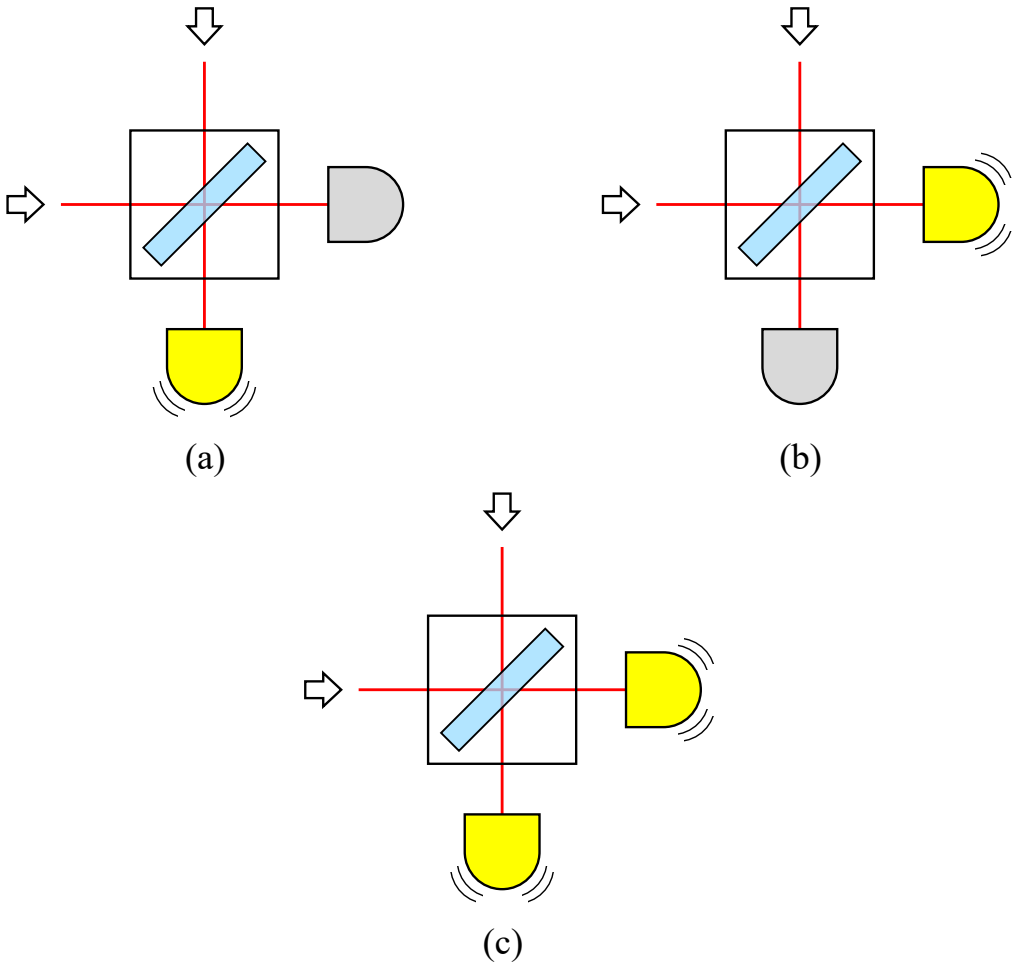


Figure 3.1: Possible outcomes when two photons exit a two-mode interferometer. (a-b) The photons exit the interferometer in the same mode. (c) The photons exit the interferometer in separate modes: a coincidence event. If the photons are indistinguishable, the Hong-Ou-Mandel effect predicts that coincidences never occur.

3.1 Preliminaries on the Hong-Ou-Mandel effect

In this section, we introduce the Hong-Ou-Mandel effect. Consider two photons fed inside a two-mode 50 : 50 beam splitter. Classically, the photons have 50% chance of exiting in separate modes, producing an event called coincidence. Quantum mechanics predicts that indistinguishable photons cannot produce coincidence events, as they must exit the interferometer always in the same mode. This is called the Hong-Ou-Mandel effect [123], and represents one the cornerstone of quantum interferometry. See Fig. 3.1 for a schematic representation. Consider two generic single-photon states, with modes a and b, each generated by a monochromatic source. Each mode reads

$$|\Psi_m\rangle = \int \mathrm{d}^2k\ \hat{\Psi}_m(k) m^\dagger(k)\,|0\rangle\ , \tag{3.1}$$

with mode $m \in \{a, b\}$. Consider the bipartite state

$$|\Psi_a\rangle \otimes |\Psi_b\rangle = \int \mathrm{d}^2k_1\,\mathrm{d}^2k_2\ \hat{\Psi}(k_1,k_2) a^\dagger(k_1) b^\dagger(k_2)\,|0\rangle\ , \tag{3.2}$$

with $\hat{\Psi}(k_1,k_2) = \Psi_a(k_1)\Psi_b(k_2)$. After feeding it into the beam splitter, the latter acts as a unitary transformation, which, in terms of the modes a an b [128], gives

$$a^\dagger \to \frac{1}{\sqrt{2}}\left(a^\dagger + b^\dagger\right)\ , \tag{3.3}$$

$$b^\dagger \to \frac{1}{\sqrt{2}}\left(a^\dagger - b^\dagger\right)\ . \tag{3.4}$$

At the output, the system state reads

$$|\Phi\rangle = \frac{1}{2}\int \mathrm{d}^2k_1\,\mathrm{d}^2k_2\ \hat{\Psi}(k_1,k_2)\left[a^\dagger(k_1) + b^\dagger(k_1)\right]\left[a^\dagger(k_2) - b^\dagger(k_2)\right]|0\rangle\ . \tag{3.5}$$

Consider two bucket detectors, i.e. with no spatial sensitivity, one for each output branch of the beam splitter. Detection of mode m is described by the projector

$$\Pi_m = \int \mathrm{d}^2k\ m^\dagger(k)\,|0\rangle\langle 0|\,m(k)\ . \tag{3.6}$$

The rate of coincidences, i.e. the probability that one and only one photon is detected in each mode, gives

$$p(1_a \cap 1_b) = \mathrm{Tr}[|\Phi\rangle\langle\Phi|\,\Pi_a \otimes \Pi_b]\ , \tag{3.7}$$

$$\text{with } \Pi_a \otimes \Pi_b = \int \mathrm{d}^2k_3\,\mathrm{d}^2k_4\ a^\dagger(k_3) b^\dagger(k_4)\,|0\rangle\langle 0|\,a(k_3) b(k_4)\ . \tag{3.8}$$

By substitution of Eq. (3.5), we obtain

$$p(1_a \cap 1_b) = \frac{1}{4}\int \prod_{i=1}^{6} \mathrm{d}^2k_i\ \hat{\Psi}(k_1,k_2)\hat{\Psi}^*(k_5,k_6) W(k_1,k_2,\ldots,k_6)\ , \tag{3.9}$$

$$W(k_1,k_2,\ldots,k_6) = W_1(k_1,k_2,k_3,k_4) W_2(k_3,k_4,k_5,k_6) \tag{3.10}$$

The first factor reads

$$\begin{aligned} W_1(k_1,k_2,k_3,k_4) = \langle 0|\, a(k_3)b(k_4)\big[a^\dagger(k_1)a^\dagger(k_2) \\ -a^\dagger(k_1)b^\dagger(k_2)+b^\dagger(k_1)a^\dagger(k_2)-b^\dagger(k_1)b^\dagger(k_2)\big]|0\rangle \end{aligned} . \tag{3.11}$$

By simplifying all the Wick products [59], it simplifies to

$$W_1(k_1,k_2,k_3,k_4) = \delta(k_2-k_3)\delta(k_1-k_4)-\delta(k_1-k_3)\delta(k_2-k_4) . \tag{3.12}$$

Similarly

$$\begin{aligned} W_2(k_3,k_4,k_5,k_6) = \langle 0|\,[a(k_6)a(k_5)-b(k_6)a(k_5) \\ +a(k_6)b(k_5)-b(k_6)b(k_5)]\,a^\dagger(k_3)b^\dagger(k_4)\,|0\rangle \end{aligned} , \tag{3.13}$$

which becomes

$$W_2(k_3,k_4,k_5,k_6) = \delta(k_3-k_6)\delta(k_4-k_5)-\delta(k_3-k_5)\delta(k_4-k_6) . \tag{3.14}$$

By integrating out the Dirac deltas of Eq. (3.9), we obtain

$$p(1_a \cap 1_b) = \frac{1}{2}\int \mathrm{d}^2k_1\,\mathrm{d}^2k_2\,\mathrm{d}^2k_5\,\mathrm{d}^2k_6\;\hat{\Psi}(k_1,k_2)\hat{\Psi}^*(k_5,k_6)\widetilde{W}(k_1,k_2,k_5,k_6) , \tag{3.15}$$

$$\widetilde{W}(k_1,k_2,k_5,k_6) = \delta(k_1-k_5)\delta(k_2-k_6)-\delta(k_1-k_6)\delta(k_2-k_5) . \tag{3.16}$$

After splitting the two contributions, the rate of coincidences yields

$$p(1_a \cap 1_b) = \frac{1}{2}\int \mathrm{d}^2k\;|\hat{\Psi}_a(k)|^2 \int \mathrm{d}^2k\;|\hat{\Psi}_b(k)|^2 - \frac{1}{2}\left|\int \mathrm{d}^2k\;\hat{\Psi}_a(k)\hat{\Psi}_b^*(k)\right|^2 . \tag{3.17}$$

Using the Parseval's and Plancherel's theorems, this equation can be compactly written as

$$p(1_a \cap 1_b) = \frac{1}{2}\left[||\Psi_a||^2||\Psi_b||^2 - |\langle\Psi_a|\Psi_b\rangle|^2\right] , \tag{3.18}$$

with $||\cdot||$ and $\langle\cdot,\cdot\rangle$ denoting the norm and inner product in L^2, i.e. the space of square-integrable functions. When normalization is guaranteed, e.g. when optical losses are negligible, this formula further simplifies to

$$p(1_a \cap 1_b) = \frac{1}{2}\left[1 - |\langle\Psi_a|\Psi_b\rangle|^2\right] . \tag{3.19}$$

When $\langle\Psi_a|\Psi_b\rangle = 1$, i.e. when the two modes are indistinguishable, the rate of coincidence vanishes, and the Hong-Ou-Mandel effect is observed. Conversely, orthogonal modes, i.e. $\langle\Psi_a|\Psi_b\rangle = 0$, leads to the outcomes reported in Fig. 3.1 with $p(1_a \cap 1_b) = 1/2$.

3.2 Quantum optical neuron

Using the Hong-Ou-Mandel effect, we design an interferometric scheme whose response function resembles that of a classical artificial neuron. Consider two single-photon states, generated by two independent monochromatic sources, fed into the left and top branches of a balanced beam splitter. We describe the branches in terms of two modes, i.e. a and b, respectively called input and probe modes. These are going to represent the neuron input and the processing layers, respectively. In the left branch, the single-photon with spectrum ϕ is generated at the longitudinal origin $z = 0$. Then, it reflects off an object with two-dimensional shape $\mathcal{O}$ and reaches the beam splitter after an optical system with transfer function $\mathcal{L}_d$, e.g a pinhole or a linear imaging apparatus. Here, we call z_o and z_i the longitudinal positions of the object and the image plane, respectively, and $d = z_i - z_o$ their displacement. We do not specialize to any particular design of the lenses. In the top branch, the single-photon state is injected through a generic device or system, encoding a set of trainable real parameters. We place a bucket detector at the interferometer output, counting the number exiting each mode without reconstructing their spatial distribution. Focusing on coincidence events, we discuss how to relate the output to that of an artificial neuron. An example of an optical scheme is reported in Fig. 3.2, where a spatial light modulator (SLM) encodes the trainable parameters in the top branch. Dealing with monochromatic light is equivalent to classifying grayscale images. Colors can be introduced by replacing the source with a multiwavelength one, or by considering multiple sets of probe parameters that account for the color decomposition of the image.

We mathematically discuss this setup. Unless differently specified, we adopt the dot symbol to denote an unused variable, e.g. $g(\cdot, y)$. The output of the left branch through the imaging optics reads

$$|\Psi_{\mathcal{I}}\rangle = \int \mathrm{d}^2k\, \hat{\mathcal{I}}_\omega(k|\mathcal{O}) a^\dagger_\omega(k)\,|0\rangle\ , \tag{3.20}$$

where

$$\hat{\mathcal{I}}_\omega(\cdot|\mathcal{O}) = [(\hat{\phi}_\omega \hat{\mathfrak{H}}_{z_o}) * \hat{\mathcal{O}}]\hat{\mathcal{L}}_d \tag{3.21}$$

is the total transfer function from the single-photon source to the image plane. Here $a^\dagger_\omega(k)$ is the creation operator of a photon in the input mode. The hat operator denotes the two-dimensional Fourier transform on the transverse coordinates plane, $*$ the convolution operation, $\mathfrak{H}_{z_o}$ the transfer function from the source to the object plane, $k = (k_x, k_y)$ the transverse momentum, and ω the frequency conjugated to the temporal degree of freedom of the electromagnetic potential. In the probe branch, the parametrized quantum state is prepared. At the beam splitter plane, the probe state reads

$$|\Psi_{\mathcal{U}}\rangle = \int \mathrm{d}^2k\, \hat{\mathcal{U}}_\omega(k|\lambda) b^\dagger_\omega(k)\,|0\rangle\ , \tag{3.22}$$

with $\lambda = \{\lambda_{i_1 \dots i_n}\}$ a collection of (trainable) parameters, $\mathcal{U}$ the spatial spectrum of the probe, and $b^\dagger_\omega(k)$ the creation operator of a photon in this mode. Both photons are

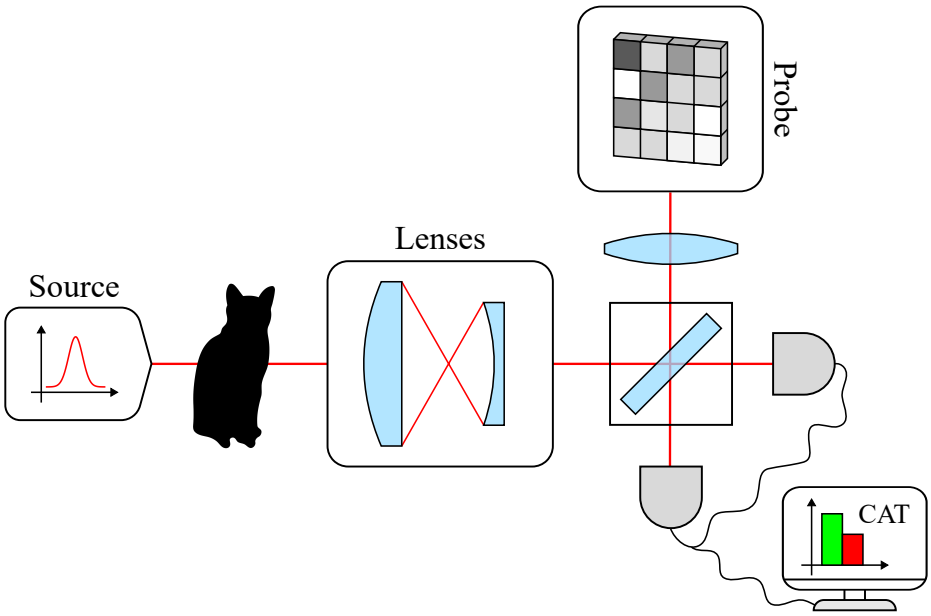


Figure 3.2: Quantum optical neuron implemented by the Hong-Ou-Mandel interferometric setup. An object is targeted by a single-photon source and classified through the rate of two-photon coincidences at the interferometer output, obtained by placing two photodetectors with no spatial resolution, i.e. without reconstructing the object image. In the top branch, an additional thin lens can translate the classification problem to the Fourier domain. Figure adapted from [2].

fed through the 50:50 beam splitter. The bucket photodetector, i.e. a photon counter with no spatial resolution, is placed at the output of each branch. Following the same discussion of the previous section, the rate of two-photon coincidences reads

$$p(1_a \cap 1_b|\lambda,\mathcal{O}) = \frac{1}{2}\left[\alpha_\lambda(\mathcal{O}) - f_\lambda(\mathcal{O})\right] , \tag{3.23}$$

$$\text{with} \quad \begin{aligned} \alpha_\lambda(\mathcal{O}) &= ||\mathcal{I}_\omega(\cdot|\mathcal{O})||^2||\mathcal{U}_\omega(\cdot|\lambda)||^2 , \\ f_\lambda(\mathcal{O}) &= |\langle \mathcal{I}_\omega(\cdot|\mathcal{O}), \mathcal{U}_\omega(\cdot|\lambda)\rangle|^2 , \end{aligned} \tag{3.24}$$

Here, $\alpha_\lambda(\mathcal{O})$ depends on the normalizability of the input and probe states. In the presence of optical losses, it can be $\alpha_\lambda < 1$. Whenever the two spectra are indistinguishable, i.e. when $\mathcal{U}$ perfectly matches $\mathcal{I}$, coincidences are not observed. On the other hand, the more distinguishable the input and the probe states are, the smaller their inner product becomes, and the rate of coincidences increases. In the next sections, we report how to obtain similar results in the Fourier domain. At the image plane I, with transverse coordinates $r = (x, y)$, Eq. (3.23) gives

$$f_\lambda(\mathcal{O}) = \left| \int_I \mathrm{d}^2 r \; \mathcal{I}_\omega(r|\mathcal{O}) \mathcal{U}^*_\omega(r|\lambda) \right|^2 . \tag{3.25}$$

This integral measures the point-wise overlap between the input image and the probe. We interpret it as the prediction of our classification model, where $f_\lambda \in [0,1]$ represents the probability that $\mathcal{I}$ belongs to the class of $\mathcal{U}$. In particular, $f_\lambda \to 0$ ($f_\lambda \to 1$)

when the class of $\mathcal{I}$ is orthogonal to (is the same of) $\mathcal{U}$. In the next section, we show how to encode a generic class in $\mathcal{U}$, by means of the optimization of the set of parameters λ. In the Appendix, we discuss how to obtain a similar response in the Fourier domain, by placing a lens in probe branch of the interferometer.

The output measurement produces a non-linear operation after the beam splitter, i.e. the squared absolute value of the scalar product between the input and the probe states. Usually, this is sufficient to obtain a predictive model [109, 110]. We improve the performance and enhance the predictability of our model, by increasing this non-linearity through the following post-processing operations. Consider the sigmoid (logistic) function

$$\sigma(x) := \frac{1}{1+\exp(-\beta x+\gamma)} , \tag{3.26}$$

where β, γ are hyperparameters, i.e. constants with respect to the training process. We introduce an additional trainable parameter $b \in \mathbb{R}$, called bias, which, combined with f_λ and σ, yields the response

$$F_{b\lambda}(\mathcal{O}) = \sigma(f_\lambda(\mathcal{O}) + b) . \tag{3.27}$$

We consider this as the post-processed label, predicted by the Hong-Ou-Mandel apparatus. The sigmoid increases the non-linearity introduced by the squared absolute value, and so the predictability of the model. In addition, the bias is introduced on heuristic motivations: it compensates the constraint introduced by the normalization, while enhancing the robustness of our protocol against optical losses (which may affect the above-mentioned normalizability, yielding $\alpha_\lambda < 1$).

3.3 Training

We discuss how to train the Hong-Ou-Mandel interferometer as a binary classifier. After subsequently feeding an ensemble of objects with known labels, we optimize the probe parameters to predict the target classes. Consider a training set, i.e. an ensemble of M objects and target labels

$$\left\{\left(\mathcal{O}_j, y_j\right) \mid 0 \le j \le M-1\right\} , \tag{3.28}$$

with $y_j \in \{0,1\}\ \forall j$. By feeding $\mathcal{O}_j$ into the input branch, we get

$$f_\lambda^{(j)} := f_\lambda(\mathcal{O}_j) = C - 2p(1_a \cap 1_b|\lambda, \mathcal{O}_j) , \tag{3.29}$$

$$F_{b\lambda}^{(j)} := F_{b\lambda}(\mathcal{O}_j) = \sigma(f_\lambda(\mathcal{O}_j)_\lambda + b) , \tag{3.30}$$

where $p \in [0, 1/2]$. For simplicity, we assumed that the optical losses are independent on both the input and the probe, that is $C := \alpha_\lambda(\mathcal{O}_j)\ \forall \lambda, j$. We define the binary cross-entropy between each target label and the predicted one as

$$H\left(y_j, F_{b\lambda}^{(j)}\right) = -y_j \log\left(F_{b\lambda}^{(j)}\right) - \left(1-y_j\right)\log\left(1-F_{b\lambda}^{(j)}\right) . \tag{3.31}$$

We train the classifier by finding the set of parameters that minimize this entropy, across the whole dataset. We use the gradient descent algorithm to optimally approximate the solution, where Eq. (3.31), averaged on the training set, is used as loss function. After each epoch, i.e. each complete pass through the dataset, the parameters are updated as

$$\lambda \to \lambda - \frac{\eta_\lambda}{M} \sum_j \partial_\lambda H\left(y_j, F_{b\lambda}^{(j)}\right) , \tag{3.32}$$

$$b \to b - \frac{\eta_b}{M} \sum_j \partial_b H\left(y_j, F_{b\lambda}^{(j)}\right) , \tag{3.33}$$

where η_λ, η_b are the learning rates of the probe and bias, i.e. the step size while moving toward the minimum of the loss function. The derivatives with respect to the parameters and the bias yield

$$\partial_\lambda H = (\partial_F H)\left(\partial_\xi \sigma\right)\partial_\lambda f , \tag{3.34}$$

$$\partial_b H = (\partial_F H)\,\partial_\xi \sigma , \tag{3.35}$$

with $\xi_{b\lambda} = f_\lambda + b$. By substitution of Eqs. (3.30) and (3.31), it reads

$$\partial_F H = \frac{F - y}{F(1-F)} , \tag{3.36}$$

$$\partial_\xi \sigma = \beta F(1-F) , \tag{3.37}$$

where β is the hyperparameter of the sigmoid. These contributions take into account the optimization strategy, i.e. the loss function, and the post-processing operations, i.e. the bias and the sigmoid. However, they are independent of the Hong-Ou-Mandel response and of the encoding type. The only model-dependent contribution is $\partial_\lambda f$, which can be computed as follows. For any complex function of real variable $h : \mathbb{R} \to \mathbb{C}$, it holds that

$$\partial_\lambda |h(\lambda)| = \frac{\mathrm{Re}\left[h(\lambda)(\partial_\lambda h(\lambda))^*\right]}{|h(\lambda)|} . \tag{3.38}$$

Hence,

$$\partial_\lambda f = 2\,\mathrm{Re}\left[\langle \mathcal{I}_\omega, \mathcal{U}_\omega\rangle \langle \mathcal{I}_\omega, \partial_\lambda \mathcal{U}_\omega\rangle^*\right] . \tag{3.39}$$

We can simplify this formula, by explicitly writing it in terms of the Hong-Ou-Mandel output. Indeed, neglecting the global phase of $\langle \mathcal{I}, \mathcal{U}\rangle$ yields

$$\partial_\lambda f \simeq 2\sqrt{f}\,\mathrm{Re}\left[\langle \mathcal{I}_\omega, \partial_\lambda \mathcal{U}_\omega\rangle\right] . \tag{3.40}$$

In our simulations, we verify this assumption under a self-consistency test. When the computation of the loss function derivative is not possible, we can replace the gradient descent method with the coordinate descent one. For each epoch, the loss function is evaluated on a discrete neighbourhood Λ in the parameters space. The update rule is

$$\lambda \to \arg\min_{\lambda \in \Lambda} H(\lambda) . \tag{3.41}$$

Notice that our model is resilient against the issue of gradient explosion [129], since it depends on physical data and functions only.

3.4 Comparison with a classical artificial neuron

Deep learning algorithms are characterized by high flexibility. They provide several trainable architectures that can solve different tasks, such as regression and classification. Examples include feedforward neural networks, convolutional neural networks, autoencoders and vision transformers. Originally introduced in the fifties as perceptron [130], the artificial neuron represents the fundamental unit of such algorithms. In this model, data are processed through a set of weighted connections, computing the scalar product between the input and a vector of trainable weights. The output is further transformed by adding a bias and an activation function, which is usually non-linear [131]. Mathematically, this process is described as

$$G_{bw}(x) = A(w \cdot x + b) \,, \tag{3.42}$$

with $x \in \mathbb{R}^N$ denoting the input data, while $w \in \mathbb{R}^N$ and $b \in \mathbb{R}$ the weights and bias, respectively. Here, A is the activation function. Examples include the linear function, the rectified linear unit, the hyperbolic tangent and the sigmoid functions. The latter takes values in the interval $[0,1]$, making it particularly suited for classifications tasks. Multiple neurons can be combined and stacked in layers, forming a neural network. For example, consider a 3-layer neural network, which takes an input $x \in \mathbb{R}^N$ (input layer), processes it through M neuron (hidden layer), and computes a scalar value (output layer). Mathematically, this reads

$$\begin{aligned} &G_{bwBW}(x) = A(w \cdot y + b) \,, \\ &\text{with } y_i = A\big(W_{ij} x_j + B_i\big) \,. \end{aligned} \tag{3.43}$$

In the last step, we employed the Einstein's summation notation. Here, $W \in \mathbb{R}^{M \times N}$, while w, $B \in \mathbb{R}^M$ and $b \in \mathbb{R}$. This network possesses universal approximation capabilities: it can be trained to simulate any real-valued function [132], i.e. it can classify an arbitrary complex pattern. We focus on the classical artificial neuron as the fundamental building block of such classifiers.

There is a formal relationship between the post-processed output of the Hong-Ou-Mandel interferometer and the operations and response of a classical artificial neuron. Consider the output $f_\lambda(\mathcal{O})$, discretized and vectorized in a mesh of N cells, either in the spatial or in the Fourier domain. Then, Eq. (3.27) corresponds to the composition of a (1) real-valued neuron, with N trainable weights, square absolute value activation function and no bias, and (2) another neuron, with a scalar unit weight, sigmoid activation function and a trainable bias. This reads

$$G_{bw}(x) = \sigma\big(|w \cdot x|^2 + b\big) \,, \tag{3.44}$$

where $x \in \mathbb{C}^N$ encodes the input data, while $w \in \mathbb{C}^N$ and $b \in \mathbb{R}$ are the weights and bias, respectively. We can formally identify $G_{bw}(x)$ with $F_{b\lambda}(\mathcal{O})$ under the substitution

$$(x, w) \overset{\sim}{\leftarrow} \big(\mathcal{I}_\omega(r|\mathcal{O}), \mathcal{U}_\omega(r|\lambda)\big) \,, \tag{3.45}$$

where $\overset{\sim}{\leftarrow}$ is the discretization and vectorization to $\mathbb{C}^N$. This analogy is schematically represented in Fig. 3.3. Indeed, a single artificial neuron represents the classical

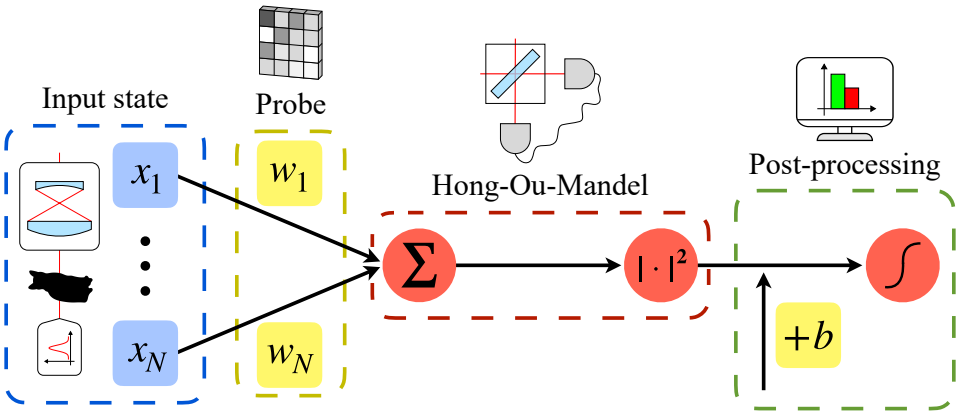


Figure 3.3: Mathematical relationship between the Hong-Ou-Mandel apparatus of Fig. 3.2 and a classical artificial neuron of Eq. (3.44). Each operation is identified with the corresponding component of the optical interferometer. Figure adapted from [2].

counterpart of our apparatus. We discuss the optical and computational advantage as the number of photons and operations required by a single image classification. Assuming that all the parameters have been previously trained with optimal accuracy, we show that our protocol requires a constant number of resources, i.e. $\mathcal{O}(1)$ complexity, independently of the input image resolution: it provides a superexponential speedup over its classical counterpart. We consider two different figures of merit: the number of mathematical operations (sums and products), which represent computational resources spent a CPU, and the number of photons injected, which describe the optical cost of the imaging or interferometric apparatus. From now on, we denote Ω, Θ and $\mathcal{O}$, respectively the lower, tight and upper bounds on the number of resources needed by a certain (optical or computational) operation.

Concerning the computational cost, we consider the best classical scenario, in which the pattern recognition problem is solved by a single artificial neuron. Consider a digital image x of N pixels, fed into a neuron

$$G_{bw}(x) = \sigma(w \cdot x + b)\,, \tag{3.46}$$

where $x, w \in \mathbb{R}^N$, $b \in \mathbb{R}$ and σ is the sigmoid activation, with hyperparameters $\beta = 1$ and $\gamma = 0$. Classically, an artificial neuron costs N operations to compute $w \cdot x$. The Hong-Ou-Mandel interferometer performs the same task in an all-optical way. Indeed, the inner product and the square absolute value activation are inherently obtained through the coincidence measurement at the output of the beam splitter. In post-processing, only the sigmoid activation function and bias requires computational resources. However, since they operate on a scalar number, i.e. the square absolute value of an inner product, they only require $\mathcal{O}(1)$ resources.

Optically, we discuss the advantage in using coincidences to classify single-photon states, instead of a neuron that processes fully reconstructed images. First, we investigate the cost of imaging by classical means. Then, we take into account the artificial neuron contribution. After targeting an object with light, a digital image x is a collection of gray levels obtained by counting the number of photons detected on different

pixels of a sensor grid, e.g. a charge-coupled device [47]. Let n_p be the average number of photons in the input state, and μ_i the average number of photons collected by the i-th pixel of the grid, with $i \in \{0,\dots,N\}$. Assuming perfect quantum efficiency and sufficiently low exposure times to neglect the saturation of the sensor, the gray values at each pixel read

$$x_i = \frac{\mu_i L}{\mu_w} \,, \tag{3.47}$$

with L the number of gray levels, i.e. the depth of the image, and $\mu_w = \max_i \mu_i$ the maximum number of photons collected in a single pixel. Indeed, $x_i \in \{0,1,\dots,L-1\}$ with 0 and L labelling the black and white colors, respectively. Each pixel has variance

$$\varsigma_i^2 = \frac{\Delta\mu_i^2 L^2}{\mu_w^2} \,, \tag{3.48}$$

with $\Delta\mu_i^2$ the variance on the number of collected photons. For coherent light, the photo-detection process undergoes the standard quantum limit (SQL) [133, 134], with Poissonian fluctuations that satisfy $\Delta\mu_i^2 \simeq \mu_i$. By averaging the variance across the image, we get

$$\varsigma^2 := \frac{1}{N}\sum_i \varsigma_i^2 \overset{\text{SQL}}{\simeq} \langle x\rangle^2 N n_p^{-1} \,, \tag{3.49}$$

$$\text{with } \langle x\rangle = \frac{1}{N}\sum_i x_i \,. \tag{3.50}$$

Here, $\langle x\rangle \in [0, L-1]$ is the average brightness of the image, which we assume to be independent of its resolution. Hence, the number of photons n_p required by a full image reconstruction with average variance ς^2 is

$$n_p = \varsigma^{-2}\langle x\rangle^2 N \,. \tag{3.51}$$

In terms of the signal-to-noise ratio $\eta = \langle x\rangle/\varsigma$, the number of photons needed simplifies to $n_p = \eta N$. This discussion involves the cost of image reconstruction only. We now consider the information propagation through the artificial neuron.

Proposition 1. Consider a neuron with sigmoid activation function. Suppose that there exists a sequence of parameters $\left\{(w_N, b_N) \in \mathbb{R}^{N+1}\right\}_{N\gg 1}$ that optimally solve the N-pixel image classification task, with b_N and the ℓ^1-norm $||w_N||_1$ asymptotically bounded for $N \to \infty$. Then, the number of photons n_p required to classify an image x with uncertainty ε, is

$$n_p = \Omega\left(\varepsilon^{-2}\langle x\rangle N\right) \,. \tag{3.52}$$

Proof. Consider the output of the neuron

$$G_{bw}(x) = \sigma(w_N \cdot x + b_N) \,, \tag{3.53}$$

and its derivative

$$\partial G(x) = G_{bw}(x)(1 - G_{bw}(x)) \,. \tag{3.54}$$

By neglecting the spatial neighbourhood correlations, which introduces at most a constant overhead in our estimation, we propagate the uncertainty of x as

$$\varepsilon^2 = \langle x \rangle (\partial G)^2(x) \sum_i (w_N)_i^2 x_i N \tilde{n}_p^{-1} \,, \tag{3.55}$$

where $\tilde{n}_p = n_r n_p$, with n_r is the number of independent image acquisition and classification. Here, black pixels do not contribute to this summation, then

$$\sum_i (w_N)_i^2 x_i \geq \sum_{i \notin \mathcal{B}} (w_N)_i^2 = ||w_N||^2 - \sum_{i \in \mathcal{B}} (w_N)_i^2 \,, \tag{3.56}$$

$$\text{where } \mathcal{B} = \{ i \in \mathbb{N} \mid x_i = 0 \text{ for } 0 \leq i \leq N-1 \} \tag{3.57}$$

represents the set of black pixels labels only. Moreover

$$||w_N||^2 > \sum_{i \in \mathcal{B}} (w_N)_i^2 \,. \tag{3.58}$$

By contradiction, $||w_N||^2 \simeq \sum_{i \in \mathcal{B}} (w_N)_i^2$ would imply either that the image is mostly black, independently of its resolution, or that any non-black pixel undergoes $(w_N)_i \simeq 0$. Both conditions prevent the learnability of the neuron. Therefore, by substitution of Eq. (3.58), Eq. (3.55) gives

$$\tilde{n}_p > \varepsilon^{-2} \langle x \rangle (\partial G)^2(x) ||w_N||^2 N \,. \tag{3.59}$$

By hypothesis, w_N is a sequence of non-trivial solutions of the classification problem. Then, its ℓ^2-norm $||w_N||^2$ cannot go to zero for $N \to \infty$. Additionally, neither $(\partial G)^2$ can converge to 0 for $N \to \infty$. Indeed,

$$(\partial G)^2(x) = \frac{\exp[-2(w_N \cdot x + b_N)]}{\{1 + \exp[-(w_N \cdot x + b_N)]\}^4} \,. \tag{3.60}$$

Since b_N is asymptotically limited, $(\partial G)^2$ converges to zero if and only if $w_N \cdot x \to \pm\infty$. Since $x_i \in [0, L-1]$, the Cauchy–Schwarz inequality combined with the Hölder's one implies that

$$|w_N \cdot x| < L ||w_N||_1 \,. \tag{3.61}$$

Since the ℓ^1-norm is limited, $(\partial G)^2$ admits strictly positive lower bound for $N \to \infty$. Finally, this gives

$$\tilde{n}_p = \Omega\left(\varepsilon^{-2} \langle x \rangle N\right) \,. \tag{3.62}$$

For a single image acquisition, i.e. for $n_r = 1$, the results read as stated above. □

In summary, two conditions lead to the above lower bound. On the one hand, that $||w_N||^2 \not\to 0$ for $N \to \infty$. This is essential to guarantee that the neuron can be trained at any resolution, i.e. that there always exists a solution to the optimization problem. On the other hand, that $||w_N||_1$ is bounded for $N \to \infty$, which is compatible with LASSO and Tikhonov's regularization techniques [135, 136]. Our method reduces this cost by substituting the imaging system with two bucket detectors, with no spatial resolution. The classifier outcome depends on the estimation of the rate

of coincidences of the Hong-Ou-Mandel interferometer. Let $\tilde{n}_p = 2n_p$ be the number input photons, and $\tilde{p} \in [0, 1/2]$ the empirical rate of coincidences. Under the normal approximation, with the 95% confidence level [83], the estimation uncertainty reads

$$\varepsilon = 2\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}_p}} \,. \tag{3.63}$$

Since $4\tilde{p}(1-\tilde{p}) \leq 1$, the total number of photons is $\mathcal{O}(\varepsilon^{-2})$. This scaling is constant, independent of the resolution of the image. In conclusion, the quantum optical neuron provides a superexponential advantage over its classical counterpart, both in the number of operations and photons saved to classify a single image. In both cases, it requires constant computational and optical resources to perform the classification, independently on the input resolution. We summarize these results and the full discussion in Table 3.1.

The above estimations do not consider the cost of training, which is present in both classical and quantum models. Depending on its implementation, it may contribute to the protocol complexity. In some cases, training may be performed offline, through the direct simulation of the optical apparatus or by “transfer learning" from an artificial neuron. Up to fine tuning, parameter transferring is supported by the mathematical analogy between the classical and the quantum response. In other cases, training might require an experimental approach. If so, the cost would be determined by the number of epochs, which depends on the pattern complexity but not on the image resolution: it is a constant overhead.

3.5 Comparison with a Mach-Zehnder interferometer

The previous discussion demonstrates the advantage of our method against a classical artificial neuron. We also compare our method against a classical optical design. We replace the artificial neuron with a Mach-Zehnder interferometer, which measures the overlap between the input and the probe states using classical, e.g. coherent, light. We describe the input as a multimode coherent state

$$\left|\xi_\phi\right\rangle = D(\xi, \phi)\,|0\rangle \,, \tag{3.64}$$

with intensity $|\xi_\phi|^2$ and spatial spectrum ϕ. Here

$$D(\xi, \phi) = \exp\left(\xi A_\phi^\dagger - \xi^* A_\phi\right), \tag{3.65}$$

$$A_\phi^\dagger = \int \mathrm{d}^2k \; \hat{\phi}(k) a_\omega^\dagger(k) \,, \tag{3.66}$$

are the multimode displacement and creation operators. In this case, the 50 : 50 beam splitter separates the source into two multimodes, A and B. Using Eqs. (3.3) and (3.4), we get

$$\left|\xi_\phi\right\rangle \to \exp\left[\frac{1}{\sqrt{2}}\left(\xi A_\phi^\dagger - \xi^* A_\phi\right) + \frac{1}{\sqrt{2}}\left(\xi B_\phi^\dagger - \xi^* B_\phi\right)\right]|0\rangle = \left|\frac{\xi_\phi}{\sqrt{2}}\right\rangle\left|\frac{\xi_\phi}{\sqrt{2}}\right\rangle, \tag{3.67}$$

where B undergoes the same definition of Eq. (3.66), with respect to mode b. In this case, the beam splitter halves the source intensity across the output branches. Consider two SLMs, each with N pixels, encoding the input image and the trainable parameters. In the simplest scenario, each pixel can be associated to a single spatial mode, by placing a lens at each branch of the interferometer. At one focal length, this performs a Fourier transform, which, for finite number of modes, reads as the unitary transformation

$$U = \frac{1}{\sqrt{N}} \sum_{ij} \zeta_{ij} e_i \otimes e_j \ , \tag{3.68}$$

with $|\zeta ij|^2 = 1 \ \forall i, j$, and $e_i \otimes e_j$ labelling the ij-th component of U [82]. Similarly to Eq. (3.67), the multimode A transforms under the action of U, splitting each branch in N spatial modes $\{B_1, \dots, B_N\}$ and yielding

$$\left| \frac{\xi_\phi}{\sqrt{2}} \right\rangle \to \left| \frac{\xi_\phi}{\sqrt{2N}} \right\rangle \dots \left| \frac{\xi_\phi}{\sqrt{2N}} \right\rangle \ . \tag{3.69}$$

Encoding the SLM, i.e. having at least one photon for each pixel, requires a coherent state with $|\xi_\phi|^2 > 2N$, namely an average number of photons that scales as $\Omega(N)$. This means that a classical interferometer would bring a computational speedup without optical advantage, undergoing the same cost of image detection and subsequent classification. In our method, single-photon states do not factorize as in Eq. (3.69). They use entanglement to encode the SLM in a superposition of N spatial mode, which, combined with the coincidence measurement, leads to the $\mathcal{O}(1)$ scaling.

3.6 Implementation with spatial light modulators

This section contains an explicit implementation of our setup. First, we replace the generic probe state with a toy model of an amplitude-only spatial light modulator, placed in the top branch of the Hong-Ou-Mandel interferometer. Experimentally, this can be implemented using a liquid crystal display with negligible losses [137]. We still describe the input as a free-space object. However, different approaches can be investigated, such as phase-only SLMs [138], which may exhibit superior resiliency against losses. In the second part, we consider this situation by using two phase-only SLMs inserted in both branches of the interferometer, injecting both the input and the probe states as digital objects.

3.6.1 Amplitude modulation

Consider an amplitude-only SLM, i.e. a pattern on a grayscale grid with N real amplitudes $\{\lambda_{\mu\nu}\}$. Following Fig. 3.2, we place the SLM in the probe branch of the interferometer. Each pixel, labelled by (μ, ν), is represented by an $L \times L$ square with center

$$r_{\mu\nu} = (\mu + 1/2, \nu + 1/2)L \ . \tag{3.70}$$

Resources		Quantum optical neuron	Imaging → Classical artificial neuron
Computational (# of mathematical operations)		$\mathcal{O}(1)$	N
Optical (# of photons)	Imaging	The image is not reconstructed	$\varsigma^{-2}\langle x\rangle^2 N$
	Classification	$\mathcal{O}(\varepsilon^{-2})$	$\Omega\left(\varepsilon^{-2}\langle x\rangle N\right)$

Table 3.1: Extended comparison of the computational and optical resources needed to classify an image x of N pixels. (Computational) An artificial neuron requires N mathematical operations to inject the parameters. Our method performs the same task by optical means, through a single Hong-Ou-Mandel interferometer. No computation is required other than scalar post-processing. (Optical) In coherent imaging, the number of photons needed depends linearly on the output resolution, i.e. $\varsigma^{-2}\langle x\rangle^2 N$. Here, ς and $\langle x\rangle$ are the standard deviation and the average brightness of the reconstructed image (which depends on the reflectivity of the object). The same also holds for the artificial neuron, after which the cost increases to $\Omega(\varepsilon^{-2}\langle x\rangle N)$, where ε is the uncertainty on the classification outcome, i.e. the error on the predicted class. Our method achieves the same task and uncertainty, with number of photons dictated only by the binomial statistics, independently of the resolution. In both cases, the quantum optical neuron requires constant resources, achieving a superexponential speedup. Table adapted from [2].

Upon an overall parameter-independent normalization, the probe can be approximated as a combination of (normalized) top-hat functions

$$\mathcal{U}_\omega(r|\lambda) = \sum_{\mu\nu} u(r - r_{\mu\nu}) \frac{\lambda_{\mu\nu}}{||\lambda||} , \tag{3.71}$$

$$\text{with } ||\lambda||^2 = \sum_{\mu\nu} \lambda_{\mu\nu}^2 . \tag{3.72}$$

Here

$$u(r) := \mathrm{He}(r + L/2) - \mathrm{He}(r - L/2) , \tag{3.73}$$

with He the two-dimensional Heaviside step function. Under this choice, the optical response, with free-space input $\mathcal{I}$, simplifies to

$$f_\lambda(\mathcal{O}) = \left| \sum_{\mu\nu} (u \star \mathcal{I}_\omega)(r_{\mu\nu}) \frac{\lambda_{\mu\nu}}{||\lambda||} \right|^2 , \tag{3.74}$$

where $\star$ is the cross-correlation operation. We introduce a bias and a sigmoid activation function, so that the post-processed output reads

$$F_{b\lambda}(\mathcal{O}) = \sigma(f_\lambda(\mathcal{O}) + b) \tag{3.75}$$

Assuming that $\mathcal{I}$ is real, Eq. (3.39) simplifies to

$$\partial_{\mu\nu} f \simeq 2\frac{\sqrt{f}}{||\lambda||}\left[(u \star \mathcal{I}_\omega)(r_{\mu\nu}) - \sqrt{f}\frac{\lambda_{\mu\nu}}{||\lambda||}\right], \tag{3.76}$$

with $\partial_{\mu\nu} f := \partial f / \partial \lambda_{\mu\nu}$. This expression can be evaluated in an all-optical way, by taking the amplitude measurement of $\mathcal{I}$ directly in the left branch of the interferometer, before the beam splitter. This operation can be done offline, and once per training object.

3.6.2 Phase modulation

In this section, we repeat the above discussion using two phase-only SLMs, which describe both the input and the probe branches of the Hong-Ou-Mandel interferometer. In this case, let $\{\lambda_{\mu\nu}\}$ be the set of trainable phases, and $\{x_{\mu\nu}\}$ the phase-encoded gray values corresponding to the input image. In both branches, we approximate the input and the probe states as a combination of (normalized) top-hat functions with disjoint support

$$\mathcal{I}_\omega(r|\mathcal{O}) = \frac{1}{\sqrt{N}}\sum_{\mu\nu} u(r - r_{\mu\nu}) \exp\big(i x_{\mu\nu}\big), \tag{3.77}$$

$$\mathcal{U}_\omega(r|\lambda) = \frac{1}{\sqrt{N}}\sum_{\mu\nu} u(r - r_{\mu\nu}) \exp\big(i \lambda_{\mu\nu}\big). \tag{3.78}$$

For simplicity, we considered two SLMs with the same number of pixels N. Notice that the normalization is automatically guaranteed and depends only on the size of the SLM. The inner product and its derivative read

$$\langle \mathcal{I}_\omega, \mathcal{U}_\omega \rangle = \frac{1}{N}\sum_{\mu\nu} \exp\big[-i(x_{\mu\nu} - \lambda_{\mu\nu})\big], \tag{3.79}$$

$$\langle \mathcal{I}_\omega, \partial_\lambda \mathcal{U}_\omega \rangle = \frac{i}{N} \exp\big[-i(x_{\mu\nu} - \lambda_{\mu\nu})\big]. \tag{3.80}$$

Here, we used that the supports of the top-hat functions are all disjoint. Finally, a substitution in Eq. (3.39) gives

$$\partial_{\mu\nu} f = \frac{2}{N^2}\sum_{ab} \sin\big(x_{ab} - x_{\mu\nu} - \lambda_{ab} + \lambda_{\mu\nu}\big). \tag{3.81}$$

By neglecting the global phase contribution, such that $\langle \mathcal{I}, \mathcal{U} \rangle \simeq \sqrt{f}$, this expression simplifies to

$$\partial_{\mu\nu} f \simeq 2\frac{\sqrt{f}}{N} \sin\big(\lambda_{\mu\nu} - x_{\mu\nu}\big). \tag{3.82}$$

Both expressions can be easily evaluated, due to the complete control of the parameter injected in the SLMs.

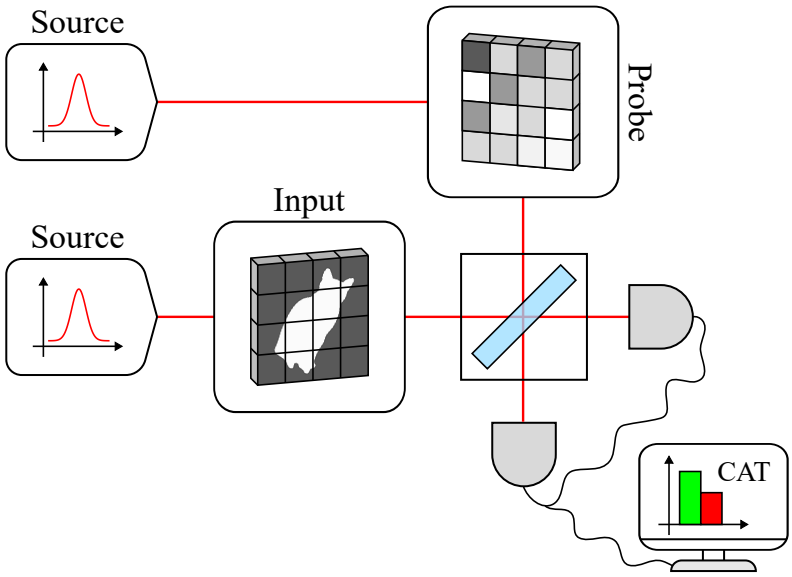


Figure 3.4: Setup used in the simulations. The target object and the optics are replaced by a second SLM. The discretized input and spectra correspond to the digital images shown on the screen. Both SLMs have the same number of pixels. Figure adapted from [2].

3.7 Simulations

In this section, we present a simulation of the quantum optical classifier, comparing its performance against classical neural network architectures. We investigate two widely recognized datasets: the MNIST, which contains 28×28 images of handwritten digits from 0 to 9, and the CIFAR-10, comprised of 32×32 color images distributed across 10 different classes. A fair comparison is guaranteed by padding the MNIST resolution to 32×32 pixels, while converting the CIFAR-10 to grayscale. Each image is identified with the discretized spectrum of a single-photon state, bypassing the simulation of the imaging apparatus. In practice, this means replacing the target object and the lenses with a second amplitude-only SLM, which displays and encodes the input image, with resolution of 32×32 pixels, both in the spatial and in the Fourier domains. We choose the binary cross-entropy as loss function, combined with the standard (non-stochastic) gradient descent optimizer, with learning rates $\eta_\lambda = 0.075$ and $\eta_b = 0.005$. We adopt the accuracy, i.e. the proportion of correct predictions over the total ones, as figure of merit of our results. The setup is schematically represented in Fig. 3.4. All the simulations are run in Python and TensorFlow [139]. The quantum optical classifier demonstrates significative performances in both datasets. In the MNIST, it achieves accuracy exceeding 99%, when discerning between zeros and ones. In the CIFAR-10, it reaches accuracy above 58%, when distinguishing between cats and dogs. This difference reflects the complexity of the two classification tasks. The results are reported in Fig. 3.5. Consider a classical artificial neuron and a convolutional neural network, commonly employed in pattern recognition tasks [96, 130, 140]. Adopting the TensorFlow notation, the

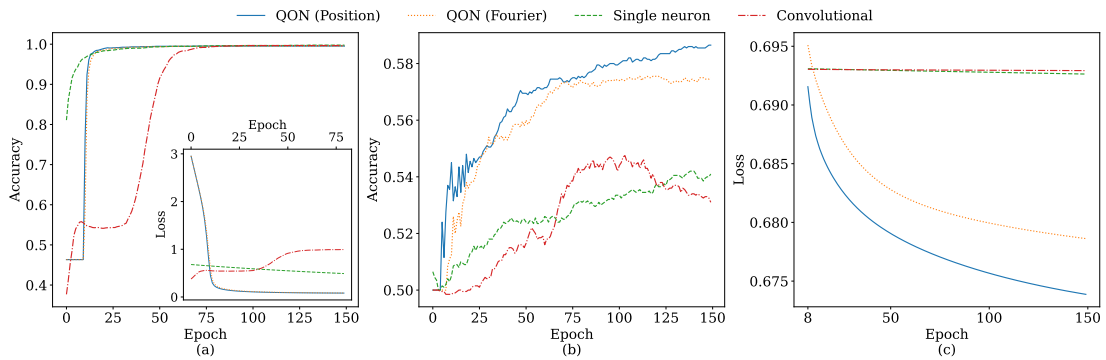


Figure 3.5: Comparison between the quantum optical neuron (QON), a single classical neuron and a convolutional network. (a) Accuracy versus the number of training epochs for the MNIST dataset. The models are trained to distinguish among images of zeros and ones, showing compatible results in terms of trainability and accuracy. The inset shows the binary cross-entropy, used as loss function in the gradient descent optimization. (b-c) Accuracy and binary cross-entropy plots versus the number of training epochs for the CIFAR-10 dataset. The models are trained to classify images of cats and dogs, with the QON showing an advantage over its classical counterparts. Figure adapted from [2].

convolutional architecture reads

$$\text{Conv2D}\,(10, 3\times 3) \rightarrow \text{Conv2D}\,(4, 2\times 2) \rightarrow \text{MaxPooling2D}\,(2\times 2)\,. \tag{3.83}$$

All the architectures are trained with ~ 1024 parameters, same optimizer and learning rates. The performances are equal in the MNIST dataset, both in terms of trainability and final accuracy. In the CIFAR-10 dataset, our classifier outperforms the conventional ones, showing superior efficiency under a strongly-constrained parameters count. These findings emphasize the competitive accuracy of our method, and its comparative advantage in pattern recognition tasks with a limited number of parameters. Additional tests can help in evaluating the effectiveness of our method. First, we repeat the above comparisons by removing the post-processing operations, i.e. the bias and the sigmoid, using the square absolute value activation only. Under the same conditions of the above simulations, the model maintains its trainability with reduced asymptotic accuracy, i.e. 88% and 57% for the MNIST and CIFAR-10 datasets, respectively. Then, we assess robustness by introducing sampling noise when estimating the coincidence rates at the Hong-Ou-Mandel output, for 100 and 1000 experimental repetitions. In the MNIST dataset, predictability remains unaffected. In the CIFAR-10 dataset, higher noise sensitivity is observed, although with compatible to superior capabilities with respect to its classical counterparts. In both cases, local fluctuations do not affect the global convergence of training. Results are reported in Figs. 3.6 and 3.7.

3.8 Discussion

In this chapter, we discussed an interferometric setup of a quantum binary classifier. We demonstrated the mathematical relation between our model, implemented through the Hong-Ou-Mandel effect, and a single artificial neuron, showing their similarity in terms of processing and response function. Being completely optical and single-photon based, the quantum optical classifier provides a superexponential speedup with respect to its classical counterpart, in terms of the number of photons and computational resources used. After its explicit implementation with SLMs, we performed a numerical comparison against conventional neural network designs, demonstrating compatible to superior capabilities in terms of accuracy and training convergence, as well as noise resiliency and trainability in absence of additional post-processing. Our method can be generalized to multiclass scenarios, adopting one-vs-one or one-vs-all strategies. In this case, multiple sets of parameters have to be considered and trained to subsequently, or adversarially, determine the target class.

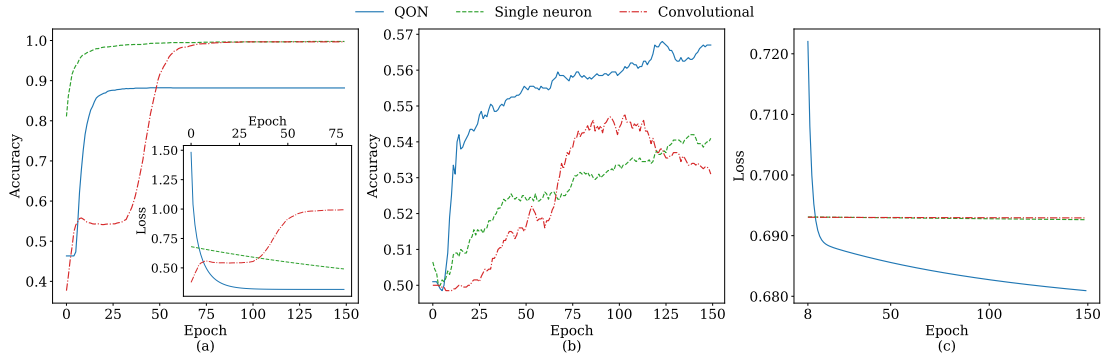


Figure 3.6: Quantum optical neuron (QON) with square absolute value activation function. Comparison against a single classical neuron and a convolutional network. (a) History plot for the MNIST dataset, when classifying zeros and ones. The inset shows the binary cross-entropy. (b-c) Accuracy and loss histories for the CIFAR-10 dataset, when classifying cats and dogs. Figure adapted from [2].

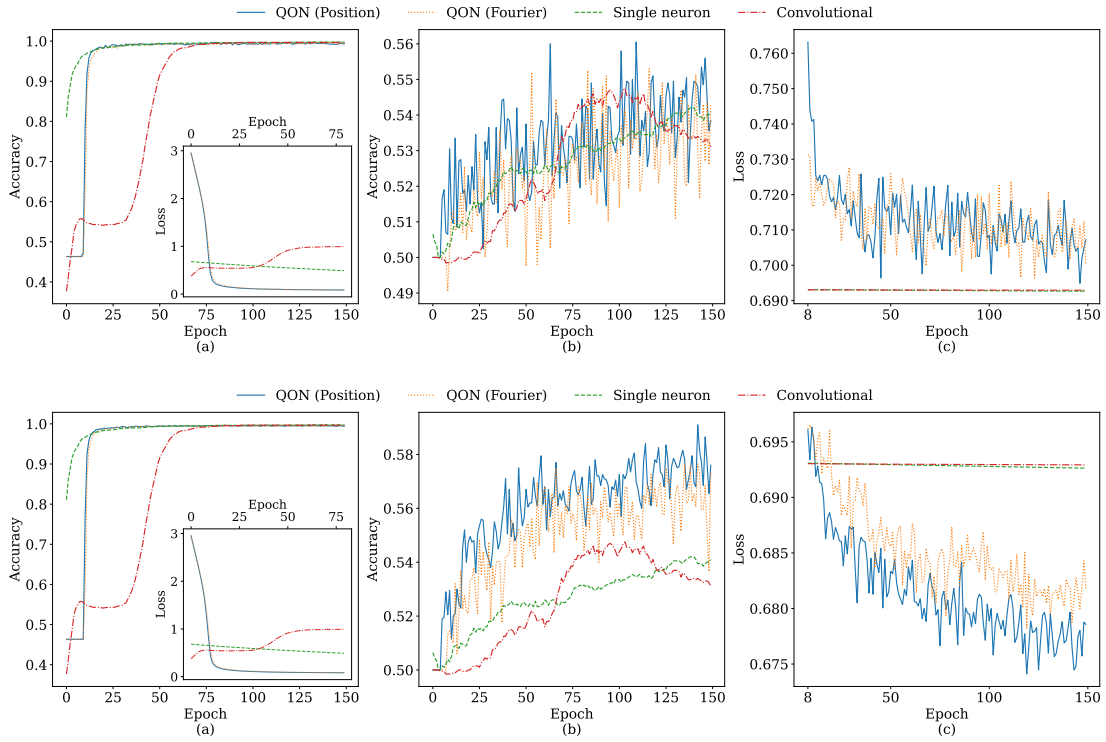


Figure 3.7: Quantum optical neuron (QON) with coincidence approximated by empirical frequencies, sampled with 100 (upper plots) and 1000 (lower plots) repetitions. Comparison against a single classical neuron and a convolutional network. (a) History plot for the MNIST dataset, when classifying zeros and ones. The inset shows the binary cross-entropy. Our method demonstrates high noise resiliency. (b-c) Accuracy and loss histories for the CIFAR-10 dataset, when classifying cats and dogs. Figure adapted from [2].

APPENDIX

3.A Classification in the frequency domain

In this section, we discuss the effect of adding a lens in the probe branch of the Hong-Ou-Mandel interferometer, translating the classification problem to the Fourier domain. Consider a thin lens, placed at one focal length ℓ from both the probe image plane and the beam splitter. In the near-field limit, the lens performs a Fourier transform of the probe state [60], yielding $|\Psi_{\mathcal{U}}\rangle \to |\Psi_{\mathcal{U}'}\rangle$, where

$$\mathcal{U}'_\omega(r|\lambda) = -i\frac{\omega}{\ell}\exp(2i\omega f)\hat{\mathcal{U}}_\omega\left(\frac{\omega}{\ell}r\middle|\lambda\right) . \tag{3.84}$$

After the beam splitter, the rate of coincidences is

$$p(1_a \cap 1_b|\lambda,\mathcal{O}) = \frac{1}{2}\left[\alpha_\omega(\mathcal{O}) - \widetilde{f}_\lambda(\mathcal{O})\right] , \tag{3.85}$$

$$\text{with } \widetilde{f}_\lambda(\mathcal{O}) = \left|\langle\mathcal{I}_\omega(\cdot|\mathcal{O}),\hat{\mathcal{U}}_\omega(\cdot|\lambda)\rangle\right|^2 . \tag{3.86}$$

This explicitly reads

$$\widetilde{F}_{b\lambda}(\mathcal{O}) = \sigma(\widetilde{f}_\lambda(\mathcal{O}) + b) , \tag{3.87}$$

$$\text{with } \widetilde{f}_\lambda(\mathcal{O}) = \left|\int_I \mathrm{d}r\,\mathrm{d}r'\;\mathcal{I}_\omega(r|\mathcal{O})\mathcal{U}^*_\omega(r'|\lambda)e^{ir\cdot r'}\right|^2 , \tag{3.88}$$

with σ and b the sigmoid activation function and bias, already introduced in the main discussion. In this case, $\widetilde{f}_\lambda(\mathcal{O})$ is not a point-wise evaluation: it combines the image spatial modes with the momentum spectrum of the probe state. Using the duality of the Fourier transform, it follows that

$$\widetilde{f}_\lambda(\mathcal{O}) = \left|\langle\hat{\mathcal{I}}_\omega(\cdot|\mathcal{O}),\mathcal{U}_\omega(\cdot|\lambda)\rangle\right|^2 , \tag{3.89}$$

which corresponds to the output illustrated above, but with the thin lens placed in the left branch, before the beam splitter. Equivalently, this takes the Fourier transform of the image, instead of that of the probe. In the next section, we leverage this symmetry to simplify both the training process and the numerical simulations. Finally, training follows the same procedure described before. After placing the lens in the top branch of the interferometer, while using the duality of the Fourier transform, we get

$$\partial_\lambda\widetilde{f} \simeq 2\sqrt{\widetilde{f}}\,\mathrm{Re}\left[\langle\hat{\mathcal{I}}_\omega,\partial_\lambda\mathcal{U}_\omega\rangle\right] . \tag{3.90}$$

We substitute the generic probe with an amplitude-only SLM. By neglecting the phase of $\langle \hat{\mathcal{I}}, \mathcal{U} \rangle$, the last two equations become

$$\widetilde{f}_{\lambda}(\mathcal{O}) = \left| \sum_{\mu,\nu} (u \star \hat{\mathcal{I}}^{*}_{\omega})(r_{\mu\nu}) \frac{\lambda_{\mu\nu}}{||\lambda||} \right|^{2} , \tag{3.91}$$

$$\partial_{\mu\nu} \widetilde{f} \simeq 2 \frac{\sqrt{\widetilde{f}}}{||\lambda||} \mathrm{Re} \left[(u \star \hat{\mathcal{I}}^{*}_{\omega})(r_{\mu\nu}) - \sqrt{\widetilde{f}} \frac{\lambda_{\mu\nu}}{||\lambda||} \right] . \tag{3.92}$$

Again, Eq. (3.92) can be evaluated in an all-optical way through the characterization of the real part of $\hat{\mathcal{I}}$, namely, by performing an amplitude and phase measurement at the output of a thin lens, placed in the left branch, before the beam splitter. In our simulations, we compare the predictability of the neuron in the spatial and Fourier domains.

PART II

QUANTUM NOISE

CHAPTER 4

QUANTUM CHANNEL TOMOGRAPHY

Many hundreds of thousands of measurements would have to be taken, [...] that would require an equal number of many hundreds of thousands of simultaneous equations to be solved.

Godfrey N. Hounsfield [141]

QUANTUM SYSTEMS are characterized by a high sensitivity to noise, which can affect the measurement outcome of any observable or experimental procedure. Quantum channels provide a dynamical framework for systems that interact with their surrounding environment. They describe a wide variety of operations, including unitary transformations, communication and teleportation protocols [142] or noisy processes, e.g. the bit-flip, the depolarizing or the amplitude damping channels, also in the presence of correlations [143–146]. A quantum channel behaves like a black box, mapping states into other states, and it is mathematically represented as a linear completely-positive and trace-preserving map [82]. Quantum process tomography is the identification of an unknown quantum channel, obtained by correlating a complete set of states to a complete set of measurements at their output. This can be achieved directly, when the measurement outcome immediately enters the channel reconstruction, eventually after post-processing the data [147–153], or indirectly, when additional techniques are employed to analyse the output state, e.g. state tomography [154–156]. In both cases, learning-based approaches have shown significant capabilities in reconstructing channels and their properties [157–162]. There are alternative representations in which process tomography can be formulated [163].

When states are vectorized, channels are mapped to matrices, called Pauli transfer matrices [148, 149]. In this representation, identifying a channel means reconstructing its Pauli transfer matrix, by combining the outcome of different experimental configurations into each of its entries. Each configuration consists of a couple of input

state and output measurement, repeatedly fed into the system for a certain number of shots, i.e. repetitions to retrieve the necessary statistics. In this chapter, we introduce a specific set of input states, for which each Pauli transfer matrix element is directly identified by few Pauli measurements performed at the channel output. In the standard approach, a full tomography requires d^4 configurations, where d is the dimension of the system. Our method uses only 2 experimental configurations for each matrix element (only 1 for unital channels), independently of the dimension of the system: an exponential speedup with respect to the minimum resources spent by standard process tomography to complete the same task.

This chapter is built upon the results of [3].

4.1 Preliminaries on the Pauli transfer matrix

Consider the Hilbert space of a n-qubit system, i.e. a system of n qubits. The Pauli operator basis reads

$$\mathcal{B} = \left\{ \sigma_{\alpha_1} \otimes \sigma_{\alpha_2} \otimes \ldots \otimes \sigma_{\alpha_n} \text{ with } \alpha_1, \alpha_2, \ldots, \alpha_n = 0, 1, 2, 3 \right\} , \tag{4.1}$$

with $\sigma_1 = X$, $\sigma_2 = Y$ and $\sigma_3 = Z$ the Pauli operators, and $\sigma_0 = \mathbb{1}_2$. We denote $\mathbb{1}$ the identity operator, with dimension specified by the subscript. We write the operator basis in the following notation

$$\mathcal{B} = \left\{ \mathcal{P}_k \text{ with } k = 0, 1, 2, 3, \ldots, d^2 - 1 \right\} , \tag{4.2}$$

with $d = 2^n$. Here, $\mathcal{P}_k$ denotes the kth basis element in lexicographic order. Let ρ be the system state. A quantum channel, or quantum operation, is a linear completely-positive trace-preserving (CPTP) map [82], that acts on the state as

$$\rho \to \Phi(\rho) . \tag{4.3}$$

There are several ways to represent a quantum channel. In the Kraus representation, the channel is described by a collection of operators $\{K_i\}$, called Kraus operators, such that

$$\begin{aligned} \Phi(\rho) &= \sum_i K_i \rho K_i^\dagger , \\ \text{with } & \sum_i K_i^\dagger K_i = \mathbb{1} . \end{aligned} \tag{4.4}$$

In this chapter, we consider an alternative representation, in which the channel is described by a $d^2 \times d^2$ matrix

$$\Gamma_{ij} = \frac{1}{d} \mathrm{Tr}\left[\mathcal{P}_i \Phi(\mathcal{P}_j) \right] , \tag{4.5}$$

called Pauli transfer matrix (PTM) [148]. This finds a natural application in the vectorized formalism [164], in which operators are mapped to $1 \times d^2$ column vectors and channels act through their PTMs, by standard matrix multiplication. Let B be an operator on the Hilbert space. Its vectorized representation reads

$$\begin{aligned} |A\rangle\rangle &= \sum_k A_k |k\rangle\rangle , \\ \text{with } A_k &= \langle\langle k|A\rangle\rangle . \end{aligned} \tag{4.6}$$

Here $\langle\langle \cdot | \cdot \rangle\rangle$ denotes the Hilbert-Schmidt product

$$\langle\langle A|B\rangle\rangle := \frac{1}{d} \mathrm{Tr}\left[A^\dagger B \right] , \tag{4.7}$$

with B any operator on the Hilbert space, and $|k\rangle\rangle$ the kth element of the Pauli basis. The normalization factor d, eventually absorbed by the substitution $\mathcal{P}_k \to \mathcal{P}_k/\sqrt{d}$,

guarantees the orthonormality of the Pauli basis with respect to the inner product. The action of the channel Φ on A reads

$$\begin{aligned} \Phi(A) &= \Gamma \left|A\right\rangle\rangle \ , \\ \text{with } \Gamma_{ij} &= \langle\langle i| \, \Gamma \, |j\rangle\rangle \ . \end{aligned} \tag{4.8}$$

Conversely, the PTM of a quantum channel can be traced back to its Kraus decomposition. Consider the Choi representation of Φ, namely

$$\Lambda = \sum_{ij} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|) \ , \tag{4.9}$$

with $\{|i\rangle \text{ with } i = 0, 1, \ldots, d\}$ the basis of the n-qubit Hilbert space [165]. A reshuffling transformation bijectively relates the PTM and the Choi matrix, whose spectrum leads back to the original Kraus representation of the channel [163]. By definition, the following properties of hold.

Proposition 2. Consider a n-qubit system, with dimension $d = 2^n$. Let Φ be a quantum channel, and Γ its PTM. The TP condition implies that

$$\Gamma_{0j} = \delta_{0j} \ . \tag{4.10}$$

with $0 \leq i, j \leq d^2 - 1$ and δ_{ij} denoting the Kronecker delta. If the channel is unital, i.e. $\Phi(\mathbb{1}) = \mathbb{1}$, then also

$$\Gamma_{i0} = \delta_{i0} \ . \tag{4.11}$$

Proof. This follows by substituting the TP condition in the PTM, combined with the properties of the Pauli basis, i.e. that $\mathrm{Tr}\big[\mathcal{P}_j\big] = d\,\delta_{0j}$ and that $\mathcal{P}_j\mathcal{P}_j = \mathbb{1}$. □

The above statement summarizes the basic properties and constraints of PTMs. Further simplifications hold for Pauli channels, namely those channels with Kraus representation generated by Pauli operators only, e.g. the bit-flip, the depolarizing or the dephasing models [82]. Importantly, the PTM of a Pauli channel is diagonal [166].

4.2 Tomographic reconstruction

Quantum process tomography seeks to reconstruct an unknown quantum channel Φ from the statistics of a collection of experimental configurations. Namely, it applies a set of measurements $\{E_i\}$ to different input states $\{\rho_j\}$. We call each couple of input state and output measurement (E_i, ρ_j) experimental configuration, or simply configuration. After gathering sufficient statistics, i.e. after many experimental repetitions over multiple copies of the same state ρ_j, the outcome associated with (E_i, ρ_j) reads

$$p_{ij} = \mathrm{Tr}\big[E_i \Phi(\rho_j)\big] \ . \tag{4.12}$$

From the completeness relation, this can be written in terms of the PTM as [148]

$$p = \alpha \Gamma \beta \ , \tag{4.13}$$

with

$$\alpha_{ij} = \frac{1}{d}\,\mathrm{Tr}\big[E_i \mathcal{P}_j\big]\,, \tag{4.14}$$

$$\beta_{ij} = \mathrm{Tr}\big[\mathcal{P}_i \rho_j\big]\,. \tag{4.15}$$

We refer to α and β as reconstruction matrices. Their inversion, if exists, yields the PTM in terms of the measurement, namely

$$\Gamma = \alpha^{-1} p \beta^{-1}\,. \tag{4.16}$$

In practical situations, the PTM is reconstructed using a tomographic fitter instead of Eq. (4.16), for example a least-squares minimization or a maximum likelihood estimation [167, 168]. For simplicity, we bypass the above approaches, by comparing our method (which can also be adapted to them) against the linear inversion solution given by Eq. (4.16).

We discuss a specific set of configurations, which provides a direct reconstruction of the PTM from the experimental data. Consider the states

$$\rho_0 = \frac{\mathbb{1}}{d}, \quad \rho_k = \frac{\mathbb{1} + \mathcal{P}_k}{d} \quad \text{for } k = 1, 2, 3, \ldots, 2^{2n} - 1\,, \tag{4.17}$$

which can be compactly written as

$$\rho_k = \frac{1}{d}\big[(1 - \delta_{0k})\mathbb{1} + \mathcal{P}_k\big]\,. \tag{4.18}$$

These states are mixed, except for $n = 1$, i.e. for a single qubit, and $k \neq 0$. In the next sections, we discuss normalization and positive definiteness, as well as different strategies for their preparation. Under the action of the channel, they evolve as

$$\Phi(\rho_k) = \frac{1}{d}\big[(1 - \delta_{0k})\Phi(\mathbb{1}) + \Phi(\mathcal{P}_k)\big]\,. \tag{4.19}$$

The expectation value of the ith element of the Pauli basis reads

$$\langle \mathcal{P}_i \rangle_{\Phi(\rho_j)} = \mathrm{Tr}\big[\mathcal{P}_i \Phi(\rho_j)\big]\,. \tag{4.20}$$

By substitution of Eq. (4.19), we write Eq. (4.20) in terms of the PTM as

$$\langle \mathcal{P}_i \rangle_{\Phi(\rho_j)} = (1 - \delta_{0j})\Gamma_{i0} + \Gamma_{ij}\,. \tag{4.21}$$

Once inverted, it gives

$$\Gamma_{ij} = \begin{cases} \langle \mathcal{P}_i \rangle_{\Phi(\rho_0)} & \text{for } j = 0 \\ \langle \mathcal{P}_i \rangle_{\Phi(\rho_j)} - \langle \mathcal{P}_i \rangle_{\Phi(\rho_0)} & \text{for } j \neq 0 \end{cases} \tag{4.22}$$

$$\Rightarrow \Gamma_{ij} = \langle \mathcal{P}_i \rangle_{\Phi(\rho_j)} - (1 - \delta_{0j})\langle \mathcal{P}_i \rangle_{\Phi(\rho_0)}\,. \tag{4.23}$$

Notice that this expression consistently satisfies the properties of Proposition 2. Moreover, it provides a direct mapping between each PTM entry and, at most, two experimental configurations. For this reason, we refer to this technique as Pauli transfer

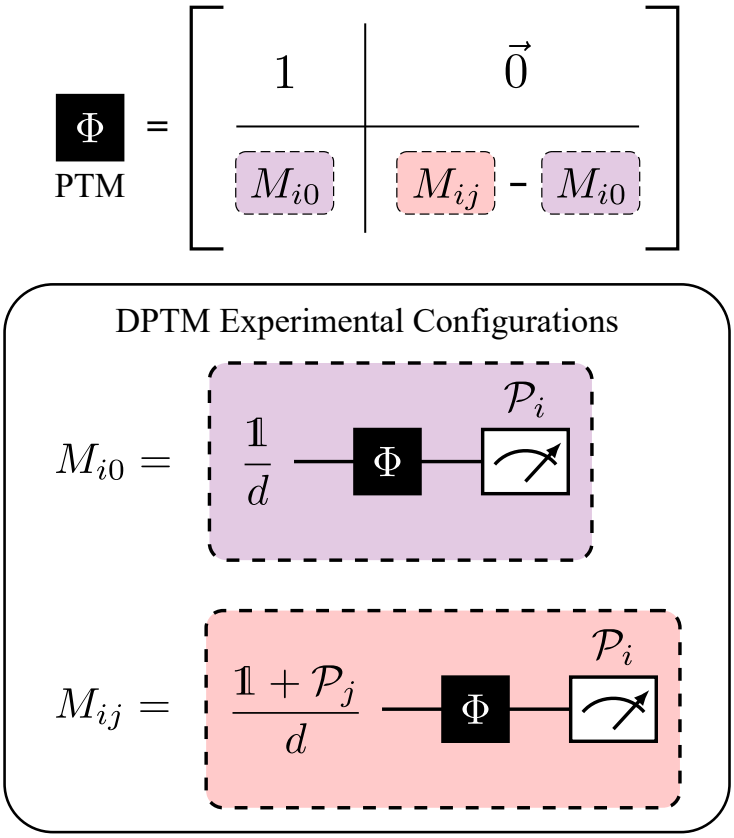


Figure 4.1: Direct reconstruction of an unknown quantum channel Φ. Two types of experimental configurations are considered. A first measurement characterizes the non-unital column of the PTM. The results in each row are then subtracted to those of the second set of experimental configurations, completing the characterization of the remaining entries. Independently of the number of qubits, the reconstruction of each entry requires at most 2 experimental configurations. Figure adapted from [3].

matrix direct reconstruction (DPTM). We summarize the protocol in Fig. 4.1. The relationship between these results and those of quantum process tomography follows by substituting Eq. (4.18) in Eq. (4.16). With $E_i = \mathcal{P}_i$ and

$$p_{ij} = \langle \mathcal{P}_i \rangle_{\Phi(\rho_j)} \,, \tag{4.24}$$

the reconstruction matrices read

$$\alpha_{ij} = \delta_{ij} \,, \quad \beta_{ij} = (1-\delta_{0j})\delta_{i0} + \delta_{ij} \,. \tag{4.25}$$

Once inverted, they yield

$$\alpha^{-1}_{ij} = \delta_{ij} \,, \quad \beta^{-1}_{ij} = \delta_{ij} - (1-\delta_{0j})\delta_{i0} \,. \tag{4.26}$$

By direct substitution, this expression reduces to Eq. (4.23). This means that our method represents a particular choice of basis and observables for quantum process tomography. Nevertheless, in the next section we show that this basis requires fewer resources, i.e. experimental configurations, when reconstructing a single PTM entry.

4.2.1 Positivity of the tomographic basis

In this section, we show that the operators of Eq. (4.18) fulfill all the requirements of density operators, i.e. Hermiticity, unit trace and positive semidefiniteness, namely, that they are eligible choices of states. Hermiticity is straightforward, while unit trace follows by observing that $\mathrm{Tr}[\mathcal{P}_k] = 0$ for $k \neq 0$. For positive semidefiniteness, we consider the representation of density operators in terms coherence vectors [169]. Let $\{S_m\}$ the coefficients of the characteristic polynomial of the matrix representation of ρ, recursively defined as

$$S_m = \frac{1}{m}\sum_{j=1}^{m}(-1)^{j-1}\,\mathrm{Tr}\left[\rho_k^j\right]S_{m-j}\,, \tag{4.27}$$

with $S_0 = 1$, $m = 0, 1, \ldots, d$ and the summation indexes explicitly reported. This relation is obtained by writing Eq. (24) from [169] as a series, with $S_0 = 1$ the coefficient of the highest-order term in the characteristic polynomial of ρ_k. A necessary and sufficient condition for ρ_k being positive semidefinite is that $S_m \geq 0\ \forall m$. We apply it below.

Proposition 3. Consider a n-qubit system, with dimension $d = 2^n$ and Pauli operator basis $\{\mathcal{P}_k \text{ with } k = 0, 1, 2, 3, \ldots, d^2 - 1\}$. The Hermitian, unit-trace operators

$$\rho_k = \frac{1 + \mathcal{P}_k}{d} \quad \text{for } k = 0, 1, 2, 3, \ldots, d^2 - 1 \tag{4.28}$$

are positive semidefinite. Hence, they are quantum states.

Proof. For any k, consider

$$A = \frac{1 + \mathcal{P}_k}{2}\,, \tag{4.29}$$

which is positive semidefinite if and only if ρ_k does. A direct computation yields $A^j = A$ and $\mathrm{Tr}\left[A^j\right] = d/2$. For A, Eq. (4.27) reads

$$S_m = \frac{d}{2m}\sum_{j=1}^{m}(-1)^{j-1}S_{m-j}\,. \tag{4.30}$$

Extracting the first term of the series and collecting a minus sign, we get

$$S_m = \frac{d}{2m}\left(S_{m-1} - \sum_{j=2}^{m}(-1)^{j-2}S_{m-j}\right)\,. \tag{4.31}$$

Translating $j \to j + 1$,with the sum now running from 1 to $m - 1$, and writing Eq. (4.30) for S_{m-1}, we obtain a recursive relation

$$S_m = \frac{d}{2m}\left(1 - \frac{2(m-1)}{d}\right)S_{m-1}\,. \tag{4.32}$$

With $\delta = 1 + d/2$, this yields

$$S_m > 0 \quad \text{for } 0 \leq m < \delta \,, \tag{4.33}$$
$$S_m = 0 \quad \text{for } \delta \leq m \leq d \,, \tag{4.34}$$

which implies that $S_m \geq 0 \ \forall m$. This result does not depend on k, nor do the coefficients S_m and the roots of the characteristic polynomial, i.e. on the eigenvalues of A. Hence, A is positive semidefinite for any choice of $\mathcal{P}_k$ and consequently so is ρ_k. □

4.2.2 Mixed state preparation

In this section, we address the problem of preparing the characterization states of Eq. (4.18). Indeed, this discussion applies to any set of input, possibly mixed, states. For $n = 1$, i.e. a single qubit, $k = 0$ labels the only mixed state of the set. Its implementation can be achieved either through an ancilla-assisted approach: by preparing the system in the maximally entangled state

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} \sum_{k=0}^{1} |k\rangle \otimes |k\rangle_a \,, \tag{4.35}$$

and tracing out the ancillary qubit

$$\rho_0 = \mathrm{Tr}_a \left[\, |\Psi_0\rangle\langle\Psi_0| \, \right], \tag{4.36}$$

or by classical random generation of its pure components

$$|\psi_0\rangle \in \{|0\rangle, |1\rangle\} \quad \text{with } p(|0\rangle) = p(|1\rangle) = \frac{1}{2} \,. \tag{4.37}$$

Notice that this is equivalent to preparing the state $\psi = |0\rangle\langle 0|$ and applying the quantum channel

$$\mathcal{E}_0(\psi) = \frac{1}{2}\psi + \frac{1}{2} X\psi X \,, \tag{4.38}$$

yielding $\rho_0 = \mathcal{E}_0(\psi)$. Starting from this, we discuss the n-qubit preparation of a set of, eventually mixed, states $\{\rho_k \text{ with } k = 0, 1, 2, 3, \ldots, d^2-1\}$. Consider $\{\psi_k \text{ with } k = 0, 1, 2, 3, \ldots, d^2-1\}$ a set of d^2, linearly independent, *pure* states. Purity is required to consider states that can be easily and procedurally generated. Let $\mathcal{E}$ be the channel that maps each state to an element of the target set, i.e. $\rho_k = \mathcal{E}(\psi_k) \ \forall k$. In the vectorized representation

$$\rho_k = \Lambda \psi_k \,, \tag{4.39}$$

with Λ denoting the PTM of $\mathcal{E}$. We split the preparation of the target set into the generation of a set of pure states and the subsequent simulation of a quantum channel. Using $\{\rho_k\}$ and $\{\psi_k\}$, we compute $\mathcal{E}$. This problem can be translated in the tomographic framework: finding a channel that provides the desired collection of input and output states. In this case, Eq. (4.16) reads

$$\Lambda = \pi B^{-1} \,, \tag{4.40}$$

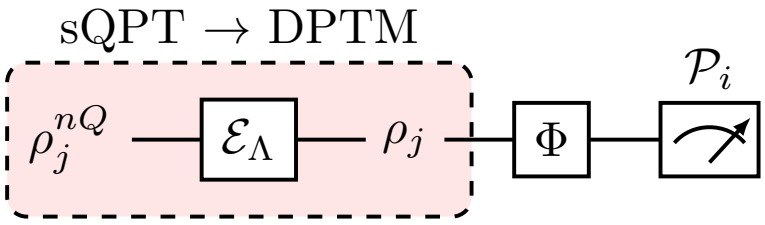


Figure 4.2: Preparation of the DPTM states from the standard quantum process tomography (sQPT) basis, implemented through the simulation of the channel $\mathcal{E}$ given by Eq. (4.40). The output is subsequently fed into the characterization of the unknown channel Φ, according to Eq. (4.23). Figure adapted from [3].

where $E_i = \mathcal{P}_i$ and

$$B_{ij} = \mathrm{Tr}\big[\mathcal{P}_i \psi_j\big]\,, \tag{4.41}$$

$$\pi_{ij} = \mathrm{Tr}\big[\mathcal{P}_i \mathcal{E}(\psi_j)\big] = \mathrm{Tr}\big[\mathcal{P}_i \rho_j\big]\,. \tag{4.42}$$

Since both $\{\rho_k\}$ and $\{\psi_k\}$ are known, we can compute Λ and procedurally simulate it, e.g. with methods like "PTM.to_instruction()" from the Qiskit package. For $n = 1$, the channel that prepares the states of Eq. (4.18), using the standard quantum process tomography basis [82], reads

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{4.43}$$

Although non-unital, this channel is almost trivial, since the two bases differ only for $k = 0$ when $n = 1$. We summarize this discussion in Fig. 4.2. In the next section, we provide a full comparison.

4.3 Comparison with standard process tomography

In this section, we perform a detailed comparison between the Pauli transfer matrix direct reconstruction (DPTM), and standard quantum process tomography (sQPT). We investigate under which conditions the former provides an advantage in terms of tomographic resources, i.e. the number of experimental configurations needed. With Eq. (4.23), we established a direct relation between PTMs and experimental configurations. When the channel is completely unknown, the resource cost is fixed by the dimensionality of the system, independently of the characterization method, i.e. d^2 input states coupled to d^2 Pauli measurements. However, when some prior knowledge about the channel is already available (e.g. unitality or its, eventually incomplete, Kraus representation), the direct reconstruction allows to easily drop those experimental configurations already fixed by the initial information. In this

case, the number of configurations reduces, and our method shows an advantage over quantum process tomography. See Table 4.1 for examples of resource costs for different channel types. Similar reductions apply to any partial extraction of the PTM, e.g. in the characterization of known theoretical models, in the estimation of unknown quantum parameters, or in assessing the unitality of a quantum channel. Consider $n = 1$. The operator basis contains four elements

$$\mathcal{P}_0 = \mathbb{1}_2 \,, \quad \mathcal{P}_1 = X \,, \quad \mathcal{P}_2 = Y \,, \quad \mathcal{P}_3 = Z \,, \tag{4.44}$$

while the computational basis states read $\{|0\rangle, |1\rangle\}$. Alternative bases can be obtained by applying the above operators, yielding

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}} \,, \tag{4.45}$$

$$|\pm i\rangle = \frac{|0\rangle \pm i\,|1\rangle}{\sqrt{2}} \,. \tag{4.46}$$

By substitution in Eq. (4.18), the DPTM states read

$$\rho_0 = \frac{1}{2}\big(|0\rangle\langle 0| + |1\rangle\langle 1|\big) \,, \tag{4.47}$$

$$\rho_1 = |+\rangle\langle +| \,, \tag{4.48}$$

$$\rho_2 = |i\rangle\langle i| \,, \tag{4.49}$$

$$\rho_3 = |0\rangle\langle 0| \,. \tag{4.50}$$

Previously, we discussed the preparation of ρ_0, which is the only mixed state in this set. The others are pure and can be prepared as

$$\{\rho_k = |\psi_k\rangle\langle\psi_k| \text{ with } k = 1,2,3\} \,, \tag{4.51}$$

$$|\psi_1\rangle = H\,|0\rangle \,, \quad |\psi_2\rangle = SH\,|0\rangle \,, \quad |\psi_3\rangle = |0\rangle \,, \tag{4.52}$$

with H and S respectively the Hadamard and the phase gates. For any state ρ, each Pauli measurement $\langle\mathcal{P}_i\rangle$ can be always obtained from a Z-measurement, i.e. by applying a unitary transformation $\rho \to U\rho U^\dagger$, so that

$$\langle\mathcal{P}_1\rangle = \mathrm{Tr}[ZH\rho H] \,, \tag{4.53}$$

$$\langle\mathcal{P}_2\rangle = \mathrm{Tr}\big[ZHS^\dagger\rho SH\big] \,, \tag{4.54}$$

$$\langle\mathcal{P}_3\rangle = \mathrm{Tr}[Z\rho] \,, \tag{4.55}$$

where we used that

$$\mathcal{P}_1 = H\mathcal{P}_3 H \,, \quad \mathcal{P}_2 = SH\mathcal{P}_3 HS^\dagger \,, \quad \mathcal{P}_3 = Z \,. \tag{4.56}$$

Each transformation translates the respective Pauli measurement to a simple count of the 0, 1 occurrences in the computational basis. For example, if $\big|\psi_f\big\rangle = c_0\,|0\rangle + c_1\,|1\rangle$ is a pure final state, eventually obtained by applying a unitary transformation to the

Classes of channels	Generic map	Trace preserving	Unital channel	Pauli channel
Resources (# of experimental configurations)	d^4	$d^2(d^2-1)$	$(d^2-1)^2$	d^2-1

Table 4.1: Resource cost for different types of constraints, in terms of the number of experimental configurations to collect all the data and complete the reconstruction. Table adapted from [3].

evolved one, then $\langle Z\rangle_{\rho_f} = |c_0|^2 - |c_1|^2$. The DPTM configurations follow by coupling each input state to the set of Pauli measurements in all the possible ways, namely

$$M^D_{ij} := \langle \mathcal{P}_i \rangle_{\Phi(\rho_j)} \, . \tag{4.57}$$

Then

$$\begin{cases} \Gamma_{i0} = M^D_{i0} \, , \\ \Gamma_{ij} = M^D_{ij} - M^D_{i0} \;\; \text{for } j \neq 0 \, . \end{cases} \tag{4.58}$$

In terms of reconstruction matrices α_D and β_D, this gives $\alpha_D^{-1} = \mathbb{1}$ and

$$\beta_D^{-1} = \begin{pmatrix} +1 & -1 & -1 & -1 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix} . \tag{4.59}$$

When the channel is unital, the identification is completely direct and $\Gamma_{ij} = M^D_{ij}$. For Pauli channels, it further simplifies to $\Gamma_{kk} = M^D_{kk}$. In both cases, the uncertainty is directly determined by that of measurement, without post-processing or further error propagation. For comparison, consider the protocol of sQPT, in which the set of Pauli measurements is usually coupled to the set of states

$$\rho^Q_0 = |1\rangle\langle 1| \;\; , \;\; \rho^Q_1 = |+\rangle\langle +| \;\; , \;\; \rho^Q_2 = |i\rangle\langle i| \;\; , \;\; \rho^Q_3 = |0\rangle\langle 0| \;\; . \tag{4.60}$$

Since the Pauli measurements are the same as above, we still obtain $\alpha_Q^{-1} = \mathbb{1}$. However, the different choice of input states modifies β_Q, whose inverse now reads

$$\beta_Q^{-1} = \frac{1}{2} \begin{pmatrix} +1 & -1 & -1 & -1 \\ 0 & +2 & 0 & 0 \\ 0 & 0 & +2 & 0 \\ +1 & -1 & -1 & +1 \end{pmatrix} . \tag{4.61}$$

The sQPT configurations are obtained by combining each input state to the set of Pauli measurements in all the possible ways, yielding

$$M^Q_{ij} := \langle \mathcal{P}_i \rangle_{\Phi\left(\rho^Q_j\right)} \, . \tag{4.62}$$

Explicitly, it follows that

$$\Gamma_{i0} = \frac{1}{2}\left(M^Q_{i0} + M^Q_{i3}\right) , \tag{4.63}$$

$$\Gamma_{i1} = \frac{1}{2}\left(-M^Q_{i0} + 2M^Q_{i1} - M^Q_{i3}\right) , \tag{4.64}$$

$$\Gamma_{i2} = \frac{1}{2}\left(-M^Q_{i0} + 2M^Q_{i2} - M^Q_{i3}\right) , \tag{4.65}$$

$$\Gamma_{i3} = \frac{1}{2}\left(-M^Q_{i0} + M^Q_{i3}\right) . \tag{4.66}$$

For each PTM entry, sQPT requires more experimental configurations than DPTM, and more resources in terms of post-processing recombinations of the data into the desired outcome. For example, a full characterization of a single-qubit Pauli channel costs 8 configurations using sQPT, but only 3 configurations with DPTM.

Finally, we compare the cost of both techniques when reconstructing a single PTM entry on n-qubit systems. On the one hand, DPTM scales always in the same way: independently of the dimension, the characterization states are given by Eq. (4.18), while the reconstruction matrix shows high sparsity, namely

$$\beta_D^{-1} = \begin{pmatrix} +1 & -1 & \dots & -1 \\ 0 & \ddots & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & +1 \end{pmatrix} . \tag{4.67}$$

On the other hand, the n-qubit basis of sQPT is obtained by taking all the possible n-fold Kronecker products of ρ_i^Q. The reconstruction matrix reads $(\beta_Q)^{\otimes n}$, with inverse $(\beta_Q^{-1})^{\otimes n}$. We compare the performance of the above methods, by defining $||\Gamma_{ij}||_D$ and $||\Gamma_{ij}||_Q$ as the number of experimental configurations respectively required by DPTM and sQPT. For $n = 1$, i.e. a single qubit, we have $||\Gamma_{i1}||_Q = 3$. We demonstrate that, independently of the number of qubits, DPTM performs exponentially better than sQPT in single-entry reconstructions. We summarize the advantage in Fig. 4.3.

Theorem 1. Consider a n-qubit quantum system, and an unknown channel Φ. Let Γ be its Pauli transfer matrix, reconstructed using DPTM and sQPT. Then

$$\max_{ij} ||\Gamma_{ij}||_D \leq \min_{ij} ||\Gamma_{ij}||_Q < \max_{ij} ||\Gamma_{ij}||_Q , \tag{4.68}$$

with the strict equality satisfied only for $n = 1$. Indeed

$$||\Gamma_{ij}||_D = \begin{cases} 1 & \text{for } j = 0 \\ 2 & \text{for } j \neq 0 \end{cases} , \tag{4.69}$$

while

$$\min_{i,j} ||\Gamma_{ij}||_Q = 2^n , \tag{4.70}$$

$$\max_{i,j} ||\Gamma_{ij}||_Q = 3^n . \tag{4.71}$$

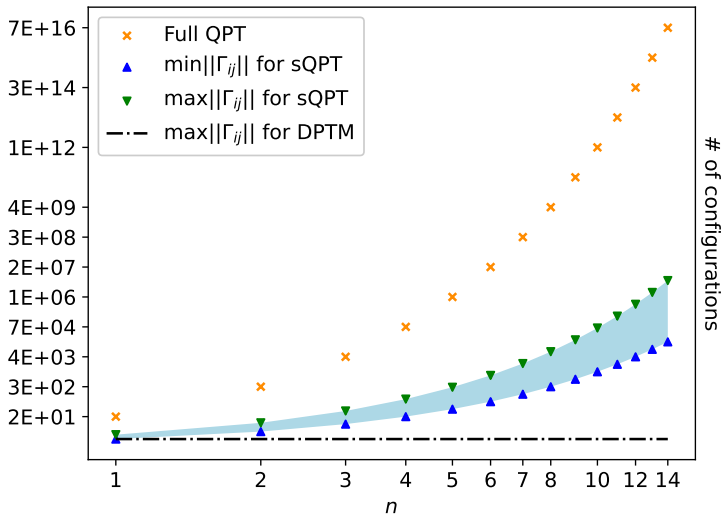


Figure 4.3: Number of experimental configurations to reconstruct a single PTM entry, with respect to the number of qubits n ($\log_2$ for both axes). DPTM provides an exponential advantage when reconstructing a single PTM entry, outperforming sQPT $\forall n > 1$. The shadowed area represents the possible cost when completing the same task with sQPT. On top, the number of resources needed for a full process tomography, i.e. the reconstruction of the whole PTM. Figure adapted from [3].

Proof. Eq. (4.69) is a direct consequence of Eq. (4.23). On the other hand, the sQPT reconstruction have

$$\Gamma = M^{Q}(\beta_Q^{-1})^{\otimes n} \,. \tag{4.72}$$

The result follows from Eq. (4.61), by respectively counting the minimum and maximum number of non-zero entries in the n-fold Kronecker product β_Q^{-1}. □

4.3.1 Example with amplitude damping channels

Consider a single-qubit amplitude damping channel, with Kraus operators

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix} , \tag{4.73}$$

$$K_1 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix} . \tag{4.74}$$

Here, p describes the transition probability from the state $|1\rangle$ to $|0\rangle$, e.g. when the system loses energy by spontaneous emission [82]. For reference, we compute the PTM of the channel, which reads

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-p} & 0 & 0 \\ 0 & 0 & \sqrt{1-p} & 0 \\ p & 0 & 0 & 1-p \end{pmatrix} . \tag{4.75}$$

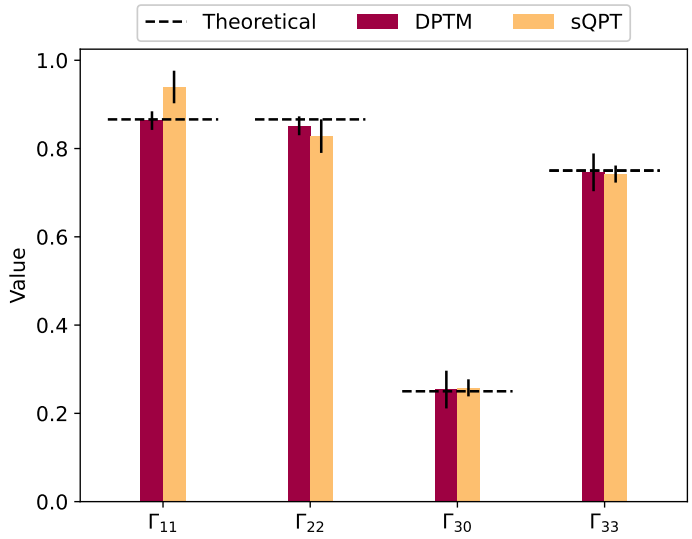


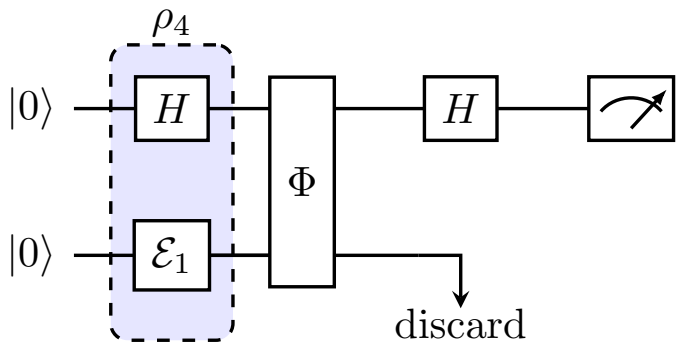
Figure 4.4: Reconstruction of a single-qubit amplitude damping channel. Simulation performed with Qiskit Aer, estimating the expectation values with 512 shots, i.e. experimental repetitions. The results of DPTM and sQPT are represented respectively by the left and right bars. The dashed lines represent analytical results. The error bars represent the standard error and are obtained by uncertainty propagation. Figure adapted from [3].

We discuss the characterization of the non-trivial components of this matrix, namely Γ_{11}, Γ_{22}, Γ_{30} and Γ_{33}, for $p = 0.25$. In practice, we simulate the experimental configurations of Eqs. (4.57) and (4.62). Both DPTM and sQPT are compatible with the theoretical prediction. However, DPTM achieves the same task using fewer experimental configurations: only 4 against the 8 required by sQPT. Additionally, the are differences in terms of statistical uncertainties. Indeed, sQPT combines more data into the same entries, potentially worsening the error propagation, while DPTM uses mixed states, which produces different statistics at the output. We report the results in Fig. 4.4.

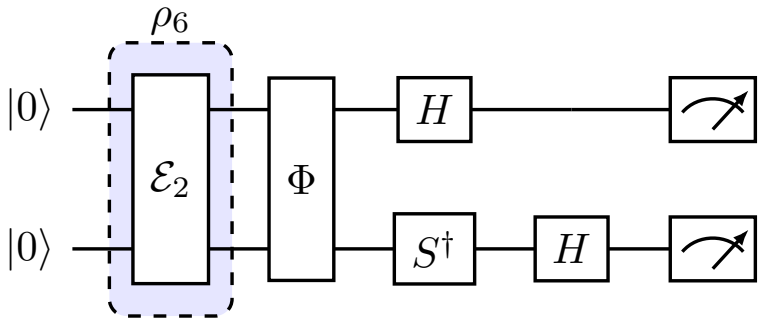
4.3.2 Example with depolarizing correlated channels

We consider a two-qubit depolarizing correlated channel, where the amount of correlation is measured by a parameter $\mu \in [0,1]$ [143, 170–175]. In general, Pauli correlated channels admit the following Kraus representation

$$\Phi(\rho) = \sum_{\alpha_1\alpha_2=0}^{3} p_{\alpha_1\alpha_2} K_{\alpha_1\alpha_2} \rho K_{\alpha_1\alpha_2}^{\dagger} , \quad \text{with } K_{\alpha_1\alpha_2} = \sigma_{\alpha_1} \otimes \sigma_{\alpha_2} . \tag{4.76}$$



(a)



(b)

Figure 4.5: Tomographic circuits for estimating (a) Γ_{44} and (b) Γ_{66}, using the Pauli transfer matrix direct reconstruction. In both cases, the mixed states are implemented through a quantum channel $\mathcal{E}_i$. Figure adapted from [3].

Correlations are modelled through the Markov chain

$$p_{\alpha_1\alpha_2} = p_{\alpha_1} p_{\alpha_2|\alpha_1} \,, \tag{4.77}$$

$$p_{\alpha_j|\alpha_i} = (1-\mu)p_{\alpha_j} + \mu\delta_{\alpha_i\alpha_j} \,, \tag{4.78}$$

where

$$\begin{aligned} \mathbf{p} &= [1-p, p_x, p_y, p_z]^T \,, \\ \text{with } p &= p_x + p_y + p_z \,. \end{aligned} \tag{4.79}$$

Here, $\mu = 0$ describes a memoryless channel, e.g. a noisy system in which the qubits are completely uncorrelated. On the other hand, $\mu = 1$ describes a full-memory channel, i.e. when the qubits exhibit complete correlation. From Eq. (4.76), a depolarizing correlated channel is obtained by choosing

$$\mathbf{p} = [1-3p/4, p/4, p/4, p/4]^T \,. \tag{4.80}$$

We compare DPTM and sQPT when extracting only the parameters p and μ, from a set of tomographic configurations. By direct computation, the PTM yields the following

relations

$$\Gamma_{44} = 1 - p \,, \tag{4.81}$$
$$\Gamma_{66} = (1-p)(\mu p - p + 1) \,. \tag{4.82}$$

Inverting them, gives the parameter as $p \leftarrow \Gamma_{44}$ and $\mu \leftarrow \Gamma_{44}, \Gamma_{66}$. On the one hand, sQPT provides the required entries using 15 experimental configurations $\{M^Q_{ij}\}$, with the Pauli basis as observables and the states obtained from all the possible 2-fold Kronecker products of the single-qubit set. The reconstruction matrix reads $(\beta_Q^{-1})^{\otimes 2}$. Mathematically, this yields

$$\Gamma_{44} = -\frac{1}{4}M^Q_{40} - \frac{1}{4}M^Q_{43} + \frac{1}{2}M^Q_{44} + \frac{1}{2}M^Q_{47} - \frac{1}{4}M^Q_{4(12)} - \frac{1}{4}M^Q_{4(15)} \,, \tag{4.83}$$

along with

$$\begin{aligned}\Gamma_{66} = \frac{1}{4}M^Q_{60} - \frac{1}{2}M^Q_{62} + \frac{1}{4}M^Q_{63} - \frac{1}{2}M^Q_{64} + M^Q_{66} + \\ -\frac{1}{2}M^Q_{67} + \frac{1}{4}M^Q_{6(12)} - \frac{1}{2}M^Q_{6(14)} + \frac{1}{4}M^Q_{6(15)} \,.\end{aligned} \tag{4.84}$$

On the other hand, DPTM provides the same components, using only 2 configurations. Namely

$$\Gamma_{44} = M^D_{44} \,, \tag{4.85}$$
$$\Gamma_{66} = M^D_{66} \,, \tag{4.86}$$

with the Pauli basis as observables and input states given by Eq. (4.18). In this case, the input states are prepared as follows. For Γ_{44}, the observable is $\mathcal{P}_4 = X \otimes I$, while the input state reads

$$\rho_4 = \frac{(\mathbb{1}_2 + X) \otimes \mathbb{1}_2}{4} = |+\rangle\langle+| \otimes \frac{\mathbb{1}_2}{2} \,, \tag{4.87}$$

with the completely mixed state is prepared through the channel

$$\begin{aligned}\mathcal{E}_1(\psi) = \psi/2 + X\psi X/2 \,, \\ \text{with } \psi = |0\rangle\langle 0| \,.\end{aligned} \tag{4.88}$$

For Γ_{66}, the observable is $\mathcal{P}_6 = X \otimes Y$, while the input state reads

$$\rho_6 = \frac{\mathbb{1} + X \otimes Y}{4} \,. \tag{4.89}$$

We prepare it by initializing the system to $\psi = |0\rangle\langle 0|$ and applying the channel

$$\mathcal{E}_2(\psi) = \frac{1}{2}U\psi U^\dagger + \frac{1}{2}VU\psi U^\dagger V^\dagger \,, \tag{4.90}$$
$$U = \text{CNOT}(S \otimes \mathbb{1})(H \otimes \mathbb{1}) \,, \tag{4.91}$$
$$V = (\mathbb{1} \otimes Z)(\mathbb{1} \otimes X) \,. \tag{4.92}$$

The circuit implementation of these configurations is reported in Fig. 4.5. We simulate these circuits, choosing $p = 0.25$ and $\mu = 0.75$, so that the expected results read $\Gamma_{44} = 0.750$ and $\Gamma_{66} \simeq 0.703$. We simulate the reconstruction, using Qiskit Aer with 2048 shots, i.e. experimental repetitions. We obtain

$$\text{DPTM} \rightarrow \begin{cases} \Gamma_{44} = & 0.749 \pm 0.015 \\ \Gamma_{66} = & 0.710 \pm 0.016 \end{cases}, \tag{4.93}$$

and

$$\text{sQPT} \rightarrow \begin{cases} \Gamma_{44} = & 0.756 \pm 0.015 \\ \Gamma_{66} = & 0.705 \pm 0.016 \end{cases}. \tag{4.94}$$

Both methods agree with the theoretical expectations. Again, DPTM performs better than sQPT: it requires fewer experimental configurations under the same number of shots and with compatible uncertainties.

4.4 Discussion

In this chapter, we considered the tomographic reconstruction of the Pauli transfer matrix of a quantum channel. In general, quantum process tomography performs a set of measurements on different experimental configurations, i.e. by changing the input state and/or the observable at the output of the channel. The results contribute to each element of the PTM. We presented an alternative technique, which directly reconstructs the PTM from the experimental outcomes. In principle, our approach differs only in the choice of the input states. Remarkably, this choice simplifies both the experimental implementation and the post-processing analysis. It exponentially reduces the number of configurations that contribute to a single matrix element, while keeping the same number of experimental repetitions. Indeed, our method requires (at most) 2 experimental configurations for each PTM entry, independently of the dimension of the system. This paves the way to more efficient (and scalable) experimental implementations of tomographic protocols, with fewer computational circuits or optical setups. This exponential gain is lost if one needs to reconstruct the whole PTM rather than few matrix elements. However, there are many situations where few entries are sufficient to recover the required channel characteristics, e.g. for multiparameter estimations in quantum metrology [176], or in assessing the unitality of a quantum channel. In contrast to other techniques like shadow tomography [177, 178], our approach does not improve the statistics of the reconstruction, since the total number of shots, or copies of the state, remains unchanged. Importantly, fewer settings can mitigate the systematics due to hardware errors while speeding up the overall experiment, by reducing the number of times the apparatus needs to be prepared in a different configuration. Our method truly shines when only a subset of the PTM has to be found, e.g. when characterizing known models like Pauli channels. In the next chapter we consider an explicit example, in which only a partial knowledge of a Pauli channel is needed to compute the noiseless expectation value of a Pauli observable.

CHAPTER 5

QUANTUM DECONVOLUTION

Excuse me! Would you mind stopping that noise?

James T. Kirk, Star Trek IV: The Voyage Home

NOISE and errors can affect the measurement outcome of any experiment. Indeed, noise sensitivity remains one of the main drawbacks that prevents quantum computers from outperforming their classical counterparts. Until quantum error correction is fully deployed [179–182], noise mitigation techniques are employed to reduce errors and improve the experimental results [39, 164, 183–185]. In this regard, we have already reviewed quantum channels, which provide a dynamical description of open quantum systems that interact with their environment, e.g. noisy systems. The characterization of such processes usually involves resource-hungry methods, e.g. quantum process tomography. Previously, we introduced a direct reconstruction technique that uses fewer tomographic resources, by leveraging the Pauli transfer matrix representation of the channel. In this chapter, we apply the same formalism to noise mitigation, i.e. to compute the noiseless expectation value of any observable from the noisy data.

We discuss a deconvolution technique [4, 186], which performs mitigation by means of a tomographic reconstruction formula that acts like a postprocessing operation on the noisy outcome, without introducing modifications to the system. Our method applies to any multiqubit (possibly correlated) channel, provided its inverse map mathematically exists, even if not physically implementable. We introduce an operational framework that is more suitable for multiqubit implementations. Our method works in general, e.g. it does not require Markovian correlations. In the specific but important case of Pauli channels [166], it provides a quadratic speedup over other reconstructions. We show that the noiseless expectation value can be tomographically reconstructed by performing local Pauli noisy measurements, whose number depends on the complexity of the target observable, after rescaling them by

the corresponding entries of the Pauli transfer matrix of the channel. This method works independently of the number qubits and does not require the complete inversion of the noise map. For this reason, the deconvolution of Pauli noise can take advantage of the direct reconstruction technique discussed in the previous chapter. Our protocol also applies to non-Pauli channels, with reduced efficiency. In this case, the deconvolution requires the complete inversion of the Pauli transfer matrix, showing the same scaling of quantum process tomography.

This chapter is built upon the results of [4].

5.1 Channel deconvolution

Consider an open system with quantum state ρ, surrounded by an environment. Ideally, a measurement of an observable O would give $\langle O \rangle_\rho = \mathrm{Tr}[O\rho]$. However, the presence of the environment modifies this outcome, by acting on the system state before the measurement. In practice, this process is described by a quantum channel, i.e. a linear CPTP operation Φ that maps the ideal, i.e. noiseless, state into a noisy one

$$\rho \to \Phi(\rho) \,. \tag{5.1}$$

In terms of expectation values, the quantum channels introduce noise in the measurement outcome as

$$\langle O \rangle_\rho \to \langle O \rangle_{\Phi(\rho)} = \mathrm{Tr}[O\Phi(\rho)] \,. \tag{5.2}$$

We summarize this setup in Fig. 5.1. Different noise models include the dephasing, the depolarizing and the amplitude damping channels [82]. In this section, we discuss the reconstruction of the ideal outcome from the noisy expectation value, i.e. the inversion of Eq. (5.2). In general, quantum channels cannot be physically inverted [187], i.e. it is not always possible to find a CPTP map Θ such that $\Theta \circ \Phi = \mathbb{1}$. We discuss a method called noise deconvolution, which provides the ideal expectation value of arbitrary operators, by post-processing the data obtained from the noisy measurement. This protocol applies to any mathematically invertible noise model, even when its inverse channel is not physically implementable. Consider the channel Φ and its adjoint Φ^*, i.e. the map satisfying

$$\mathrm{Tr}\big[A^\dagger \Phi(B)\big] = \mathrm{Tr}\big[(\Phi^*(A))^\dagger B\big] \,, \tag{5.3}$$

for any operators A and B. Equivalently, given the Kraus operators $\{K_i\}$, a channel and its adjoint respectively read

$$\Phi(B) = \sum_i K_i B K_i^\dagger \,, \tag{5.4}$$

$$\Phi^*(A) = \sum_i K_i^\dagger A K_i \,. \tag{5.5}$$

The inverse map satisfies $\Phi^{-1} \circ \Phi = \mathbb{1}$, yielding

$$\begin{gathered} \langle O \rangle_\rho = \mathrm{Tr}\big[O(\Phi^{-1} \circ \Phi)(\rho)\big] = \mathrm{Tr}\big[\widetilde{O}\Phi(\rho)\big] \,, \\ \text{with } \widetilde{O} = (\Phi^{-1})^*(O) \,, \end{gathered} \tag{5.6}$$

where, on the right-hand side, we used the definition of adjoint map. By substituting back the definition of expectation value, it follows that

$$\langle O \rangle_\rho = \big\langle (\Phi^{-1})^*(O) \big\rangle_{\Phi(\rho)} \,. \tag{5.7}$$

This expression, originally derived in [186], yields the ideal, i.e. noiseless, expectation value of an arbitrary observable (or even of a non-observable operator) by evaluating, instead, the inverted adjoint map over the noisy state. In other proposals

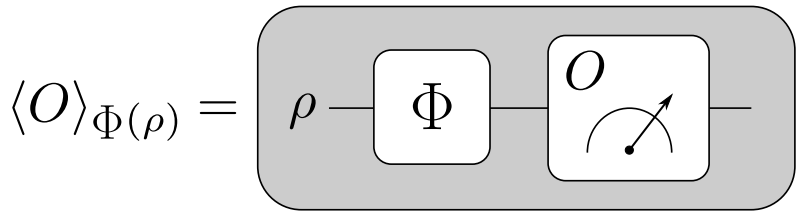


Figure 5.1: Expectation value of an observable O affected by measurement noise Φ. Figure adapted from [4].

[188], similar inversions are implemented physically (often only approximately) by modifying the original channel. Instead, Eq. (5.7) uses the noisy data to calculate the noiseless value entirely and solely at the data processing stage. The above discussion is independent of the system dimension and of the choice of basis. In the next section, we specialize our discussion to multiqubit systems in the Pauli operator basis.

5.2 Deconvolution using the Pauli transfer matrix

We discuss the deconvolution technique for a system of n qubits, using the Pauli transfer matrix (PTM) representation of the channel, introduced in the previous chapter. Consider the Pauli operator basis $\{\mathcal{P}_k\}$, expressed in the vectorized representation

$$\mathcal{B} = \left\{ |k\rangle\rangle \text{ with } k = 0, 1, 2, 3, \ldots, d^2 - 1 \right\} , \tag{5.8}$$

with $d = 2^n$. The PTM of Φ is the $d^2 \times d^2$ matrix

$$\begin{aligned} \Gamma &= \sum_{ij} \Gamma_{ij} \, |i\rangle\rangle\langle\langle j| \ , \\ \text{with } \Gamma_{ij} &= \frac{1}{d} \operatorname{Tr}\left[\mathcal{P}_i \Phi(\mathcal{P}_j)\right] , \end{aligned} \tag{5.9}$$

For a generic operator A, the action of the channel becomes a standard matrix multiplication, namely

$$\Gamma_\Phi \, |A\rangle\rangle = \sum_{ij} \Gamma_{ij} A_j \, |i\rangle\rangle \ . \tag{5.10}$$

In such formalism, the adjoint channel undergoes the following definition

$$\begin{aligned} \langle\langle A|\Phi(B)\rangle\rangle &= \langle\langle \Phi^*(A)|B\rangle\rangle \ , \\ \text{with } \langle\langle A|B\rangle\rangle &= \frac{1}{d} \operatorname{Tr}\left[A^\dagger B\right] . \end{aligned} \tag{5.11}$$

Hence, the adjoint PTM, denoted Γ^*, is precisely the Hermitian conjugate of Γ. In this representation, the deconvolution formula reads as

$$\langle\langle \rho | O \rangle\rangle = \langle\langle \Phi(\rho)| \, {\Gamma^*}^{-1} \, |O\rangle\rangle \ . \tag{5.12}$$

By expanding the observable in the Pauli operator basis

$$\begin{aligned} |O\rangle\rangle &= \sum_k O_k |k\rangle\rangle \ , \\ \text{with } O_k &= \frac{1}{d} \mathrm{Tr}[\mathcal{P}_k O] \end{aligned} \tag{5.13}$$

the deconvolution formula simplifies to

$$\langle O \rangle_\rho = \sum_{ij} \left(\Gamma^{*^{-1}}\right)_{ij} O_j \langle \mathcal{P}_i \rangle_{\Phi(\rho)} \ . \tag{5.14}$$

Here, the noiseless outcome is obtained at the post-processing stage, by multiplying the expectation values of the Pauli components of O, measured in the noisy state $\Phi(\rho)$, by the inverse adjoint PTM of the channel. Eq. (5.14) only requires that Γ is mathematically invertible. For a Pauli observable $O = \mathcal{P}_k$, the above expression becomes

$$\langle \mathcal{P}_k \rangle_\rho = \sum_i \left(\Gamma^{*^{-1}}\right)_{ik} \langle \mathcal{P}_i \rangle_{\Phi(\rho)} \ . \tag{5.15}$$

In general, the deconvolution is computationally expensive, since Eq. (5.14) requires the complete inversion of the channel, even for simple observables. For Pauli channels, however, the deconvolution definitely simplifies: it is achievable without completely inverting the PTM, depending on the complexity of the observable. Consider the Kraus representation of a n-qubit Pauli channel [166], which reads

$$\begin{aligned} \Phi(\rho) &= \sum_i \beta_i \mathcal{P}_i \rho \mathcal{P}_i \ , \\ \text{with } \beta_i \geq 0 \ \forall i \text{ and } &\sum_i \beta_i = 1 \ . \end{aligned} \tag{5.16}$$

This map belongs to the class of random unitary channels [189], which consist in randomly applying the ith Kraus operator with probability β_i. In the Pauli basis, this produces subsequent depolarizing contractions of the Bloch hypersphere, whose intensity and direction depend on the channel components [190]. The PTM of a Pauli channel is diagonal, namely

$$\begin{aligned} \Gamma &= \sum_i \lambda_i |i\rangle\rangle\langle\langle i| \ , \\ \text{with } \lambda_i &= \langle\langle i| \Gamma |i\rangle\rangle \ . \end{aligned} \tag{5.17}$$

Assume that Γ is mathematically invertible, i.e. that $\lambda_i \neq 0 \ \forall j$. As a counterexample, the completely depolarizing channel is not invertible when $p = 1$, for which the Bloch hypersphere collapses to a single point. Since all the Kraus operators are Hermitian, the adjoint of the Pauli channel is the channel itself, which implies that $\Gamma^{*^{-1}} = \Gamma^{-1}$. Moreover, the inverse PTM is diagonal too, yielding

$$\begin{aligned} \Gamma^{*^{-1}} &= \sum_i \frac{1}{\lambda_i} |i\rangle\rangle\langle\langle i| \ , \\ \text{with } \Gamma^{-1}_{ii} |i\rangle\rangle &= \frac{1}{\lambda_i} |i\rangle\rangle \ . \end{aligned} \tag{5.18}$$

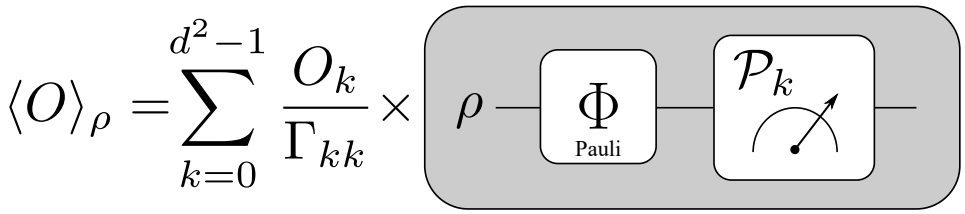


Figure 5.2: Expectation value of an observable O affected by Pauli noise Φ. Deconvolution is achieved by rescaling the noisy data with the corresponding diagonal entry of the Pauli transfer matrix of the channel. Figure adapted from [4].

Consider a generic observable O, expanded in the Pauli operator basis as in Eq. (5.13). Then, the deconvolution formula simplifies to

$$\langle O \rangle_\rho = \sum_k \frac{d O_k \langle \mathcal{P}_k \rangle_{\Phi(\rho)}}{\mathrm{Tr}[\mathcal{P}_k \Phi(\mathcal{P}_k)]} \,. \tag{5.19}$$

For Pauli channels, each component of the observable is rescaled by the corresponding diagonal element of the inverse PTM. Hence, the ideal noiseless outcome (expected over ρ) is obtained by processing the noisy expectation values (measured on $\Phi(\rho)$) of only and only those Pauli basis elements that contribute to the expansion of O. This can lead to further simplifications. For example, consider a Pauli observable, which in the vectorized notation corresponds to a column vector with only one non-zero component, i.e. $|O\rangle\rangle = |k\rangle\rangle$. Then, $\langle\langle \rho | k \rangle\rangle = \lambda_k^{-1} \langle\langle \Phi(\rho) | k \rangle\rangle$, which in standard non-vectorized notation reads

$$\langle \mathcal{P}_k \rangle_\rho = \frac{d \langle \mathcal{P}_k \rangle_{\Phi(\rho)}}{\mathrm{Tr}[\mathcal{P}_k \Phi(\mathcal{P}_k)]} \,. \tag{5.20}$$

This states that the deconvolution of a Pauli observable is always a single rescaling of the noisy outcome, which depends on the kth diagonal element of the PTM diagonal. We summarize these results in Fig. 5.2.

Our protocol works for any number of qubits. Even in the presence of correlations, it requires only local measurements of each element of the operator basis. For Pauli channels, the resource complexity, i.e. the number of PTM elements to be computed, scales with the number of non-zero components of the expansion of O on the operator basis. When O is a Pauli observable, the deconvolution always requires a single measurement and one PTM entry, independently of the number of qubits. On the other hand, when O is a generic observable, the deconvolution requires a measurement on the entire basis and the computation of all the d^2 diagonal PTM entries. In the worst-case scenario, it considerably reduces the number of PTM entries needed over the inversion-based method of [186], which always requires d^4 operations. This is a quadratic speedup. When noise cannot be modelled by a Pauli channel, as it often occurs in real devices, the speedup is not immediately achievable. However, generic noise models are usually split in terms of Pauli and non-Pauli (but simpler) contributions. See for example [186], where the decoherence of Rigetti Aspen-9 is as a composition of a dephasing (Pauli) and an amplitude damping (non-Pauli) channel.

In the vectorized representation, the composition of multiple channels reduces to the product of their PTM. Even in this case, their deconvolution can be treated separately, with still a quadratic advantage on the Pauli terms, which then acts as a rescaling of the non-Pauli rows or columns. Alternatively, twirling techniques can be employed to remove the off-diagonal elements of the PTM [166, 191], which effectively map the original channel to a Pauli one.

So far, we have assumed that the Pauli channel is known a priori. This assumption simplifies the analysis, but it is not necessary: we can still estimate the PTM using the direct reconstruction technique discussed in the previous chapter [3]. In this case, the number of tomographic resources, i.e. the number of experimental configurations needed, depends on the number of components of the observable O, expanded in the Pauli operator basis. In Algorithm 5.1, we show how to combine the two methods, providing a parallel characterization of the PTM. For non-diagonal PTMs, the deconvolution requires the full knowledge of the channel, which can be obtained through standard quantum process tomography.

Algorithm 5.1 Channel deconvolution

Input: Observable O **Optional:** Pauli channel Φ

Result: Expectation value

get $\{O_k$ with $k = 0, 1, 2, 3, \ldots, d^2 - 1\}$ ▷ See Eq. (5.13)
for k such that $O_k \neq 0$ **do**
    measure $\langle \mathcal{P}_k \rangle_{\Phi(\rho)}$
    if Φ is known **then**
        get Γ_{kk} ▷ See Eq. (5.9)
    else
        prepare $\rho_k = (\mathbb{1} + \mathcal{P}_k)/d$
        measure $\langle \mathcal{P}_k \rangle_{\Phi(\rho_k)}$
        get Γ_{kk} ▷ See DPTM
compute $\langle O \rangle_\rho$ ▷ See Eq. (5.19)

In the next section, we discuss the deconvolution for several examples of Pauli and non-Pauli channels. We also take into account correlations, by demonstrating our method on a specific class of maps, in which a parameter $\mu \in [0, 1]$ measures the amount of memory of the channel [143, 170–175].

5.2.1 Example with full Kraus decomposition

Before taking advantage of the PTM representation, we illustrate the deconvolution of a specific class of two-qubit correlated channels, following the method illustrated in [186]. This approach consists in an explicit calculation of the Kraus decomposition of the inverse channel. In the next sections, we show some examples on how Eq. (5.19) can simplify such task, bypassing the Kraus decomposition, independently of the number of qubits. Consider the Kraus representation of a two-qubit (bit, bit-

phase, phase) flip correlated channel

$$\Phi_k(\rho) = \sum_{\alpha\beta\in\{0,k\}} p_{\alpha\beta} K_{\alpha\beta}\rho K_{\alpha\beta}^{\dagger} \ , \quad \text{with } K_{\alpha\beta} = \sigma_\alpha \otimes \sigma_b \text{ and } k \in \{1,2,3\} \ . \tag{5.21}$$

Here, different values of k yield a different Pauli generator of the noise, e.g. the bit flip ($k=1$), the bit-phase flip ($k=2$) and the phase flip ($k=3$) channels. The noise process is parametrized by the coefficients $p_{\alpha\beta}$, which the condition

$$\sum_{\alpha\beta} p_{\alpha\beta} = 1 \ . \tag{5.22}$$

Correlations are described by the Markov chain

$$p_{\alpha\beta} = p_\alpha p_{\beta|\alpha} \ , \tag{5.23}$$

$$p_{\beta|\alpha} = (1-\mu)p_\beta + \mu\delta_{\alpha\beta} \ . \tag{5.24}$$

The parameter μ represents the degree of correlation between the two qubits. For $k=1$, the bit-flip correlated channel reads

$$\Phi_1(\rho) = p_{00}\mathcal{P}_0\rho\mathcal{P}_0 + p_{01}\mathcal{P}_1\rho\mathcal{P}_1 + p_{10}\mathcal{P}_4\rho\mathcal{P}_4 + p_{11}\mathcal{P}_5\rho\mathcal{P}_5 \ . \tag{5.25}$$

For simplicity, assume $p_1 = p$ and $p_0 = 1-p$. We compute the PTM of the channel and write the result as a block matrix. Namely

$$\Gamma_1 = \begin{pmatrix} M_1 & 0 & 0 & 0 \\ 0 & M_1 & 0 & 0 \\ 0 & 0 & Q_1 & 0 \\ 0 & 0 & 0 & Q_1 \end{pmatrix} . \tag{5.26}$$

Here

$$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & q_1 & 0 \\ 0 & 0 & 0 & q_1 \end{pmatrix} , \quad Q_1 = \begin{pmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_1 & 0 & 0 \\ 0 & 0 & q_2 & 0 \\ 0 & 0 & 0 & q_2 \end{pmatrix} , \tag{5.27}$$

where

$$q_1 = 1-2p \ , \tag{5.28}$$

$$q_2 = 4(p-p^2)(\mu-1)+1 \ . \tag{5.29}$$

Its inverse is obtained by taking the reciprocal of each diagonal element. Besides the coefficient values, the inverse PTM has the same form of the original one. Then, the inverse adjoint map shares the same generator of Φ. Indeed, we get

$$(\Phi_1^{-1})^*(O) = q_{00}\mathcal{P}_0 O\mathcal{P}_0 + q_{01}\mathcal{P}_1 O\mathcal{P}_1 + q_{10}\mathcal{P}_4 O\mathcal{P}_4 + q_{11}\mathcal{P}_5 O\mathcal{P}_5 \ , \tag{5.30}$$

with

$$q_{00} = \frac{1}{4q_2} + \frac{1}{2q_1} + \frac{1}{4}\,, \tag{5.31}$$

$$q_{01} = q_{10} = \frac{1}{4} - \frac{1}{4q_2}\,, \tag{5.32}$$

$$q_{11} = \frac{1}{4q_2} - \frac{1}{2q_1} + \frac{1}{4}\,. \tag{5.33}$$

We perform the same computation for the bit-phase ($k = 2$) and phase ($k = 3$) flip correlated channels, getting

$$\Phi_2(O) = p_{00}\mathcal{P}_0 O\mathcal{P}_0 + p_{02}\mathcal{P}_2 O\mathcal{P}_2 + p_{20}\mathcal{P}_8 O\mathcal{P}_8 + p_{22}\mathcal{P}_{10} O\mathcal{P}_{10}\,, \tag{5.34}$$

$$\Phi_3(O) = p_{00}\mathcal{P}_0 O\mathcal{P}_0 + p_{03}\mathcal{P}_3 O\mathcal{P}_3 + p_{30}\mathcal{P}_{12} O\mathcal{P}_{12} + p_{33}\mathcal{P}_{15} O\mathcal{P}_{15}\,. \tag{5.35}$$

With $p_2 = p_3 = p$ and $p_0 = 1-p$, the PTMs respectively read

$$\Gamma_2 = \begin{pmatrix} M_2 & 0 & 0 & 0 \\ 0 & Q_2 & 0 & 0 \\ 0 & 0 & M_2 & 0 \\ 0 & 0 & 0 & Q_2 \end{pmatrix}\,,\quad \Gamma_3 = \begin{pmatrix} M_3 & 0 & 0 & 0 \\ 0 & Q_3 & 0 & 0 \\ 0 & 0 & Q_3 & 0 \\ 0 & 0 & 0 & M_3 \end{pmatrix}\,, \tag{5.36}$$

where

$$M_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & q_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & q_1 \end{pmatrix}\,,\quad Q_2 = \begin{pmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_1 & 0 \\ 0 & 0 & 0 & q_2 \end{pmatrix}\,, \tag{5.37}$$

and

$$M_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & q_1 & 0 & 0 \\ 0 & 0 & q_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\,,\quad Q_3 = \begin{pmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_2 & 0 \\ 0 & 0 & 0 & q_1 \end{pmatrix}\,. \tag{5.38}$$

Similarly, we recover the Kraus representation of the inverted-adjoint maps, yielding

$$(\Phi_2^{-1})^*(O) = q_{00}\mathcal{P}_0 O\mathcal{P}_0 + q_{01}\mathcal{P}_2 O\mathcal{P}_2 + q_{10}\mathcal{P}_8 O\mathcal{P}_8 + q_{11}\mathcal{P}_{10} O\mathcal{P}_{10}\,, \tag{5.39}$$

$$(\Phi_3^{-1})^*(O) = q_{00}\mathcal{P}_0 O\mathcal{P}_0 + q_{01}\mathcal{P}_3 O\mathcal{P}_3 + q_{10}\mathcal{P}_{12} O\mathcal{P}_{12} + q_{11}\mathcal{P}_{15} O\mathcal{P}_{15}\,. \tag{5.40}$$

Consider the set of Pauli observables $\{\sigma_i \otimes \sigma_j \text{ with } i,j = 1,2,3\}$. Before explicitly discussing the deconvolution of the observables, we review some limit cases that we expect to retrieve in our final results. From to the CPTP condition, it follows that $\Phi_k(\sigma_k \otimes \sigma_k) = \sigma_k \otimes \sigma_k$, i.e. that the noise does not affect the measurement outcome along its own direction. When the two qubits are completely correlated, the noise has no effect on *any* measurement of the Pauli matrices. Namely, when $\mu = 1$ then $p_{01} = 0$. This, together with the normalization condition, ensures that $\Phi_k(\sigma_i \otimes \sigma_j) = \sigma_i \otimes \sigma_j$. When the two qubits are uncorrelated, the noise factorizes in terms of its single qubit

contributions [186]. Namely, for $\mu = 0$ we get $\Phi_k(\sigma_i \otimes \sigma_j) = (1-2p)^2 \sigma_i \otimes \sigma_j$. For the expectation value of the observable $\sigma_i \otimes \sigma_j$, the deconvolution formula reads

$$\langle \sigma_i \otimes \sigma_j \rangle_\rho = \langle (\Phi_k^{-1})^* (\sigma_i \otimes \sigma_j) \rangle_{\Phi_k(\rho)} \,. \tag{5.41}$$

Since

$$\sigma_k \sigma_i \sigma_k = (2\delta_{ki} - 1)\sigma_i \,, \tag{5.42}$$

we obtain

$$\begin{aligned} \langle \sigma_i \otimes \sigma_j \rangle_\rho = \big[& q_{00} + q_{01}(2\delta_{kj} - 1) + q_{10}(2\delta_{ki} - 1) \\ & + q_{11}(2\delta_{ki} - 1)(2\delta_{kj} - 1) \big] \langle \sigma_i \otimes \sigma_j \rangle_{\Phi_k(\rho)} \,. \end{aligned} \tag{5.43}$$

Notice that the deconvolution is a rescaling of the noisy outcome. In particular, consider the observable $\sigma_3 \otimes \sigma_3$ affected by a bit-flip channel, i.e. for $k = 1$. We get

$$\langle \sigma_3 \otimes \sigma_3 \rangle_\rho = \frac{\langle \sigma_3 \otimes \sigma_3 \rangle_{\Phi(\rho)_1}}{4(p - p^2)(\mu - 1) + 1} \,. \tag{5.44}$$

These results correctly reproduce the above limit cases, for uncorrelated ($\mu = 0$) and completely correlated ($\mu = 1$) channels. In the next section, we perform the same task using a broader class of channels, by leveraging their PTM representation instead of the full Kraus decomposition of their inverses.

5.2.2 Example with Pauli correlated channels

We consider the class of Pauli correlated channels, for a system of n qubits. In general, they admit the following Kraus representation

$$\begin{aligned} \Phi(\rho) &= \sum_\alpha^3 p_\alpha K_\alpha \rho K_\alpha^\dagger \,, \\ \text{with } K_\alpha &= \sigma_{\alpha_1} \otimes \sigma_{\alpha_2} \otimes \ldots \otimes \sigma_{\alpha_n} \end{aligned} \tag{5.45}$$

with multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. Correlations are modelled through the Markov chain

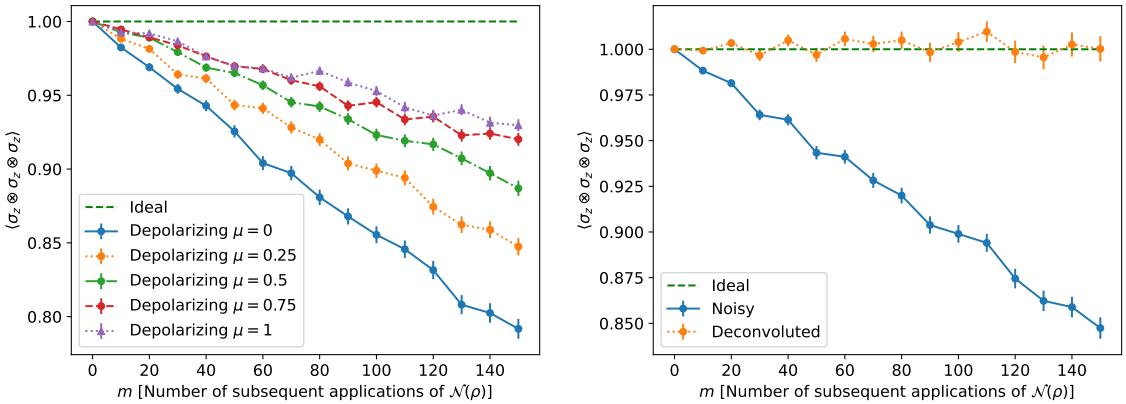
$$p_\alpha = p_{\alpha_1} p_{\alpha_2|\alpha_1} \cdots p_{\alpha_n|\alpha_{n-1}} \,, \tag{5.46}$$

$$p_{\alpha_j|\alpha_i} = (1-\mu) p_{\alpha_j} + \mu \delta_{\alpha_i \alpha_j} \,, \tag{5.47}$$

where

$$\begin{aligned} \mathbf{p} &= [1-p, p_x, p_y, p_z]^T \,, \\ \text{with } p &= p_x + p_y + p_z \,. \end{aligned} \tag{5.48}$$

This type of correlation provides only an example: it is not a requirement of our deconvolution technique. Here, μ represents the degree of correlation between couple of qubits. On the one hand, $\mu = 0$ represents a memoryless channel, that is when the qubits are completely uncorrelated. On the other hand, $\mu = 1$ describes a channel with perfect memory, i.e. with completely correlated qubits. By a proper choice of $\mathbf{p}$, we can use Eq. (5.45) to introduce correlations in processes such as the bit-flip, the



(a) (b)

Figure 5.3: (a) Depolarizing correlated channel, with $n = 3$, degree of correlations μ and probability $p = 0.00052$. Simulation performed using Qiskit Aer for 8192 shots, plotted with respect to the number m of subsequent applications of the noise map. The dashed horizontal line represents the ideal, i.e. noiseless outcome. (a) Noisy outcome plotted for different values of μ. (b) Deconvolution for $\mu = 0.25$. Figure adapted from [4].

dephasing and the depolarizing noise. For example, the bit-flip correlated channel corresponds to

$$\mathbf{p} = [1-p, p, 0, 0]^T \,, \tag{5.49}$$

while the depolarizing channel can be obtained by choosing

$$\mathbf{p} = [1-p, p/3, p/3, p/3]^T \,. \tag{5.50}$$

Consider the observable $O = \sigma_z \otimes \ldots \otimes \sigma_z$, affected by Pauli correlated noise Φ. In this case, only one entry of the PTM is needed to recover the noiseless expectation value. Indeed, the deconvolution formula yields

$$\left\langle \sigma_z^{\otimes n} \right\rangle_\rho = f_n(p,\mu) \left\langle \sigma_z^{\otimes n} \right\rangle_{\Phi(\rho)} \,, \tag{5.51}$$

where the reconstruction factor reads

$$f_n(p,\mu) = \frac{2^n}{\mathrm{Tr}\left[\sigma_z^{\otimes n} \Phi(\sigma_z^{\otimes n})\right]} \,. \tag{5.52}$$

We explicitly compute the reconstruction factor for different number of qubits. For a bit-flip correlated channel, $n = 1, 2, 3, 4$ gives

$$f_1(p,\mu) = \frac{1}{1-2p}\,, \tag{5.53}$$

$$f_2(p,\mu) = \frac{1}{1+4(\mu-1)(1-p)p}\,, \tag{5.54}$$

$$f_3(p,\mu) = \frac{1}{(1-2p)[1+4(\mu-1)^2(p-1)p]}\,, \tag{5.55}$$

$$f_4(p,\mu) = \frac{1}{4(\mu-1)(1-p)p[\mu^2+4(\mu-1)^2p^2-4(\mu-1)^2p-\mu+2]+1}\,. \tag{5.56}$$

For a depolarizing correlated channel, the same computation yields

$$f_1(p,\mu) = \frac{1}{1-p}\,, \tag{5.57}$$

$$f_2(p,\mu) = \frac{1}{1+(\mu-1)(2-p)p}\,, \tag{5.58}$$

$$f_3(p,\mu) = \frac{1}{(1-p)[1+(\mu-1)^2(p-2)p]}\,, \tag{5.59}$$

$$f_4(p,\mu) = \frac{1}{(\mu-1)(2-p)p[\mu^2+(\mu-1)^2p^2-2(\mu-1)^2p-\mu+2]+1}\,. \tag{5.60}$$

For completely uncorrelated qubits ($\mu = 0$), the deconvolution factorizes in terms of its single-qubit contributions. When the qubits are completely correlated ($\mu = 1$), the noise has no effect for n even, while it is corrected by a single-qubit term for n odd. Consider $n = 3$ and the system initialized in $\rho = |000\rangle\langle 000|$. In Fig. 5.3, we plot the effect of the depolarizing correlated channel, with respect to the number of subsequent applications of the noise map. When the correlation among qubits increases, so does the noisy intensity. Then, we simulate the deconvolution technique, successfully reproducing the noiseless expectation value.

5.2.3 Example with non-Pauli channels

For discussing non-Pauli channels, we consider a two-qubit amplitude damping correlated channel. For $n = 1$, the amplitude damping noise is defined by the Kraus operators

$$E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{p} \end{pmatrix}\,, \tag{5.61}$$

$$E_1 = \begin{pmatrix} 0 & \sqrt{1-p} \\ 0 & 0 \end{pmatrix}, \tag{5.62}$$

where $1-p$ represents the probability that the system loses a qubit, e.g. by emitting a photon [82], and it plays the role of channel transmissivity, e.g. in the case of optical fibers. When $n = 2$, the amplitude damping correlated channel can be obtained as a

convex combination of a memoryless amplitude damping channel with a memoryful one

$$\begin{aligned}\Phi(\rho) &= (1-\mu)\Phi^{(0)}(\rho) + \mu\Phi^{(1)}(\rho)\,, \\ \text{with } \Phi^{(\ell)}(\rho) &= \sum_{j=0}^{3} K_j^{(\ell)} \rho K_j^{(\ell)\dagger}\,,\end{aligned} \tag{5.63}$$

and $\ell \in \{0,1\}$. For the memoryless and memoryful components, the Kraus operators read

$$K_0^{(0)} = E_0 \otimes E_0\,,\ K_1^{(0)} = E_0 \otimes E_1\,,\ K_2^{(0)} = E_1 \otimes E_0\,,\ K_3^{(0)} = E_1 \otimes E_1\,, \tag{5.64}$$

$$K_0^{(1)} = |0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2| + \sqrt{p}\,|3\rangle\langle 3|\,, \tag{5.65}$$

$$K_1^{(1)} = \sqrt{1-p}\,|0\rangle\langle 3|\,. \tag{5.66}$$

Consider the set of Pauli observables $\{\sigma_x \otimes \sigma_x,\ \sigma_y \otimes \sigma_y,\ \sigma_z \otimes \sigma_z\}$, of which we measure the expectation value on the noisy state $\Phi(\rho)$. In this case, computing the noiseless expectation value requires the complete knowledge and the inversion of the channel PTM. Then, the deconvolution formula yields

$$\begin{aligned}\langle \sigma_x \otimes \sigma_x \rangle_\rho = f(p,\mu)\big\{&\mu(\sqrt{p}-1)\langle \sigma_y \otimes \sigma_y \rangle_{\Phi(\rho)} \\ &+ [2p(1-\mu) + \mu(\sqrt{p}+1)]\langle \sigma_x \otimes \sigma_x \rangle_{\Phi(\rho)}\big\}\,,\end{aligned} \tag{5.67}$$

Notice that, in contrast to what happens with Pauli channels, the deconvolution of a Pauli observable is not a simple rescaling of the noisy outcome. The same computation yields

$$\begin{aligned}\langle \sigma_z \otimes \sigma_z \rangle_\rho \ &= g(p,\mu)\big[(\mu-1)^2(p-1)^2 \\ &+ \langle \sigma_z \otimes \sigma_z \rangle_{\Phi(\rho)} - (\mu-1)(p-1)\langle \mathbb{1} \otimes \sigma_z + \sigma_z \otimes \mathbb{1} \rangle_{\Phi(\rho)}\big]\,.\end{aligned} \tag{5.68}$$

Analogously, $\langle \sigma_y \otimes \sigma_y \rangle_\rho$ can obtained by exchanging the indexes $x \longleftrightarrow y$ in Eq. (5.67). Here

$$f(p,\mu) = \frac{1}{2\big[\mu\big(p-\sqrt{p}\big)-p\big][\mu(p-1)-p]}\,, \tag{5.69}$$

$$g(p,\mu) = \frac{1}{[p+\mu(1-p)]^2}\,. \tag{5.70}$$

When the qubits are completely uncorrelated ($\mu = 0$) the noise deconvolution factorizes in terms of the single-qubit contributions (See [186] for the deconvolution of the single-qubit amplitude damping channel, with $p = 1-\gamma$). When the two qubits are completely correlated ($\mu = 1$) the noise has no effect on $\langle \sigma_z \otimes \sigma_z \rangle$.

5.3 Discussion

In this chapter, we discussed a method for the deconvolution of multiqubit noise models, described by mathematically invertible quantum channels. This procedure

returns the ideal expectation value of arbitrary observables from the noisy data. For Pauli channels, our prescription bypasses the inversion of the noise map, as well as its Kraus decomposition, providing the noiseless outcome through a sequence of rescalings of the noisy Pauli measurements: using only few components of the PTM. In this case, the number of entries needed quadratically reduces with respect to the general scenario. Whenever the channel is unknown, this technique can be combined with the direct reconstruction method discussed in the previous chapter [3], bypassing the full quantum process tomography of the system. As shown in the simulation, our technique can be applied to any multiqubit Pauli channel, e.g. the bit-flip or the depolarizing correlated channels. For channels that do not belong to this class, the deconvolution still works with reduced efficiency, requiring the complete computation (and inversion) of the channel PTM. Mathematical invertibility remains a fundamental assumption. It could be bypassed by discussing deconvolution with pseudoinverse PTMs or by projecting the channel into an injective domain, e.g. as in Petz recovery maps [192]. This can overcome the limitations posed by non-singular channels: the deconvolution of a completely depolarizing channels, which collapse the entire Bloch sphere to a single point, represents a greater challenge. Unlike other approaches, our protocol can be passively implemented in post-processing: it does not require further experimental configurations, nor active modifications of the original system. For this reason, its range of applicability includes and surpasses quantum computing. Indeed, it applies to any quantum measurements on noisy states, e.g. it can be used to reverse open quantum dynamics [193, 194].

PART III

QUANTUM PHASE AND TIME

CHAPTER 6

QUANTUM BINARY PHASE ESTIMATION

I don't know half of you half as well as I should like; and I like less than half of you half as well as you deserve.

Bilbo Baggins, The Fellowship of The Ring by John R. R. Tolkien

MEASUREMENTS on quantum systems can be performed directly or by interaction with a probe. Usually, the result of this process is encoded in a phase transformation, whose retrieval provides information by indirect means. Several tasks undergo this description, such as quantum metrology protocols [195–197], which are typically modelled in terms of an interferometric phase shift. Examples include displacement interferometry [198], clock synchronization [199], key distribution [200, 201] and imaging [202, 203]. There are multiple strategies to estimate a quantum phase, including computational approaches [82, 204, 205], interferometry [49, 206–209] and Bayesian methods [210].

In this chapter, we focus on information encoded in phase. We introduce an optical protocol for the estimation of an unknown phase value, by means of a two-mode interferometric setup, whose response function approximates a square wave. In the ideal case, the mode where the photon exits identifies a single bit of the binary expansion of the unknown phase, whose significance is controlled by the response period, i.e. by the number of phase shifts applied to the photon. Combining all the binary outputs provide a full phase estimation at arbitrary uncertainty. This step can be performed sequentially, or in parallel without feedforward propagation. Our strategy does not require training [211–213]. The shape of the response function is determined by a fixed sequence of parameters. This ensures modularity in controlling the square wave approximation, giving the possibility to extend or shorten the interferometer without further optimization. A protocol is considered optimal

when it reaches the Heisenberg limit, i.e. when it undergoes the precision scaling set by the quantum bounds [214–217]. We show that our method is equivalent to a binary search protocol, which undergoes the Heisenberg scaling in the number of phase shifts experienced by each photon. Such scaling is attained immediately, and not asymptotically, since statistics is not required to complete the search. The protocol requires no prior information, i.e. it is a global estimation method.

This chapter is built upon the results of [5]. The underlying code that generated the data for this study is openly available in GitHub [218].

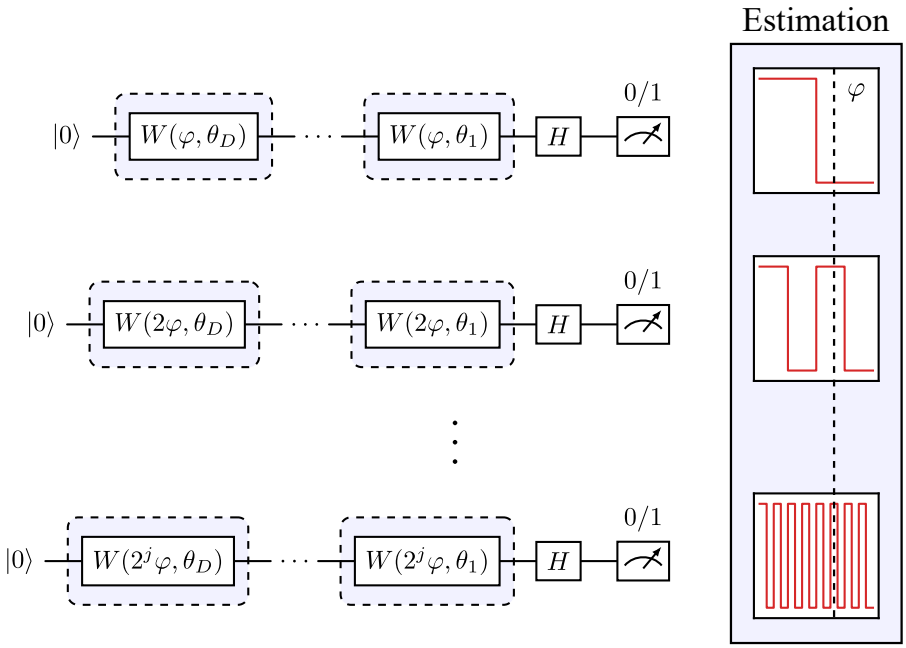


Figure 6.1: Binary phase estimation. Each line represents one photon going through the interferometer and providing a single bit of the binary expansion of the unknown phase φ. At the jth step (line), the interferometer reproduces a square-wave response with period $[0, 2\pi/2^j)$, and Fourier expressivity determined by the number D of subsequent applications of W (whose optical design is reported in Fig. 6.2). The true phase value is sketched by the dashed vertical line. Combining all the outputs completes the estimation, with uncertainty determined by the number of iterations performed. The number of unknown phase transformations, i.e. the resource cost of the protocol, is controlled by the depth of the interferometer and by the number of iterations (which determines the resolution of the binary expansion). See Section 6.3 for a discussion. Figure adapted from [5].

6.1 Square wave interferometer

We consider the problem of estimating an unknown phase $\varphi \in [0, 2\pi]$, encoded by the unitary operator $P(\varphi)$. We combine a set of beam splitters and phase transformations into a two-mode interferometer, which takes single-photon states as input. We do so by appending a sequence of unitary blocks W (called modules), with behaviour controlled by the phase parameters $\{\theta_i\}$. This provides a modular, i.e. extendible, structure. Ideally, the interferometric response function reproduces a square wave. When a photon goes through it, the probability of exiting the first mode yields one bit of the binary expansion of φ. The square wave period, controlled by the number of subsequent applications of $P(\varphi)$, determines which of the bits is returned. In practice, the interferometric response approximates the truncated Fourier series of the square wave. In this case, the truncation order is determined by the number of subsequent applications of W (with parameters changed at each composition). The deeper the interferometer, the richer is the Fourier expressivity. Alternatively, a multiround implementation can be attained, where the truncation order is fixed by the

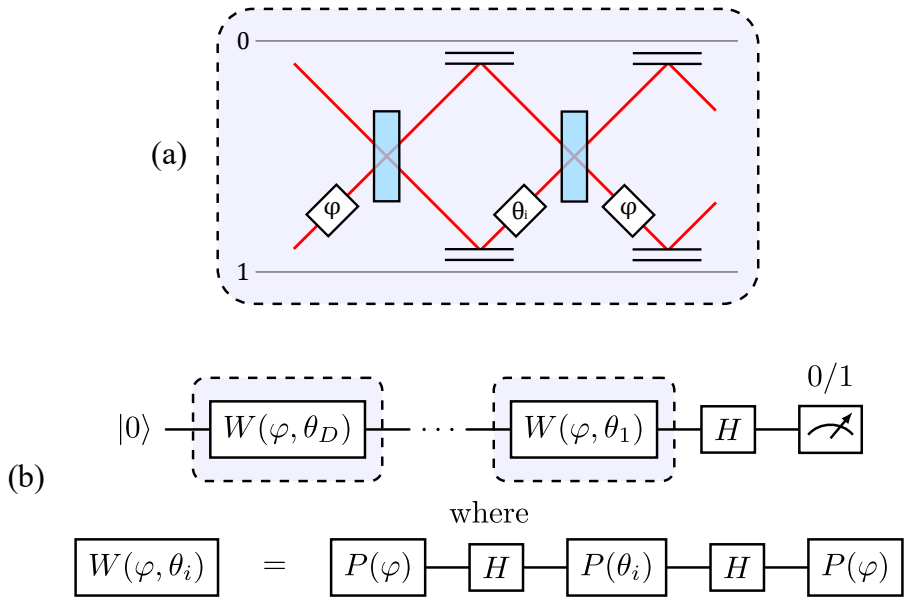


Figure 6.2: Interferometric module. (a) Optical implementation of $W(k_j\varphi,\theta_i)$, for $k_j = 1$. In mode 1, the photon undergoes three phase transformations, each separated by a beam splitter. Subsequent applications of W (with different θ_i) approximate the truncated Fourier series of the square wave. (b) First iteration ($k_j = 1$) after D subsequent applications of W, represented as a single-qubit circuit. The circuit depth determines the Fourier truncation order. Subsequent iterations are reported in Fig. 6.1, for $k_j = 1, 2, \ldots, 2^j$. Figure adapted from [5].

number of rounds a single photon makes on the same module, with phase parameters changed at each passage. See Fig. 6.1, which illustrates the estimation protocol, and Fig. 6.2 for the optical design of the interferometric module.

We mathematically discuss this setup. Consider a binary estimation made of $n \in \mathbb{N}^+$ iterations, each associated with an interferometer, whose response function approximates a square wave with period $[0, 2\pi/k_j)$, where $j \in \{0, \ldots, n-1\}$ and $k_j = 2^j$. We adopt the dual-rail encoding, where photons can only propagate in two orthogonal modes, labelled by the quantum states $|0\rangle$ and $|1\rangle$, i.e. a qubit. The photon is initially injected in mode 0. Consider a fixed set of known phase parameters

$$\begin{aligned} \theta &= \{\theta_i\}_{i\in\mathbb{N}} \,, \\ &\text{with } i \in \{1, \ldots, D\} \,. \end{aligned} \tag{6.1}$$

We define the following operator, which constitutes the building block of our interferometric setup. Namely

$$W(k_j\varphi,\theta_i) = P(k_j\varphi)HP(\theta_i)HP(k_j\varphi) \,, \tag{6.2}$$

where H is the Hadamard gate (implemented by a 50:50 beam splitter), and

$$P(k_j\varphi) = \begin{pmatrix} 1 & 0 \\ 0 & \exp\left(ik_j\varphi\right) \end{pmatrix} , \tag{6.3}$$

a phase shift transformation. Optically, this operator is equivalent to applying three phase shifts to mode 1, each interspersed by a beam splitter [219]. Consider D subsequent applications of W (each with a different parameter θ_i), followed by a final beam splitter, i.e. a Hadamard gate. We measure the expectation value of the projector $\Pi_0 = |0\rangle\langle 0|$, i.e. the ratio of photons p_{jD} in mode 0, yielding

$$\begin{aligned} p_{jD}(\varphi) &= \langle 0| U^\dagger(k_j\varphi,\theta)\Pi_0 U(k_j\varphi,\theta)|0\rangle \ , \\ \text{with } U(k_j\varphi,\theta) &= H\prod_{i=1}^{D} W(k_j\varphi,\theta_i) \ . \end{aligned} \tag{6.4}$$

From $P(k_j\varphi)|0\rangle = |0\rangle$, it follows that

$$U(k_j\varphi,\theta)|0\rangle = HP(k_j\varphi)\widetilde{U}(k_j\varphi,\theta)|0\rangle \ , \tag{6.5}$$

where

$$\begin{aligned} \widetilde{U}(k_j\varphi,\theta) &= L(\theta_1)P(2k_j\varphi)L(\theta_2)\cdots P(2k_j\varphi)L(\theta_D) \ , \\ \text{with } L(\theta_i) &= HP(\theta_i)H \ . \end{aligned} \tag{6.6}$$

By Stone's theorem [220], we write the Phase gate in term of its generator, namely

$$P(2k_j\varphi) = \exp\big(2k_j\varphi\,|1\rangle\langle 1|\big) \ , \tag{6.7}$$

which has spectrum $\{0, k_j\}$. Using the spectral decomposition [221], the interferometric response reads

$$\begin{aligned} p_{jD}(\varphi) &= \sum_{q\in I} c_q(\theta)\exp\big(ik_jq\varphi\big) \ , \\ \text{with } I &\subseteq \{-2D+1, -2D+2, \ldots, 2D-1\} \end{aligned} \tag{6.8}$$

This is a truncated Fourier series, with period $\big[0, 2\pi/k_j\big)$ and coefficients fixed by θ. Notice that the depth of the interferometer determines the truncation point of the series, i.e. the higher D is, the more accurate the approximation of the square wave. We summarize the whole setup in Fig. 6.2, which represents the first iteration of the binary search estimation. The subsequent iterations are obtained by progressively doubling k_j.

6.2 Binary phase search

We combined the bits, obtained at the interferometer output at each iteration $j \in \{0, \ldots, n-1\}$, and perform a binary estimation of the unknown phase φ, with arbitrary uncertainty ε. Consider a sequence of square waves $\{f_j\}_{j\in\mathbb{N}}$, obtained through the periodic extension of the step function

$$f_j(\varphi) = \begin{cases} 1 \text{ for } \varphi \in \left[0, \frac{\pi}{k_j}\right) \\ 0 \text{ for } \varphi \in \left[\frac{\pi}{k_j}, \frac{2\pi}{k_j}\right) \end{cases} \ , \tag{6.9}$$

which, as we show in our simulations below, approximates the interferometer response $p_{jD}(\varphi)$. At each iteration, f_j partitions the interval $[0, 2\pi)$ in 2^{j+1} subsets of size π/k_j, namely

$$\left\{ \left[\frac{m\pi}{k_j}, \frac{(m+1)\pi}{k_j} \right) \text{ with } m = 0, 1, \ldots, k_{j+1} - 1 \right\} . \tag{6.10}$$

Each subset is associated with a classical bit, 0 or 1, which corresponds to the value of $f_j(\varphi)$, and approximated by $p_{jD}(\varphi)$. Combining the bits identifies which subset includes the true value of the unknown phase. In practice, we obtain each bit by measuring the jth ratio of photons in mode 0 at the output of the interferometer. The result is rounded to the nearest binary digit, with threshold at $1/2$. The process is repeated, each time doubling the value of k_j, which subsequently halves the response period as $f_{j+1}(\varphi) = f_j(2\varphi)$. The estimation halts as soon as the partition size becomes double the uncertainty, namely

$$2\varepsilon = \frac{\pi}{k_{n-1}} , \tag{6.11}$$

and the number of iterations reads

$$n = \log_2 \left(\frac{\pi}{\varepsilon} \right) . \tag{6.12}$$

If $n \notin \mathbb{N}^+$ we round it down to the first integer, such that $2\varepsilon = \inf_j \pi/k_j$. Denoting $\widetilde{m}$ the index of the decision outcome, we estimate the unknown phase φ as

$$\widetilde{\varphi} = \frac{(2\widetilde{m}+1)\pi}{2^n} \pm \frac{\pi}{2^n} . \tag{6.13}$$

The full method is outlined in Algorithm 6.1. Alternatively, the above procedure can be parallelized using n interferometers, each set to a different value of k_j. Although the parallel strategy brings no advantage in terms of resource counting, it does not rely on the feedforward propagation of the sequential one.

Algorithm 6.1 Binary phase estimation

Input: Phase φ, uncertainty ε

Output: Estimation $\widetilde{\varphi}$

initialize $b \leftarrow (0, \ldots, 0)$
initialize $j \leftarrow 0$
while $\pi/2^j \geq 2\varepsilon$ **do**
    measure $p_{jD}(\varphi)$ ▷ See Eq. (6.4)
    if $p_{jD}(\varphi) \geq 1/2$ **then** set $b_j \leftarrow 1$
    else set $b_j \leftarrow 0$
    increment $j \leftarrow j + 1$
get $b \leftarrow \mathrm{NOT}(b)$
get $b \leftarrow \mathrm{Reverse}(b)$
convert $\widetilde{m} \leftarrow b_0 2^0 + b_1 2^1 + \ldots + b_{n-1} 2^{n-1}$
estimate $\widetilde{\varphi}$ ▷ See Eq. (6.13)

6.3 Metrological resources and Heisenberg bound

In resource counting, the metrological figure of merit is represented by the number n_p of unknown phase shifts $P(\varphi)$ underwent by each photon [206, 222]. For $k_j = 1$, a single application of W costs 2 resources. Since $P(k_j\varphi)$ corresponds to k_j compositions of single phase shifts, $W(k_j\varphi, \theta_i)$ requires $2k_j$ resources. After D subsequent applications of W, the cost of approximating a binary iteration f_j is $2Dk_j$. The binary search method consists in repeating the above protocol, doubling the value of k_j until condition $\varepsilon = \pi/2^n$ is met. The total cost becomes

$$n_p = 2D\sum_j k_j = 2D\left(\frac{\pi}{\varepsilon} - 1\right), \tag{6.14}$$

where we used that $k_j = 2^j$, Namely, the protocol requires $\mathcal{O}(D\pi\varepsilon^{-1})$ resources, with uncertainty scaling as $\mathcal{O}(n_p^{-1})$. Hence, our strategy reaches optimality with respect to the Heisenberg bound, with uncertainty obtained by averaging the mean squared error over a single interval of the partition. Indeed, consider the ideal response f_j, restricted to a single period of size $2L = 2\pi/k_j$. For simplicity, we neglect the Gibbs fluctuations of the interferometric response function. The mean squared error is defined as

$$\Sigma^2(\widetilde{\varphi}) = \frac{1}{L}\int_0^{2L} \mathrm{d}\varphi\ f_j(\varphi)(\varphi - \widetilde{\varphi})^2\ , \tag{6.15}$$

where we introduced a factor, which guarantees that f_j/L is normalized on $[0, 2L)$. Since f_j vanishes on $[L, 2L)$, this explicitly reads

$$\Sigma^2(\widetilde{\varphi}) = \frac{1}{L}\int_0^{L} \mathrm{d}\varphi\ \left(\varphi^2 - 2\widetilde{\varphi}\varphi + \widetilde{\varphi}^2\right) = \frac{L^2}{3} - L\widetilde{\varphi} + \widetilde{\varphi}^2\ . \tag{6.16}$$

By taking the average on $[0, L]$, i.e. on a single partition created by the square wave, we get

$$\langle\Sigma^2\rangle = \frac{1}{L}\int_0^{L} \mathrm{d}\widetilde{\varphi}\ \Sigma^2(\widetilde{\varphi}) = \frac{L^2}{6}\ , \tag{6.17}$$

At the jth step $2\varepsilon = L$, which yields $\langle\Sigma^2\rangle \simeq n_p^{-2}$. Besides Eq. (6.14), our protocol undergoes the Heisenberg bound also with respect to the mean squared error of the estimator. Notice that this is not automatically guaranteed: an estimation strategy may be Heisenberg-limited in the number of resources while undergoing the standard quantum limit in the variance of the estimator [206] (see also binary elimination [223]). Finally, our results are also connected to the π-corrected Heisenberg bound [216], up to a constant overhead because the square wave is achieved only approximately. Notice that protocols that perform below the Heisenberg bound are usually not attainable, since they would require more prior information, making the whole estimation strategy ineffective [224].

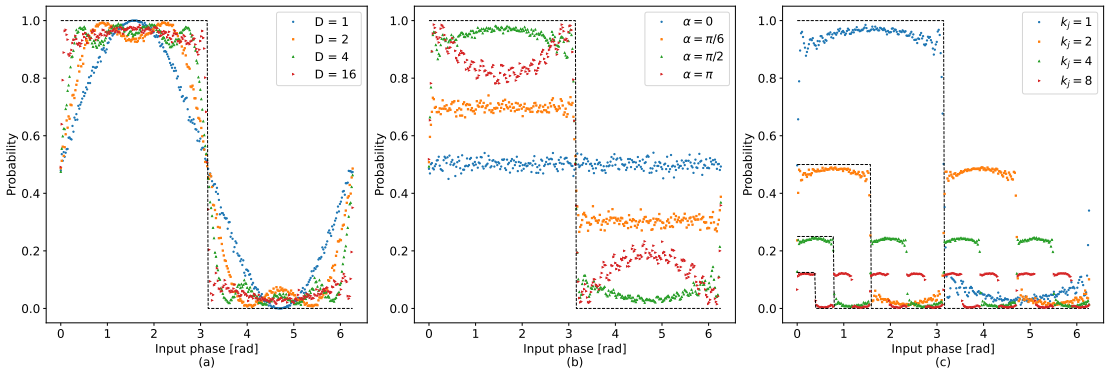


Figure 6.3: Comparison between the square wave and the optical response p_{jD}, obtained as the ratio of photons observed in mode 0. Simulation performed with Qiskit Aer, sampled for 300 evenly-spaced phases $\varphi \in [0, 2\pi)$ and 1000 shots, for different choices of depth D, α and k_j. (a) Simulation performed with $\alpha = \pi/2$, $k_j = 1$, and different depths D. The response function shows a binary behaviour, with approximation accuracy that increases with the interferometer depth. (b) Simulation with $D = 16$ and $k_j = 1$, plotted for different values of α. As α increases, the extrema of p_{jD} approximates 0 and 1, up to a critical value, after which the output departs from the expected behaviour. (c) Response function simulated with $D = 16$ and $\alpha = \pi/2$, plotted for different values of k_j. As k_j increases, the response period is progressively halved, representing different iterations of the binary search. To enhance visualization, each response is renormalized to k_j: all the outputs have true range in $[0, 1]$. Figure adapted from [5].

6.4 Simulations of the interferometric response function

We consider a specific (and fixed) set of parameters that simulates the square wave. We evaluate the interferometer response by simulating the circuit of Fig. 6.2 when

$$\theta_i = \frac{\alpha}{2i-1} \,, \quad \text{with } i \in \{1, \dots, D\} \text{ and } \alpha \in \mathbb{R}^+ \,. \tag{6.18}$$

The results are reported in Fig. 6.3. Under this choice of parameters, the output reproduces a square wave, with accuracy depending on the interferometer depth. Indeed, the identification between the square wave and the interferometric response holds only approximately. The truncated Fourier series of the square wave admits both negative and greater-than-one values. Such values are strictly forbidden at the interferometer output, which is probabilistic, i.e. $p_{jD}(\varphi) \in [0, 1]\ \forall \varphi \in [0, 2\pi)$. This constraint is counterbalanced by α, which can be set to saturate the infimum and supremum of $p_{jD}(\varphi)$ to 0 and 1, respectively. Consider the square wave f_j with period $[0, 2\pi/k_j)$ and range $[0, 1]$. By direct computation [225], the Fourier series of the square wave reads

$$f_j(\varphi) = \frac{1}{2} + \sum_q \frac{2}{\pi q} \sin\left(q k_j \varphi\right) \,, \tag{6.19}$$

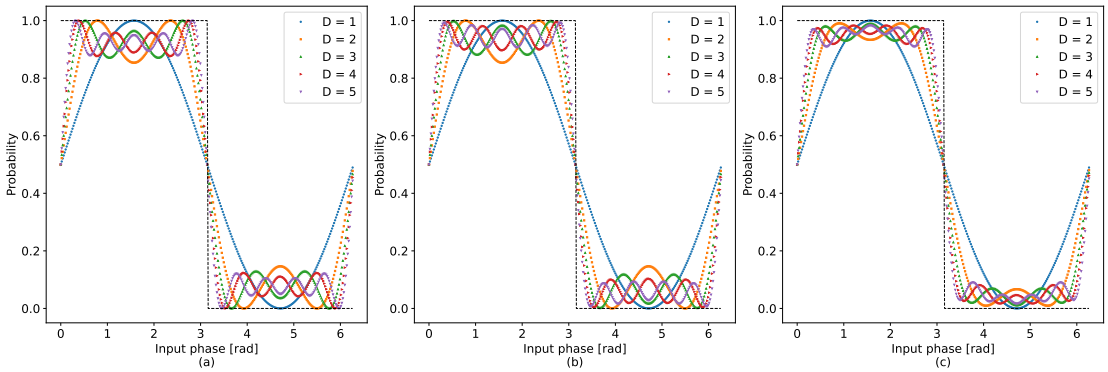


Figure 6.4: Comparison between the square wave f_j and the optical response function p_{jD}. Results obtained by direct numerical computation, for 300 evenly-spaced phases $\varphi \in [0, 2\pi)$, and different truncation orders D. The interferometric response is computed exactly, with no statistical sampling method involved. (a) Truncated Fourier series of f_j, renormalized to the range $[0, 1]$. (b) Interferometer response p_{jD}, evaluated for $\beta = 1$ and $\alpha = \pi/2$. (c) Interferometer response p_{jD}, evaluated for $\beta = 2$ and $\alpha = \pi/2$. Figure adapted from [5].

with $q \in 2\mathbb{N} + 1$. Its first-order truncation yields

$$f_j(\varphi) \simeq \frac{1}{2} + \frac{2}{\pi} \sin\left(k_j \varphi\right) . \tag{6.20}$$

We compare this expression with the interferometer response p_{jD}, computed for $D = 1$. To this extent, consider a broader class of fixed phase parameters

$$\begin{gathered} \theta_i = \alpha/(\beta i + 1 - \beta) , \\ \text{with } \alpha, \beta \in \mathbb{R}^+ \text{ and } i \in \{1, \ldots, D\} , \end{gathered} \tag{6.21}$$

which reduces to Eq. (6.18) for $\beta = 2$. Then

$$\begin{aligned} p_{j1}(\varphi) &= \frac{1}{2} + \frac{1}{8} \operatorname{Re}\left[\exp\left(i k_j \varphi\right)(\exp(-i\alpha) - \exp(i\alpha))\right] \\ &= \frac{1}{2} + \frac{1}{2} \sin\alpha \sin\left(k_j \varphi\right) . \end{aligned} \tag{6.22}$$

Independently of the choice of β, we guarantee that $\max p_{j1}(\varphi) = 1$ and $\min p_{j1}(\varphi) = 0$, by choosing

$$\alpha = \pi/2 \mod 2\pi , \tag{6.23}$$

We numerically evaluate f_j and p_{jD}, for different truncation orders D, and ensure a representative comparison by renormalizing the truncated Fourier series of f_j, so that its range becomes $[0, 1]$. The results are reported in Fig. 6.4, for different values of β. For $\beta = 1$, p_{jD} better approximates f_j. When $\beta = 2$, p_{jD} better saturates its extrema to 0 and 1, while mitigating the Gibbs phenomenon.

6.5 Discussion

We introduced an interferometric setup for the estimation of an unknown phase value. In principle, it can approximate any response function, with accuracy determined by the interferometer depth, which controls the number of Fourier modes that contribute to the response function. In practice, we considered an ad hoc architecture, whose response is a square wave, bypassing the optimization of the interferometer parameters. This choice rephrases the estimation task as a binary search problem, which can estimate the unknown phase at arbitrary uncertainty. Our protocol is compatible with parallel, sequential and multiround strategies. Importantly, it undergoes the Heisenberg scaling in the number of phase transformations, with no prior information required.

As a generalization of our method, it is possible to train the parameters θ to approximate an arbitrary response function, with Fourier expressivity controlled by the circuit architecture [75]. Instead, we focused on a fixed set of parameters. This choice brings several advantages. On the one hand, it bypasses the training procedure, i.e. the additional cost of repeating the protocol for enough optimization epochs. On the other hand, it provides modularity, i.e. it makes possible to extend the interferometer, improving the truncation accuracy without having to reoptimize the former parameters.

CHAPTER 7

QUANTUM ARRIVAL TIME

Duration and extension are attributes of things, but time and space are taken to be outside of things and serve to measure them.

Gottfried W. Leibniz [226]

CLASSICALLY, the time at which an object occupies a specific position can be computed by inverting its trajectory, which is uniquely determined from the dynamics and initial conditions. When the position is that of a detector, the time at which the detector clicks is known as time of flight or arrival time. The computation of the arrival time of a quantum particle is far more complicated. Quantum mechanics requires each observable to be described by a self-adjoint operator, which acts on the Hilbert space of the system, or by a positive-operator valued measure. The quantization of a classical system is usually performed by means of the correspondence principle, i.e. by promoting each classical observable to a self-adjoint operator on the Hilbert space. This procedure cannot be directly employed for time, which is a (scalar) parameter in textbook quantum mechanics. Several reasons lie behind this obstruction, e.g. the Pauli objection [44]: since the energy is responsible for time translations, any time operator must be conjugated to an energy operator whose spectrum then, must be continuous and unbounded. However, the Hamiltonian does not usually satisfy such a condition, preventing the possibility to directly construct a self-adjoint time operator. Nevertheless, experiments usually compute the arrival time using the semiclassical approximation [227], which holds only in regimes where quantum effects can be neglected, namely when the particle is approximated by a well localized wave packet, with negligible spatial uncertainty compared to the detector size [228].

To overcome the limitations of the semiclassical approximation, several quantum time and arrival time formulations have been formulated [229, 230], some within

the framework of standard quantum mechanics, e.g. through canonical quantization [231–233] or advocating different constructions [234–240], others derived from extensions or alternative formulations of the textbook theory [241–248], or using Bohmian mechanics [249–252]. In this chapter, we review and compare some of these proposals, by analysing the interference patterns in the arrival time distribution, for a superposition of Gaussian wave packets in different regimes, schematically represented in Fig. 7.1. We discuss a possible experimental implementation to discriminate between these theories, with the superposition generated by applying Bragg diffraction [253] to a Bose-Einstein condensate trapped in an accelerator ring [254].

This chapter is built upon the results of [6].

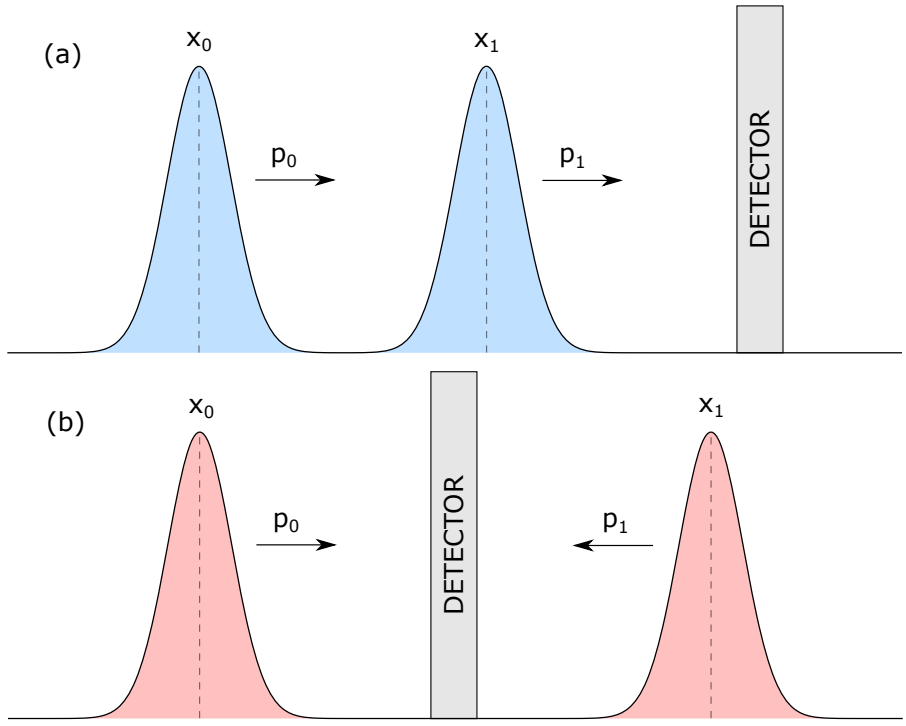


Figure 7.1: Qualitative representation of the regimes considered in our comparison. A particle is initially prepared in a superposition of two Gaussian wave packets. When the packets overlap near the detector, arrival time interference is observed. (a) In this regime, two packets with initial average positions x_0, x_1 cross the detector from the same side, with average momenta p_0, p_1. A limit case is provided by the overtaking condition, in which the packets completely overlap at the detector with the same average arrival time. (b) In this scenario two wave packets cross the detector from opposite directions. We refer to this regime as counterpropagating. Figure adapted from [6].

7.1 Preliminaries on Gaussian packets

Consider a Gaussian wave packet, which we use to evaluate the proposals in different regimes and for different superpositions. Denote x_0 the initial position, p_0 the momentum and σ_0 the standard deviation [255]. The wave function of the packet reads

$$\psi(x,0) = \frac{1}{(\pi\sigma_0^2)^{1/4}} \exp\left[-\frac{(x-x_0)^2}{2\sigma_0^2} + \frac{ip_0(x-x_0)}{\hbar}\right]. \tag{7.1}$$

Its time evolution, controlled by the free particle Hamiltonian, reads

$$\psi(x,t) = A(t) \exp\left[\frac{-(x-x_0-p_0t/m)^2}{2\sigma_0^2\left(1+i\hbar t/m\sigma_0^2\right)}\right] \exp\left[\frac{ip_0}{\hbar}\left(x-x_0-\frac{p_0t}{2m}\right)\right], \tag{7.2}$$

with

$$A(t) = \left[\sqrt{\pi}\left(\sigma_0 + \frac{i\hbar t}{m\sigma_0}\right)\right]^{-1/2}. \tag{7.3}$$

The probability density reads

$$|\psi(x,t)|^2 = \exp\left[\frac{-(x-x_0-p_0t/m)^2}{\sigma_0^2+\hbar^2t^2/m^2\sigma_0^2}\right]\left[\pi\left(\sigma_0^2+\frac{\hbar^2t^2}{m^2\sigma_0^2}\right)\right]^{-1/2}. \tag{7.4}$$

From now on, and unless explicitly stated, we work in units of length, time and energy respectively given by $l_0 = \sqrt{\hbar/m\omega}$, $t_0 = 1/\omega$ and $E_0 = \hbar\omega$, with ω the frequency of the harmonic trap where the particle is initially located, in the ground state. After the preparation, the trap is turned off and the particle is kicked so that its probability density is Eq. (7.4).

7.2 Quantum arrival time distributions

In this section, we review some of proposed arrival time distributions: the semiclassical approximation [227], Kijowski's axiomatic construction [234, 235], as well as the proposals in agreement with it [231–233], the quantum flux [236, 249] and the quantum clock proposals [246, 247]. We denote each arrival time distribution as Π, specifying the corresponding proposal in the subscript.

7.2.1 Semiclassical far-field regime

Experimentally, time measurements are semiclassically treated as a momentum measurement [249], from which the arrival time distribution reads

$$\Pi_S(t) = \frac{mL}{t^2}\left|\hat{\Psi}\left(\frac{mL}{t}\right)\right|^2, \tag{7.5}$$

with $\hat{\Psi}(p)$ denoting Fourier transform of $\Psi(x,0)$, and L the distance between the source of the wave and the detector. This distribution gives sensible results only for those states that possess a classical trajectory interpretation (e.g. Gaussian wave packets), and for which the width of the wave function at $t = 0$ can be considered negligible with respect to the source-detector distance.

7.2.2 Kijowski's axiomatic and geometrical approach

One of the possible solutions were proposed by Kijowski [234, 235], who adopts an axiomatic approach. Kijowski's axioms can be potentially satisfied by infinitely many candidates of arrival time distributions Π. However, there exists a unique choice of Π which best approximates any other possible candidate Π'. This best candidate provides the same expectation value of any Π' and a lower bound to its second moment. Consider a one-dimensional particle with either positive or negative momenta p, and a detector placed in the origin. The particle is described by a wave function Ψ with momentum representation supported for either positive or negative values of p. Respectively, two probability distributions Π_+, Π_- are obtained

$$\Pi_{\pm}(t) = \frac{1}{4\pi^2 m\sqrt{\hbar}} \left| \int_0^{\pm\infty} \mathrm{d}p \; \hat{\Psi}(p,t)\sqrt{|p|} \right|^2 , \tag{7.6}$$

with m the mass of the particle and

$$\hat{\Psi}(p,t) = \exp\left(-\frac{itp^2}{2m\hbar}\right)\hat{\Psi}(p) \, . \tag{7.7}$$

Here, $\hat{\Psi}(p)$ denotes the Fourier transform of the wave function at $t = 0$. Alternatively, Eq. (7.6) can be written in position representation. This result is due to Leavens [256], and reads

$$\Pi_{\pm}(t) = \frac{\hbar}{32\pi m} \left| \int_{-\infty}^{+\infty} \mathrm{d}x \; [\Psi(x,t) - \Psi(0,t)] \frac{1 \pm i\,\mathrm{sign}(x)}{|x|^{3/2}} \right|^2 . \tag{7.8}$$

Independently of the representation, for a generic particle whose momentum is not supported only for positive or negative momenta, Π_+ and Π_- are combined so that the total arrival time distribution yields

$$\Pi_K(t) = \frac{\Pi_+(t) + \Pi_-(t)}{N_K} \, , \tag{7.9}$$

with normalization constant

$$N_K = \int_{-\infty}^{+\infty} \mathrm{d}t \; [\Pi_+(t) + \Pi_-(t)] \, . \tag{7.10}$$

See [257] and references therein. Kijowski's distribution reproduces the same results obtained by Aharonov and Bohm through the quantization of the classical arrival time by the correspondence principle [231]. See [258] for a comparison between these two results. Other proposals reproduce or are at least in agreement with the predictions of Kijowski. For example, in [232] the quantization of the classical arrival time uses a different symmetric ordering, with respect to the approach of Aharonov and Bohm. When considering a particle with either positive or negative momentum, this arrival time distribution has the same form of $\Pi_\pm$. An alternative approach consists

in confining the motion within a given space interval. This idea has been followed by Galapon, deriving the discrete probability distribution of a particle quantized in a finite box [233]. In the limit of an infinite interval, this reproduces $\Pi_{\pm}$. In [257], Leavens discusses some paradoxical behaviours arising from $\Pi_{\pm}$. Indeed, for a particle that propagates towards an infinite potential barrier, the arrival time distribution predicts non-vanishing probabilities even in prohibited regions. In this regard, further comments are given in [259, 260], while other paradoxes and issues are analysed in [258]. Finally, Kijowski's distribution can be obtained by geometrical means, removing the axioms, and defining the arrival time self-adjoint operator as the Lie derivative of the vector field that transports wave functions in the momentum space [261]. In this case, the split into positive and negative momenta signs is induced by the fact the coordinates, with respect to which the above field is constant, do not cover the entire momentum space.

7.2.3 Probability current and quantum flux

A straightforward way to obtain an arrival time distribution starts from the definition of the Schrödinger current operator, or quantum flux, as

$$J = \frac{p}{2m} |x\rangle\langle x| + \frac{1}{2m} |x\rangle\langle x| p \ , \tag{7.11}$$

with m the mass of the particle, x the eigenvalues of the position operator and p the momentum operator and its eigenvalues, with a little abuse of notation. By taking the expectation value of the current, with respect to $\Psi(t)$, it follows that

$$\Pi_F(x,t) = \frac{\hbar}{N_F m} \operatorname{Im}\left[\Psi^*(x,t)\partial_x \Psi(x,t)\right] \ , \tag{7.12}$$

with normalization constant

$$N_F = \int_{-\infty}^{+\infty} \mathrm{d}t \ \Pi_F(t) \ . \tag{7.13}$$

When the detector is placed in the origin, we denote the arrival time distribution as $\Pi_F(t) = \Pi_F(0,t)$. In general, J is not a positive-definite operator, so it cannot be interpreted as a probability current [234]. In [236], Delgado and Muga shows that this interpretation is possible under a specific choice of states. To this extent, they construct a self-adjoint operator whose orthonormal and complete set of eigenstates are well-behaved under time translations, while being conjugated not to the Hamiltonian of the system but to another operator with dimensions of energy. Such energy-like operator is the Hamiltonian multiplied by the sign of the momentum operator. They show that the current is positive-definite only for states with definite momentum sign, thus admitting a probabilistic interpretation. The arrival time distribution is then obtained by projecting such states along the basis of this self-adjoint temporal-like operator. Alternatively, Eq. (7.12) can also be derived in the framework of Bohmian mechanics [228, 249, 258]. In this case, there exist free particle

states which forbid the probabilistic interpretation of the quantum flux, leading to negative probability contributions even for definite momenta signs [249]. This eventuality takes the name of quantum backflow, historically introduced by Bracken and Melloy [262]. In the next sections, we show that even in presence of backflow in the quantum flux, the other proposals take only positive values. Finally, the quantum flux can be also obtained by stochastic means. In this case, the distribution is obtained by applying the method of transformation to position and time, which are treated as random variables [263, 264].

7.2.4 Quantum clocks in the Page-Wooters mechanism

We have considered proposals that treat time as a label attached to the particle. Alternatively, time may arise as a property of a quantum reference frame [241], i.e. the quantum clock of a physical system [245]. In the Page and Wooters mechanism, time and dynamics emerge from the entanglement between the system and the clock [245]. Several criticisms have been addressed to this point of view [265], which can be overcome by modifying the original mechanism [247]. This extension correctly reproduces the standard quantum mechanics predictions such as measurements statistics and propagators, under proper conditioning of the entangled state. In this framework, a well-defined time operator can be constructed, which gives rise to a time-energy uncertainty relation and bypasses the above-mentioned Pauli objection [266]. In particular, this formalism can be applied to describe generic quantum time measurements, such as the arrival time of a particle at the detector. Consider the Hilbert space of the physical system $\mathcal{H}_S$, equipped with an ancillary one $\mathcal{H}_C$ with Hamiltonian linear in momentum, which describes time as the quantity measured by the clock. Notice that linearity is necessary to obtain a Schrödinger dynamics. The global Hilbert space is constructed as

$$\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_C, \tag{7.14}$$

so do the global states $|\Phi\rangle$ that are the eigenvectors of the total Hamiltonian, under a constraint with the form of the Wheeler–DeWitt equation [267]. These states are static, simultaneously encoding the history of the system in a given time interval, namely

$$\begin{gathered} |\Phi\rangle = \frac{1}{\sqrt{T}} \int_{\mathcal{T}} \mathrm{d}t \ \ |t\rangle \otimes |\Psi(t)\rangle \ , \\ \text{with } \mathcal{T} = [-T/2, T/2] \, , \end{gathered} \tag{7.15}$$

and $T \to \infty$ a regularization parameter. Here, $|\Psi(t)\rangle$ denotes the time-dependent state, considered so far, obtained by conditioning $|\Phi\rangle$ on t, i.e. on projecting $|\Phi\rangle$ on $|t\rangle$. Consider a dimensionless detector placed at x, a positive-operator-valued measure (POVM) describes a joint measurement on the particle and on the clock as

$$\Theta_t = |t\rangle\langle t| \otimes |x\rangle\langle x| \ \ \forall t \in \mathcal{T} \, , \tag{7.16}$$

$$\Theta^* = 1 - \int \mathrm{d}t \ \Theta_t \, , \tag{7.17}$$

where each value of t labels a different element of the POVM. For $x = 0$, combining the Born rule with the Bayes formula, yields the arrival time distribution as

$$\Pi_C(t) = \frac{1}{N_C(T)} |\Psi(0,t)|^2 \,. \tag{7.18}$$

Indeed, this is the conditioned probability that the particle arrives at the detector at time t, with normalization constant $N_C(T)$ that depends on the regularization parameter T as

$$N_C(T) = \int_{-T/2}^{+T/2} \mathrm{d}t \; |\Psi(0,t)|^2 \,. \tag{7.19}$$

The arrival time distribution becomes independent of regularization if the probability that the particle remains at the detector at arbitrary large times is sufficiently low, namely whenever the integral converges for $T \to +\infty$. Alternatively, Fredenhagen and Brunetti obtained the same distribution in [248], introducing a generalization of quantum states in term of particle weights [268], consistently with the Gelfand–Naimark–Segal representation theorem. In [269], Höhn, Smith and Lock show that the Page-Wootters mechanism can be equivalently formulated in the framework of relational dynamics, originally introduced by Rovelli [243]. In [270], Gambini and Pullin combine the Page and Wooters mechanism with Rovelli's evolving constant of motions, both for a clock Hamiltonian that is unbounded or bounded-below. In the former case, i.e. when the clock Hamiltonian is linear in momentum, this yields the same arrival time distribution of the quantum clock, normalized on the whole time interval. In this case, and in contrast with Kijowski's proposal, no paradoxical behaviour occurs for a particle travelling towards an infinite potential barrier: Π_C predicts no arrivals in the prohibited regions.

7.2.5 Normalizability

We discuss the normalizability of the arrival time proposals. Consider a Gaussian wave packet $\psi(x,t)$. In Kijowski's proposal, the normalization is guaranteed axiomatically, independently from the choice of the packet. For the quantum flux, it follows that

$$\partial_x \psi(x,t) = \left[\frac{x_0 - x + i p_0}{1 + i t} \right] \psi(x,t) \,. \tag{7.20}$$

At later times, this yields

$$\mathrm{Im}\left[\Psi^*(0,t)\partial_x \Psi(0,t)\right] \simeq \frac{p_0}{t^3} \,. \tag{7.21}$$

Namely, the normalization integral converges on the whole time interval. For the quantum clock, the probability density of the packet at later times reads

$$|\psi(0,t)|^2 \simeq \frac{1}{t} \exp\left(-p_0^2\right) \,. \tag{7.22}$$

Such contribution diverges logarithmically when integrated on the whole time domain, namely

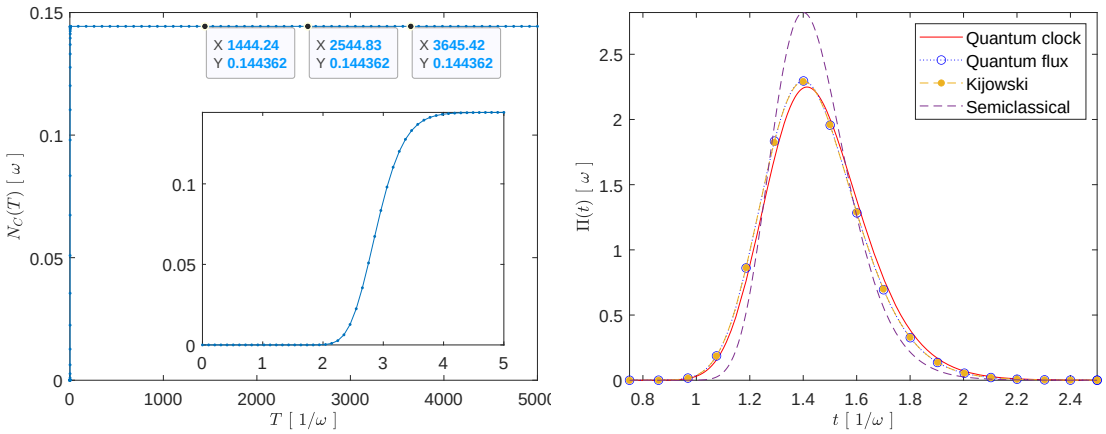
$$N_C(T) = \int_{-T/2}^{+T/2} \mathrm{d}t \ |\psi(0,t)|^2 \simeq 2\exp\left(-p_0^2\right)\log\left(\frac{T}{2}\right) . \tag{7.23}$$

In this case, the quantum clock distribution is not normalizable. Unlike the other proposals, Π_C directly depends on the probability density of the wave function, thus it inherits its properties, including the non-normalizability in time. Such an issue arises also classically. Consider the motion of a classical cloud that propagates towards a detector from the left, while spreading due to diffusion. We can describe the probability density of the cloud using the (square modulus) of the Gaussian wave packet. When the diffusion of the cloud is non-negligible, with respect to its translation, the left tail may spread faster than the way it is approaching the detector. This behaviour is reproduced by Eq. (7.23), where the divergence increases exponentially with the reciprocal of the momentum. Hence, it is impossible to completely detect the cloud even by keeping the detector switched on indefinitely.

The above considerations suggest that it is ill-advised to postulate that the arrival time distribution should be normalized. There are trivial situations in which normalization should not hold at all, such as when the packet is stationary at the detector position or in the above scenario, when the it spreads so quickly that the trailing tail has sufficiently large probability to be found at the detector position at arbitrarily large times. Nevertheless, non-normalizability is not appreciable in usually reasonable scenarios, since N_C is almost constant with respect to the total duration of the experiment. We report a numerical calculation in Fig. 7.2a. Initially, the normalization constant abruptly increases, i.e. when the bulk of the packet approaches the detector. Then, the integral shows a slow logarithmic growth, produced by the trailing tale of the packet that spreads while crossing the detector. A reasonable choice of the integration domain produces no numerical difference. From now on, we consider $T = 100\ 1/\omega$.

7.3 Comparison using Gaussian packets

In this section, we compare the arrival time proposals, introduced above, in different regimes. Focusing on the interference patterns produced by Gaussian wave packets that overlap near a detector, we analyse the distribution of times at which the particle occupies the detector position, without specializing to any detection model, experimental apparatus or measurement protocol. Further analysis will be conducted in future research. The detector cross-sectional area is assumed much greater than the transversal width of each packet, so that a click can happen whenever a particle reaches its longitudinal position. Moreover, we restrict to those regimes in which each packet can be detected at most once, so that its subsequent collapse provides no contribution in the arrival time distribution.



(a) (b)

Figure 7.2: Arrival time distribution of a single Gaussian wave packet. (a) Quantum clock normalization constant. (b) Comparison of the different proposals. Figure adapted from [6].

7.3.1 Single Gaussian packet

Consider the arrival time distribution of a single Gaussian wave packet that travels towards the detector from the left. Single packets allow a trajectory interpretation. Hence, we safely employ also the semiclassical approximation. We report the numerical results in Fig. 7.2b, using a packet with initial conditions $x_0 = -10\, l_0$, $p_0 = 7\, \hbar/l_0$ and $\sigma_0 = 1\, l_0$. Here, the quantum clock produces similar predictions, although not equivalent, to that of Kijowski and of the quantum flux, which overlap in this regime.

7.3.2 Gaussian wave train

Consider a sequence of n Gaussian wave packets, described by the superposition

$$\Psi(t) = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \Psi_k(t)\,, \tag{7.24}$$

where $\Psi_k(t)$ describe a single Gaussian wave packet. We denote x_k and p_k the average position and momentum of the kth wave function, respectively. We call $\Psi(t)$ Gaussian wave train. All the proposals are well-behaved for a Gaussian train that satisfies the no-spreading condition, i.e. a train for which quantum diffusion is negligible and the propagation of each packet is approximately a rigid translation [228]. Such can be employed in Kijowski's proposal by using the Leavens prescription of Eq. (7.8). In this regime, each distribution yields an equiprobable sequence of peaks centred around the classical arrival time $|x_k m/p_k|$. When diffusion is taken into account, a similar behaviour is obtained with a sequence of peaks that show an overall exponential decay. In the next comparisons, we never make such an approximation and always consider Gaussian trains that propagate while spreading. We report the numerical

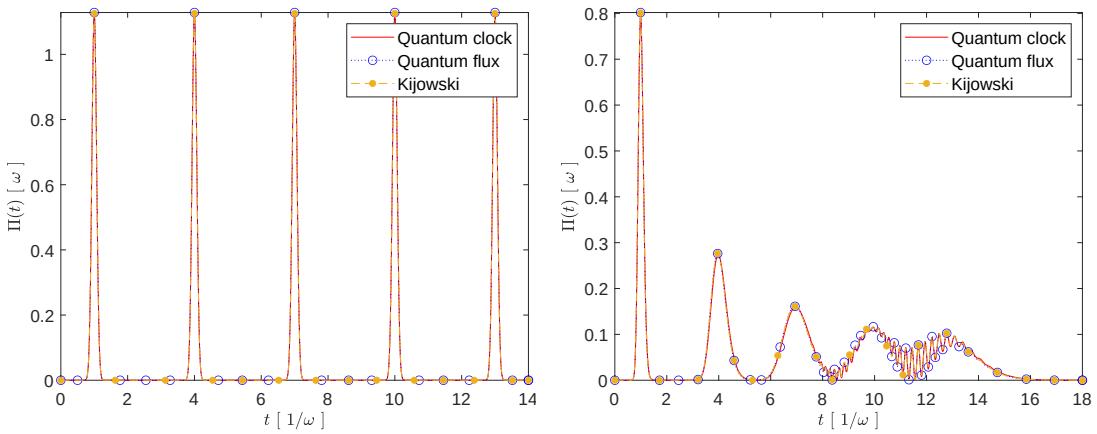


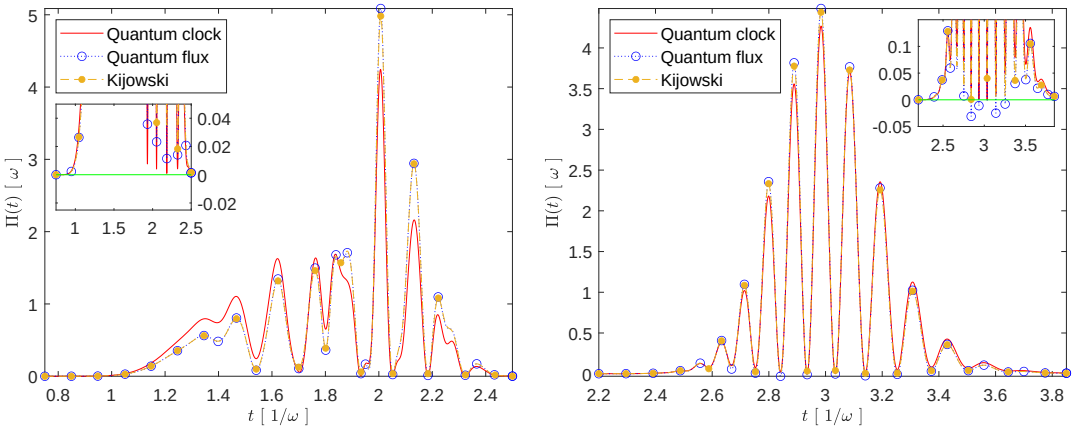
Figure 7.3: Arrival time distribution of a train of $n = 5$ Gaussian wave packets, propagating towards the detector from the left. (a) The wave train satisfies the no-spreading condition, such that no overlapping happens near the detector. (b) When spreading is considered, the peak sequence undergoes an exponential decay. In both cases, the proposals predict compatible results (missing markers are due to sampling differences).

results in Fig. 7.3, using packets with parameters $x_k = -10(1+3k)\ l_0$, $p_k = 10\ \hbar/l_0$ and $\sigma_0 = 1\ l_0$. In this regime, all the proposals produce similar results.

7.3.2.1 Copropagating packets

Consider a train of $n = 4$ Gaussian wave packets that propagate towards the detector in the same direction. We adjust the parameters of each packet such that they partially overlap near the detector. We compute the arrival time distribution according to each proposal. We report the numerical results in Fig. 7.4a, using packets with parameters $x_k = -2(5+4k)\ l_0$, $p_k = (7+3k)\ \hbar/l_0$ and $\sigma_k = 1\ l_0$. In this case, all the distributions show an interference pattern. However, the quantum clock behaves differently from the others, in a way that cannot be compensated by a different choice of the normalization constant.

Consider a train of $n = 2$ Gaussian wave packets, respectively localized at x_0 and x_1, with momenta p_0 and p_1. Given $p_1 > p_0$, we adjust x_1 in such a way that both wave packets arrive at the detector with the same average arrival time, that is when the faster but delayed packet overtakes the other one at $x = 0$. We refer to this last requirement as an overtaking condition, namely $x_1/p_1 = x_0/p_0$. We report the numerical results in Fig. 7.4b, initializing the packets with $x_0 = -30\ l_0$, $p_0 = 10\ \hbar/l_0$ and $p_1 = 15\ \hbar/l_0$ and $\sigma_0 = \sigma_1 = 1\ l_0$, yielding $x_1 = -45\ l_0$. In this regime, interference arises with some marked differences among the proposals. This behaviour cannot be justified by a proper redefinition of the quantum clock normalization factor. Moreover, the quantum flux distribution is not positive-definite, hence its probabilistic interpretation is not allowed here. This issue is due to the presence of quantum



(a) (b)

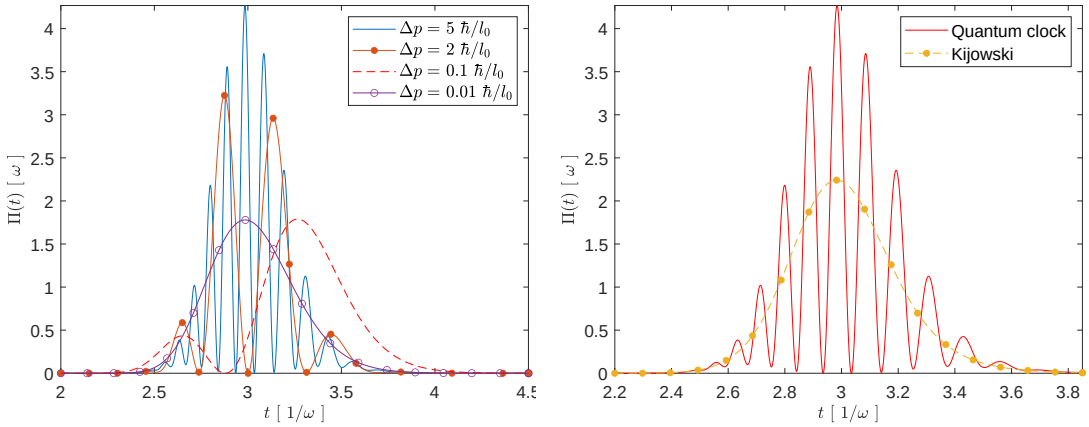
Figure 7.4: Arrival time distribution of a train of n Gaussian wave packets, approaching the detector from the same side. (a) Superposition of $n = 4$ packets, which partially overlap near the detector. All the distributions take positive value only, admitting a probabilistic interpretation. The inset is a zoom-in of the main plot close to the horizontal axis. (b) Superposition of $n = 2$ packets, which completely overlap at the detector under the overtaking condition. The proposals produce significantly different results. The inset shows that negative probabilities taken by the quantum flux, due to the presence of backflow. Figure adapted from [6].

backflow [262]. Although both packets have support for positive momenta only, i.e. they approach the detector from the left, interference may produce a temporary inversion of the motion at the detector, i.e. both right and left-moving contributions in the current operator. This yields an arrival time distribution with negative probabilities. In this same regime, backflow can also be analysed in terms of Bohmian trajectories [249]. Conversely, both Kijowski's and the quantum clock distributions remain positive-definite, although they still produce quantitatively different results.

Under the overtaking condition, the number of interference fringes is determined by the momentum separation between the two packets $\Delta p = |p_0 - p_1|$. The lower Δp is, the wider and fewer are the fringes in the arrival time pattern. When this separation is sufficiently low, interference disappears and a result similar to the single packet regime of Fig. 7.2b is obtained. In Fig. 7.5a, we show this behaviour for the quantum clock.

7.3.2.2 Counterpropagating packets

We repeat the same analysis of Fig. 7.4b, using a superposition of packets supported for both positive and negative momenta values. To this extent, we consider a couple of counterpropagating Gaussian wave packets, i.e. that cross the detector from two opposite directions with the same average arrival time. We report the numerical result in Fig. 7.5b, where the packets are initialized with $x_0 = -30\ l_0$, $p_0 = 10\ \hbar/l_0$, $x_1 = 45\ l_0$, $p_1 = -15\ \hbar/l_0$ and $\sigma_0 = \sigma_1 = 1\ l_0$. We compare the predictions of



(a) (b)

Figure 7.5: Arrival time distribution of a train of $n = 2$ Gaussian wave packets. (a) Quantum clock prediction for different momentum separation. The lower is the separation, the wider and less numerous are the interference fringes, until a minimum value is reached, and interference completely disappears. This behaviour is also reproduced by the other proposals, constrained to their regimes of validity. (b) Comparison for $n = 2$ counterpropagating packets that meet at the detector with the same average arrival time. Kijowski's proposal lacks the interference fringes, while the quantum flux cannot be applied in this regime. Figure adapted from [6].

Kijowski's and the quantum clock distributions, while the quantum flux gives meaningless results in this regime. In this case, the anomalous behaviour of the quantum flux is not related to the presence of quantum backflow, it rather arises from our choice of packets with opposite momenta signs. Kijowski's proposal shows no interference pattern. This is due to how positive and negative momenta independently contribute to Eq. (7.6), which are separately treated and combined in the final distribution. For the quantum clock, the arrival time distribution is exactly the same as the one obtained in the copropagating regime, where the packets approach the detector from the same side. Indeed, two Gaussian wave packets yield

$$\Pi_C(t) = \frac{1}{N_C}\left|\sum_{k=0}^{1}\frac{\exp\left[-(x_k^2 + 2ix_kp_k\sigma_k^2 + itp_k^2\sigma_k^2)/2(\sigma_k^2 + it)\right]}{\sqrt{\sigma_k + it/\sigma_k}}\right|^2 , \tag{7.25}$$

with x_k and p_k the initial average position and momentum of each packet. Here, the substitution

$$\begin{cases} x_1 \to -x_1 \\ p_1 \to -p_1 \end{cases} , \tag{7.26}$$

preserves the right-hand side of Eq. (7.25), while translating the counterpropagating regime into the copropagating (or the other way round). This transformation does not affect the quantum clock distribution, which predicts the equivalence of the two regimes, when the detector is placed in the origin. Importantly, this symmetry is not

present in Kijowski's distribution and cannot be observed in the quantum flux, which is not well defined under such transformations.

7.4 Experimental proposal

In this section, we discuss the possibility to experimentally observe the arrival time interference patterns described above. Consider the experimental parameters of a Bose-Einstein condensate trapped in an accelerator ring [254]. Prepare the condensate by using a cloud of ^{87}Rb subject to a harmonic potential. After releasing the trap, the condensate is loaded and accelerated along a ring of radius $R = 443\,\mu\text{m}$. We describe each atom of the cloud using a Gaussian wave packet that propagates on the circle of length $d = 2\pi R$, with periodic boundary conditions in the interval $[-d/2, d/2]$, thus both its momentum and energy take discrete eigenvalues

$$p_n = \frac{2\pi n\hbar}{d}\,, \tag{7.27}$$

$$E_n = \frac{2\pi^2 n^2\hbar^2}{md^2}\,. \tag{7.28}$$

We focus only on the longitudinal degree of freedom, assuming that the cloud is sufficiently confined and that its transversal spreading is negligible with respect to length of the ring. As long as the confinement is sufficiently strong, the atoms occupy only the lowest energy level of the transverse degree of freedom. In this case, the condensate truly behaves like a one-dimensional system. Hence, each wave packet is described by

$$\psi(x,t) = \sum_{n=-N}^{N} a_n \exp\left[\frac{ip_n(x-\bar{x}_0)}{\hbar}\right]\exp\left[-\frac{iE_n t}{\hbar}\right]\,, \tag{7.29}$$

where x labels the curvilinear coordinate on the ring. For a packet with initial spatial width $\sigma_0 \ll d$, the wave plane coefficients read

$$a_n = \left(\frac{4\pi\sigma_0^2}{d^4}\right)^{1/4}\exp\left[\frac{-(p_n-\bar{p}_0)^2\sigma_0^2}{2\hbar^2}\right]\,. \tag{7.30}$$

Here. $\bar{x}_0$ and $\bar{p}_0$ respectively denote the initial average position and momentum of the packet in the period $[-d/2, d/2]$. For $\sigma_0 \ll d$ the tails of the Gaussian are negligible outside the period, so we can approximate the wave plane coefficients by integrating the packet on the whole space interval, i.e. by substituting the discrete momentum eigenvalues in the Fourier transform of the packet. We prepare the superposition of the above packets, by letting the condensate interact with a standing light wave. Due to Bragg diffraction, this coherently splits each packet with a momentum transfer that depends on the angle of incidence of light [253]. Detection is then indirectly achieved, targeting the condensate with a focused laser beam, which is kept on for the entire duration of the experiment. The interaction with the beam extracts some atoms out of the condensate (with almost negligible feedback on the condensate).

By collecting these atoms (or the photons scattered during the process), we can use the corresponding extraction rate to reconstruct the arrival time distribution. In this case, the arrival of the particle is conditioned at the longitudinal position where the focused laser beam intersects the condensate. This corresponds to virtually placing a detector in the ring, with temporal resolution controlled by the beam waist of the laser.

In practice, we place the detector in the origin. Then, we consider a packet with $\bar{x}_0 = -d/2$, $\sigma_0 = 100\,\mu\text{m}$ and $\bar{p}_0 = p_{450}$. Using Bragg diffraction, we coherently split it into another one with same initial position and width, but momentum $\bar{p}_1 = p_{350}$. In Fig. 7.6a, we plot the arrival time distributions given by the superposition of these two packets. With these initial conditions a partial overlap occurs, in a similar way it does in Fig. 7.4a. The quantum clock produces significantly different results from the other proposals. In this case, the contributions from negative momentum components are negligible, hence the differences between it and Kijowski's distribution cannot be attributed to the incoherent sum of Eq. (7.9), since $\Pi_- \simeq 0$. We perform the same analysis for two counterpropagating packets with initial positions $\bar{x}_0 = \bar{x}_1 = -d/2$, and momenta $\bar{p}_0 = p_{450}$, $\bar{p}_1 = -p_{400}$ with same initial standard deviation. We report the numerical results in Fig. 7.6b. Similarly to what happened in Fig. 7.5b, Kijowski's predicts no interference fringes with respect to the quantum clock proposal, so also in this regime an experimental comparison is desirable.

Finally, taking into account experimental errors, we estimate the minimum amount of data to achieve the results of this section. Let Π_1 and Π_2 be two arrival time distributions. Consider M detector clicks temporally distributed in a histogram with a certain number of bins. In the kth bin there are two possible outcomes, which are described by a Bernoulli distribution: the click falls within the bin Δt_k with probability f_k, or it does not with probability $1-f_k$. Discrimination is possible whenever the error bar in the kth bin is sufficiently smaller than the vertical separation between the two distribution

$$D = \left| \int_{\Delta t_k} \mathrm{d}t \; [\Pi_1(t) - \Pi_2(t)] \right| , \tag{7.31}$$

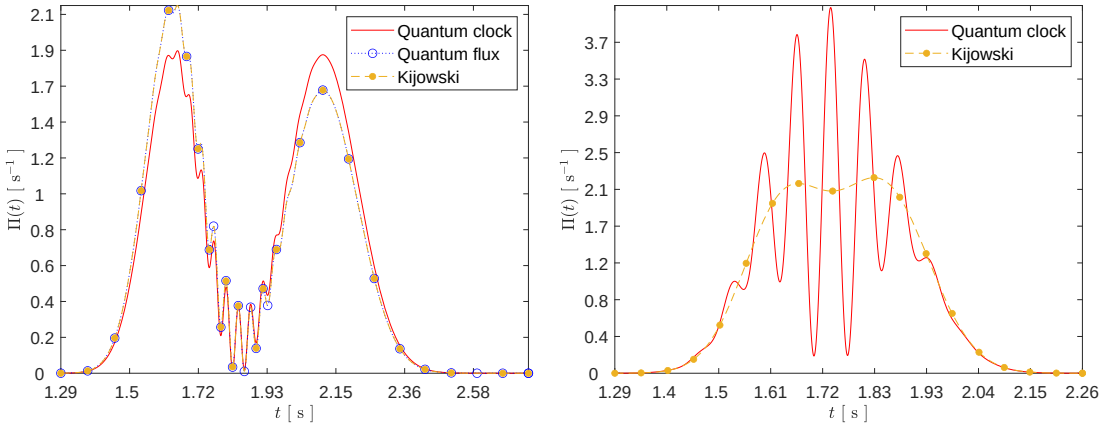
namely when $\varepsilon_k < D/2$, with ε_k the standard error on the kth bin. For a Bernoulli distribution

$$\varepsilon_k = \sqrt{f_k(1-f_k)/M} \, , \tag{7.32}$$

which implies that the condition to achieve experimental discrimination becomes

$$M > \frac{4 f_k}{D^2} (1 - f_k) \, . \tag{7.33}$$

Here, f_k is approximated by the experimental sample in terms of the relative frequencies M_k/M, with M_k the number of clicks in the kth bin. Then, through a conventional χ^2 test we can compare the experimental data with the different theoretical predictions to determine which of the arrival time proposals correctly reproduces the measurement outcomes. In Fig. 7.6a, we choose as bin $\Delta t_k = [1.29\,\text{s}, 1.83\,\text{s}]$. We theoretically estimate f_k by integrating the arrival time distribution that maximises the lower bound of Eq. (7.33) in Δt_k. In this regime $M \simeq 300$ clicks (i.e. single packet



(a) (b)

Figure 7.6: Arrival time distributions for a superposition of $n = 2$ Gaussian wave packets with periodic boundary conditions on a ring, which mimics the properties of a Bose-Einstein condensate split using Bragg diffraction. The distributions are normalized in their respective plot intervals. (a) The detector is located at $x = 0$, the two packets are placed at $\bar{x}_0 = \bar{x}_1 = -1.39\,\mathrm{mm}$ and propagate clockwise with $\bar{p}_0 = 1.07 \times 10^{-25}\,\mathrm{mm\,kg\,s^{-2}}$ and $\bar{p}_1 = 0.83 \times 10^{-25}\,\mathrm{mm\,kg\,s^{-2}}$. Under these initial conditions, the packets partially overlap at the detector. For this reason, no quantum backflow is observed. (b) The second packet is initially placed at $\bar{x}_1 = -1.39\,\mathrm{mm}$, but it propagates counterclockwise with $\bar{p}_1 = -0.95 \times 10^{-25}\,\mathrm{mm\,kg\,s^{-2}}$. Kijowski's distribution shows no interference fringes, while the quantum flux cannot be applied in this non-definite momentum sign regime. Both (a) and (b) require a temporal resolution of 0.1 s, which approximately corresponds to a detector size of 50 µm. Figure adapted from [6].

detections) represent a sufficient amount of data to discriminate between Kijowski's and the quantum clock distributions.

Photonic platforms represent a potential alternative to Bose-Einstein condensates; however, they pose further challenges. Mass is responsible for the spreading mechanism, which leads to numerical discriminability. Spreading can be optically achieved by letting the packets propagating in a dispersive medium or in a waveguide that supports transverse standing modes. This leads to a trade-off between numerical discriminability and technological constraints in detector resolution, which may result in a lower visibility.

7.5 Discussion

In this chapter, we reviewed some of the theories that deal with the quantum arrival time problem. We compared Kijowski's, the quantum flux and the quantum clock proposals, using different superpositions of Gaussian wave packets in multiple regimes. Our discussion is not limited to predicting a specific experimental result,

nor it specializes to any particular model of particle detection: it highlights those regimes in which different theoretical approaches lead to contradicting predictions, by analysing the distribution of times at which the particle occupies a certain position, i.e. where the detector should be placed. Although inequivalent, Kijowski's and the quantum flux distributions behave in a similar way, except when the quantum backflow contributions are not negligible. Instead, the quantum clock distribution produces significantly different results, particularly in the counterpropagating regime. In this case, Kijowski's distribution predicts no interference fringes, while the quantum clock yields exactly the same pattern that occurs for two overtaking packets. This symmetry is not reproduced by Kijowski's proposal, while it cannot be observed in the quantum flux. Ultimately, different proposals would give conflicting predictions if they corresponded to different measurement protocols or arrival time definitions. This possibility will be investigated in future research. Such differences can be tested in the laboratory. Finally, we discussed this possibility for an experimental implementation that combines interferometry for Bose-Einstein condensate with Bragg diffraction. This design is capable of experimentally discriminating Kijowski's and quantum flux distribution from the quantum clock. It also provides a useful test for the symmetry possessed by the latter, since it allows to compare the overtaking and the counterpropagating regimes, for two packets with the same absolute value of initial position and momentum.

CONCLUSIONS AND FUTURE DEVELOPMENTS

Not quite. There's room for a little more.

Frodo Baggins, The Return of The King
in Peter Jackson's Adaptation

In this thesis, we presented the main results that appeared in [1–6], which primarily concerned the topics of quantum imaging and image processing, quantum channels and tomography, as well as quantum phase and time measurements.

In Part I, we presented a hardware encoder that, at the output of an imaging apparatus, maps the light intensity at each point on the image plane to the probability amplitude of the corresponding qubit of being excited by a single-photon interaction [1]. This setup can perform amplitude encoding of optical spatial information into a multiqubit quantum register. Then, we introduced a quantum algorithm that can downsample the spatial information stored in a multiqubit state [1], obtained using the previous hardware encoder or by direct state preparation. Using the quantum Fourier transform, the algorithm discards the high spatial-frequency components of the input image, reducing the number of encoding qubits without altering the overall input pattern. Although image reconstruction can severely affect the overall complexity, the algorithm displays an advantage over classical, interpolation-based, implementations. Finally, we considered the task of image classification [2]. Using the Hong-Ou-Mandel effect, we designed an interferometric scheme that can classify images without direct image reconstruction, using the amplitude information stored in a single-photon state. Such an apparatus can be trained in recognizing any category of images, since its response function matches that of a classical artificial neuron, over which it provides a superexponential speedup. There are multiple future developments concerning this research. On the hardware side, multiple entangled sensors could be combined in a single image acquisition, which could potentially operate below the Rayleigh bound. The downsampling algorithm can be generalized to the multidimensional case, by separately treating each axis of the input data,

or inverted, providing a quantum upsampling algorithm to increase the resolution of an input image. Such generalizations are the object of [7]. Different encoding scheme could be considered, bypassing the amplitude-based one in favour of deterministic approaches, protecting the advantage from the cost of image reconstruction. Concerning the Hong-Ou-Mandel classifier, further research will be primarily concentrated on adapting the interferometric scheme to multiple neurons, combined into a quantum optical neural network.

In Part II, we considered the problem of characterizing and mitigating multiqubit noisy processes. Focusing on the Pauli transfer matrix representation of a channel, we introduced a tomographic reconstruction technique which can directly estimate the channel components from the experimental data [3]. For single matrix element extractions, such method requires at most two experimental configurations, independently of the number of qubits and of experimental repetitions, providing an exponential speedup over standard quantum process tomography. Then, we discussed a deconvolution technique for obtaining the noiseless expectation values of observables, by post-processing the data obtained through noisy measurements [4]. This protocol can be combined with the characterization one, taking the deconvolution from a selective knowledge of the channel. For Pauli channels, this provides at least a quadratic speedup over standard tomographic methods. Further generalizations can be considered, which are expected to retain their advantage even in the case of quantum systems with arbitrary dimension, e.g. qudits. Besides the advantage in terms of tomographic resources, possible statistical speedup could be considered, combining the above techniques with shadow tomography.

In Part III, we discuss the estimation of quantities that are not associated to observables in quantum mechanics, e.g. phase and time. First, we considered the problem of estimating an unknown phase at arbitrary uncertainty. We introduced an interferometric setup, whose response function approximates a square wave [5]. This translates the estimation problem into a binary search task, which provides a global estimation strategy that can reach the Heisenberg bound without the use of entanglement and independently of the amount of statistics collected at the output. Then, we consider the task of measuring the time at which a particle reaches a detector, i.e. the arrival time problem. We compared the proposals that tried to solve this issue [6], by either using quantum-mechanical techniques or by extending the standard quantum theory. We explored multiple regimes in which such proposals provide conflicting and inequivalent predictions, both from a qualitative and quantitative point of view. Finally, we analysed a potential experimental implementation, using a Bose-Einstein condensate trapped in an accelerator ring.

BIBLIOGRAPHY

[1] S. Roncallo, L. Maccone, and C. Macchiavello, Quantum JPEG, AVS Quantum Sci. **5**, 043803 (2023).

[2] S. Roncallo, A. R. Morgillo, C. Macchiavello, L. Maccone, and S. Lloyd, Quantum optical classifier with superexponential speedup (2024), arXiv:2404.15266 [quant-ph] .

[3] S. Roncallo, L. Maccone, and C. Macchiavello, Pauli transfer matrix direct reconstruction: channel characterization without full process tomography, Quantum Sci. Technol. **9**, 015010 (2023).

[4] S. Roncallo, L. Maccone, and C. Macchiavello, Multiqubit noise deconvolution and characterization, Phys. Rev. A **107**, 022419 (2023).

[5] S. Roncallo, X. Lu, and L. Maccone, Interferometric binary phase estimations (2024), arXiv:2407.10966 [quant-ph] .

[6] S. Roncallo, K. Sacha, and L. Maccone, When does a particle arrive?, Quantum **7**, 968 (2023).

[7] E. Tumbiolo, S. Roncallo, C. Macchiavello, and L. Maccone, Quantum frequency resampling (2025), arXiv:2501.16283 [quant-ph] .

[8] L. Maccone, S. Roncallo, and K. Sacha, Quantum measurements of time: A reply to criticisms (2025), arXiv:2501.10416 [quant-ph] .

[9] P. Meda, N. Pinamonti, S. Roncallo, and N. Zanghì, Evaporation of four-dimensional dynamical black holes sourced by the quantum trace anomaly, Class. Quant. Grav. **38**, 195022 (2021).

[10] A. M. Turing, Computing machinery and intelligence, Mind **LIX**, 433 (1950).

[11] A. Purohit, M. Kaur, Z. C. Seskir, M. T. Posner, and A. Venegas-Gomez, Building a quantum-ready ecosystem, IET Quantum Commun. **5**, 1 (2024).

[12] T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko, Optical imaging by means of two-photon quantum entanglement, Phys. Rev. A **52**, R3429 (1995).

[13] A. Paniate, G. Massaro, A. Avella, A. Meda, F. Pepe, M. Genovese, M. D'Angelo, and I. Ruo-Berchera, Light-field ghost imaging, Phys. Rev. Appl. **21**, 024032 (2024).

[14] M. D'Angelo, F. V. Pepe, A. Garuccio, and G. Scarcelli, Correlation plenoptic imaging, Phys. Rev. Lett. **116**, 223602 (2016).

[15] F. D. Lena, G. Massaro, A. Lupo, A. Garuccio, F. V. Pepe, and M. D'Angelo, Correlation plenoptic imaging between arbitrary planes, Opt. Express **28**, 35857 (2020).

[16] G. Massaro, F. Di Lena, D. Giannella, F. Pepe, F. Scattarella, S. Vasiukov, and M. D'Angelo, Quantum-enhanced plenoptic imaging, in *Quantum Communications and Quantum Imaging XX* (2022).

[17] A. F. Abouraddy, M. B. Nasr, B. E. A. Saleh, A. V. Sergienko, and M. C. Teich, Quantum-optical coherence tomography with dispersion cancellation, Phys. Rev. A **65**, 053817 (2002).

[18] M. B. Nasr, B. E. A. Saleh, A. V. Sergienko, and M. C. Teich, Demonstration of dispersion-canceled quantum-optical coherence tomography, Phys. Rev. Lett. **91**, 083601 (2003).

[19] M. C. Teich, B. E. A. Saleh, F. N. C. Wong, and J. H. Shapiro, Variations on the theme of quantum optical coherence tomography: a review, Quantum Inf. Process. **11**, 903 (2012).

[20] M. I. Kolobov and C. Fabre, Quantum limits on optical resolution, Phys. Rev. Lett. **85**, 3789 (2000).

[21] S. M. Barnett, C. Fabre, and A. Maıtre, Ultimate quantum limits for resolution of beam displacements, Eur. Phys. J. D **22**, 513 (2003).

[22] V. Giovannetti, S. Lloyd, L. Maccone, and J. H. Shapiro, Sub-Rayleigh-diffraction-bound quantum imaging, Phys. Rev. A **79**, 013827 (2009).

[23] C. A. Pérez-Delgado, M. E. Pearce, and P. Kok, Fundamental limits of classical and quantum imaging, Phys. Rev. Lett. **109**, 123601 (2012).

[24] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Appl. Phys. Rev. **6**, 021318 (2019).

[25] R. Manenti and M. Motta, *Quantum Information Science* (Oxford University Press, 2023).

[26] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Trapped-ion quantum computing: Progress and challenges, Appl. Phys. Rev. **6**, 021314 (2019).

[27] L. Henriet, L. Beguin, A. Signoles, T. Lahaye, A. Browaeys, G.-O. Reymond, and C. Jurczak, Quantum computing with neutral atoms, Quantum **4**, 327 (2020).

[28] S. Slussarenko and G. J. Pryde, Photonic quantum information processing: A concise review, Appl. Phys. Rev. **6**, 041303 (2019).

[29] A. Y. Kitaev, Quantum measurements and the abelian stabilizer problem (1995), arXiv:quant-ph/9511026 [quant-ph] .

[30] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (1996) p. 212–219.

[31] P. W. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994) pp. 124–134.

[32] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, Phys. Rev. Lett. **103**, 150502 (2009).

[33] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195–202 (2017).

[34] F. Yan and S. E. Venegas-Andraca, *Quantum Image Processing* (Springer, 2020).

[35] A. Gyenis, A. Di Paolo, J. Koch, A. Blais, A. A. Houck, and D. I. Schuster, Moving beyond the transmon: Noise-protected superconducting quantum circuits, PRX Quantum **2**, 030101 (2021).

[36] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[37] S. Severini, Bye NISQ. Hello LISQ?, https://www.linkedin.com/pulse/bye-nisq-hello-lisq-simone-severini-ybkmc (2023).

[38] J. M. Arrazola and Xanadu, From NISQ to ISQ, https://pennylane.ai/blog/2023/06/from-nisq-to-isq (2023).

[39] Y. Kim, A. Eddins, S. Anand, K. Wei, E. Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, Nature **618**, 500 (2023).

[40] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin, Logical quantum processor based on reconfigurable atom arrays, Nature **626**, 58 (2024).

[41] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, High-threshold and low-overhead fault-tolerant quantum memory, Nature **627**, 778 (2024).

[42] M. P. da Silva, C. Ryan-Anderson, J. M. Bello-Rivas, A. Chernoguzov, J. M. Dreiling, C. Foltz, F. Frachon, J. P. Gaebler, T. M. Gatterman, L. Grans-Samuelsson, D. Hayes, N. Hewitt, J. Johansen, D. Lucchetti, M. Mills, S. A. Moses, B. Neyenhuis, A. Paz, J. Pino, P. Siegfried, J. Strabley, A. Sundaram, D. Tom, S. J. Wernli, M. Zanner, R. P. Stutz, and K. M. Svore, Demonstration of logical qubits and repeated error correction with better-than-physical error rates (2024), arXiv:2404.02280 [quant-ph] .

[43] Google Quantum AI and Collaborators, Quantum error correction below the surface code threshold (2024), arXiv:2408.13687 [quant-ph] .

[44] W. Pauli, *General Principles of Quantum Mechanics* (Springer, 1980).

[45] B. Mielnik, The screen problem, Found. Phys. **24**, 1113 (1994).

[46] W. S. Boyle, Nobel Lecture: CCD - An extension of man's view, Rev. Mod. Phys. **82**, 2305 (2010).

[47] W. S. Boyle and G. E. Smith, Charge coupled semiconductor devices, Bell Syst. Tech. J. **49**, 587 (1970).

[48] J. D. Jackson, *Classical Electrodynamics* (Wiley, 1998).

[49] C. Gerry and P. Knight, *Introductory Quantum Optics* (Cambridge University Press, 2004).

[50] M. O. Scully and M. S. Zubairy, *Quantum Optics* (Cambridge University Press, 1997).

[51] L. Maccone, Y. Zheng, and C. Ren, Gaussian beam quantum radar protocol (2023), arXiv:2309.11834 [quant-ph] .

[52] M. Srednicki, *Quantum Field Theory* (Cambridge University Press, 2007).

[53] S. Weinberg, *The Quantum Theory of Fields. Vol. 1: Foundations* (Cambridge University Press, 2005).

[54] F. Strocchi, *An Introduction to Non-Perturbative Foundations of Quantum Field Theory* (Oxford University Press, 2013).

[55] R. Brunetti, C. Dappiaggi, K. Fredenhagen, and J. Yngvason, eds., *Advances in Algebraic Quantum Field Theory* (Springer, 2015).

[56] C. J. Fewster and M. J. Pfenning, A quantum weak energy inequality for spin-one fields in curved space–time, J. Math. Phys. **44**, 4480–4513 (2003).

[57] S. Murro and G. Schmid, The quantization of Maxwell theory in the Cauchy radiation gauge: Hodge decomposition and Hadamard states (2024), arXiv:2401.08403 [math.AP] .

[58] S. N. Gupta, Theory of longitudinal photons in quantum electrodynamics, Proc. Phys. Soc. A **63**, 681 (1950).

[59] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, 1995).

[60] M. Rezai and J. A. Salehi, Fundamentals of quantum Fourier optics, IEEE Trans. Quantum Eng. **4**, 1 (2022).

[61] S. M. Barnett and P. M. Radmore, *Methods in Theoretical Quantum Optics* (Oxford University Press, 2002).

[62] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (Wiley, 1991).

[63] E. Jaynes and F. Cummings, Comparison of quantum and semiclassical radiation theories with application to the beam maser, Proceedings of the IEEE **51**, 89 (1963).

[64] C. Emary, Dark-states in multi-mode multi-atom Jaynes-Cummings systems, J. Phys. B: At. Mol. Opt. Phys. **46**, 224008 (2013).

[65] A. Wickenbrock, M. Hemmerling, G. R. M. Robb, C. Emary, and F. Renzoni, Collective strong coupling in multimode cavity QED, Phys. Rev. A **87**, 043817 (2013).

[66] M. Tavis and F. W. Cummings, Exact solution for an N-molecule—radiation-field Hamiltonian, Phys. Rev. **170**, 379 (1968).

[67] B. Chen, H. Wu, H. Yuan, L. Wu, and X. Li, A logarithm depth quantum converter: From one-hot encoding to binary encoding (2022), arXiv:2206.11166 [quant-ph] .

[68] A. Bärtschi and S. Eidenbenz, Deterministic preparation of Dicke states, in *Fundamentals of Computation Theory* (2019) pp. 126–139.

[69] S. E. Venegas-Andraca and S. Bose, Storing, processing, and retrieving an image using quantum mechanics, in *Quantum Information and Computation*, Vol. 5105 (SPIE, 2003) pp. 137 – 147.

[70] F. Yan, A. M. Iliyasu, and S. E. Venegas-Andraca, A survey of quantum image representations, Quantum Inf. Process. **15**, 1 (2016).

[71] Z. Wang, M. Xu, and Y. Zhang, Review of quantum image processing, Arch. Computat. Methods Eng. **29**, 737 (2022).

[72] M. Lisnichenko and S. Protasov, Quant.image representation: a review, Quantum Mach. Intell. **5**, 2 (2022).

[73] P. Q. Le, F. Dong, and K. Hirota, A flexible representation of quantum images for polynomial preparation, image compression, and processing operations, Quantum Inf. Process. **10**, 63 (2011).

[74] Y. Zhang, K. Lu, Y. Gao, and M. Wang, NEQR: a novel enhanced quantum representation of digital images, Quantum Inf. Process. **12**, 2833 (2013).

[75] M. Schuld and F. Petruccione, *Machine Learning with Quantum Computers* (Springer, 2021).

[76] X.-W. Yao, H. Wang, Z. Liao, M.-C. Chen, J. Pan, J. Li, K. Zhang, X. Lin, Z. Wang, Z. Luo, W. Zheng, J. Li, M. Zhao, X. Peng, and D. Suter, Quantum image processing and its application to edge detection: Theory and experiment, Phys. Rev. X **7**, 031041 (2017).

[77] D. Salomon and G. Motta, *Handbook of Data Compression* (Springer, 2010).

[78] *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines*, ISO/IEC 10918-1:1994 (International Organization for Standardization, Geneva, Switzerland, 1994).

[79] S. Roncallo, https://github.com/simoneroncallo/quantum-jpeg.

[80] L. Grover and T. Rudolph, Creating superpositions that correspond to efficiently integrable probability distributions (2002), arXiv:quant-ph/0208112 [quant-ph] .

[81] D. Ramacciotti, A. I. Lefterovici, and A. F. Rotundo, Simple quantum algorithm to efficiently prepare sparse states, Phys. Rev. A **110**, 032609 (2024).

[82] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[83] A. Rotondi, P. Pedroni, and A. Pievatolo, *Probability, Statistics and Simulation: With Application Programs Written in R* (Springer, 2022).

[84] L. A. Shepp and B. F. Logan, The Fourier reconstruction of a head section, IEEE Trans. Nucl. Sci **21**, 21 (1974).

[85] L. Hales and S. Hallgren, An improved quantum Fourier transform algorithm and applications, in *Proceedings 41st Annual Symposium on Foundations of Computer Science* (2000) pp. 515–525.

[86] E. O. Brigham and R. E. Morrow, The fast Fourier transform, IEEE Spectr. **4**, 63 (1967).

[87] S. Eddins, https://github.com/mathworks/zone-plate-test-image.

[88] J. I. Latorre, Image compression and entanglement (2005), arXiv:quant-ph/0510031 [quant-ph] .

[89] N. Jiang, X. Lu, H. Hu, Y. Dang, and Y. Cai, A novel quantum image compression method based on JPEG, Int. J. Theor. Phys. **57**, 611 (2018).

[90] R.-G. Zhou, W. Hu, P. Fan, and H. Ian, Quantum realization of the bilinear interpolation method for NEQR, Sci. Rep. **7**, 2511 (2017).

[91] F. Yan, S. Zhao, S. E. Venegas-Andraca, and K. Hirota, Implementing bilinear interpolation with quantum images, Digit. Signal Process. **117**, 103149 (2021).

[92] Z. Cui, S. Jin, A. Sone, and X. Wang, Quantum advantages for image filtering on images with efficient encoding and lower-bounded signal-to-noise ratio, Sci. China Phys. Mech. Astron. **67**, 290362 (2024).

[93] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum random access memory, Phys. Rev. Lett. **100**, 160501 (2008).

[94] L. Gyongyosi, S. Imre, and H. V. Nguyen, A survey on quantum channel capacities, IEEE Commun. Surv. Tutor. **20**, 1149 (2018).

[95] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE **86**, 2278 (1998).

[96] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM **60**, 84–90 (2017).

[97] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '16, p. 770.

[98] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in *International Conference on Learning Representations*, ICLR '21, arXiv:2010.11929 [cs.CV] .

[99] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, Photonics for artificial intelligence and neuromorphic computing, Nat. Photon. **15**, 102–114 (2021).

[100] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, All-optical machine learning using diffractive deep neural networks, Science **361**, 1004–1008 (2018).

[101] Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, All-optical neural network with nonlinear activation functions, Optica **6**, 1132 (2019).

[102] S. Colburn, Y. Chu, E. Shilzerman, and A. Majumdar, Optical frontend for a convolutional neural network, Appl. Opt. **58**, 3179 (2019).

[103] S. Li, B. Ni, X. Feng, K. Cui, F. Liu, W. Zhang, and Y. Huang, All-optical image identification with programmable matrix transformation, Opt. Express **29**, 26474 (2021).

[104] Y. Luo, Y. Zhao, J. Li, E. Çetintaş, Y. Rivenson, M. Jarrahi, and A. Ozcan, Computational imaging without a computer: seeing through random diffusers at the speed of light, eLight **2**, 4 (2022).

[105] P. L. McMahon, The physics of optical computing, Nat. Rev. Phys. **5**, 717–734 (2023).

[106] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum algorithms for supervised and unsupervised machine learning (2013), arXiv:1307.0411 [quant-ph] .

[107] X.-D. Cai, D. Wu, Z.-E. Su, M.-C. Chen, X.-L. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, Entanglement-based machine learning on a quantum computer, Phys. Rev. Lett. **114**, 110504 (2015).

[108] V. Trávníček, K. Bartkiewicz, A. Černoch, and K. Lemr, Phys. Rev. Lett. **123**, 260501 (2019).

[109] F. Benatti, S. Mancini, and S. Mangini, Continuous variable quantum perceptron, Int. J. Quantum Inf. **17**, 1941009 (2019).

[110] F. Tacchino, C. Macchiavello, D. Gerace, and D. Bajoni, An artificial neuron implemented on an actual quantum processor, Npj Quantum Inf. **5**, 26 (2019).

[111] S. Mangini, F. Tacchino, D. Gerace, C. Macchiavello, and D. Bajoni, Quantum computing model of an artificial neuron with continuously valued input data, Mach. Learn.: Sci. Technol. **1**, 045008 (2020).

[112] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625–644 (2021).

[113] A. Senokosov, A. Sedykh, A. Sagingalieva, B. Kyriacou, and A. Melnikov, Quantum machine learning for image classification, Mach. Learn.: Sci. Technol. **5**, 015040 (2024).

[114] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing (2024), arXiv:2312.09121 [quant-ph] .

[115] C. Bowie, S. Shrapnel, and M. J. Kewming, Quantum kernel evaluation via Hong–Ou–Mandel interference, Quantum Sci. Technol. **9**, 015001 (2023).

[116] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, Quantum optical neural networks, Npj Quantum Inf. **5**, 60 (2019).

[117] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, Phys. Rev. Res. **1**, 033063 (2019).

[118] K. Bartkiewicz, C. Gneiting, A. Černoch, K. Jiráková, K. Lemr, and F. Nori, Experimental kernel-based quantum machine learning in finite feature space, Sci. Rep. **10**, 12356 (2020).

[119] X. Sui, Q. Wu, J. Liu, Q. Chen, and G. Gu, A review of optical neural networks, IEEE Access **8**, 70773 (2020).

[120] A. Zhang, H. Zhan, J. Liao, K. Zheng, T. Jiang, M. Mi, P. Yao, and L. Zhang, Quantum verification of NP problems with single photons and linear optics, Light Sci. Appl. **10**, 169 (2021).

[121] D. Stanev, N. Spagnolo, and F. Sciarrino, Deterministic optimal quantum cloning via a quantum-optical neural network, Phys. Rev. Res. **5**, 013139 (2023).

[122] C. Wood, S. Shrapnel, and G. J. Milburn, A Kerr kernel quantum learning machine (2024), arXiv:2404.01787 [quant-ph] .

[123] C. K. Hong, Z. Y. Ou, and L. Mandel, Measurement of subpicosecond time intervals between two photons by interference, Phys. Rev. Lett. **59**, 2044 (1987).

[124] J. C. Garcia-Escartin and P. Chamorro-Posada, SWAP test and Hong-Ou-Mandel effect are equivalent, Phys. Rev. A **87**, 052330 (2013).

[125] S. Sadana, D. Ghosh, K. Joarder, A. N. Lakshmi, B. C. Sanders, and U. Sinha, Near-100% two-photon-like coincidence-visibility dip with classical light and the role of complementarity, Phys. Rev. A **100**, 013839 (2019).

[126] M. Hiekkamäki and R. Fickler, High-dimensional two-photon interference effects in spatial modes, Phys. Rev. Lett. **126**, 123601 (2021).

[127] A. R. Morgillo and S. Roncallo, https://github.com/simoneroncallo/quantum-optical-neuron.

[128] A. M. Brańczyk, Hong-Ou-Mandel interference (2017), arXiv:1711.00080 [quant-ph] .

[129] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, PMLR, Vol. 9 (2010) pp. 249–256.

[130] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, Psychol. Rev. **65**, 386 (1958).

[131] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).

[132] R. Hecht-Nielsen, Kolmogorov's mapping neural network existence theorem, in *Proceedings of International Conference on Neural* (1987) pp. 11–13.

[133] M. I. Kolobov, The spatial behavior of nonclassical light, Rev. Mod. Phys. **71**, 1539 (1999).

[134] V. Delaubert, N. Treps, C. Fabre, H. A. Bachor, and P. Réfrégier, Quantum limits in image processing, EPL **81**, 44001 (2008).

[135] F. Santosa and W. W. Symes, Linear inversion of band-limited reflection seismograms, SIAM J. Sci. Stat. Comput. **7**, 1307–1330 (1986).

[136] A. Tikhonov and V. Glasko, Use of the regularization method in non-linear problems, USSR Comput. Math. Math. Phys. **5**, 93 (1965).

[137] J. Neff, R. Athale, and S. Lee, Two-dimensional spatial light modulators: a tutorial, Proc. IEEE **78**, 826 (1990).

[138] Z. Zhang, Z. You, and D. Chu, Fundamentals of phase-only liquid crystal on silicon (LCOS) devices, Light Sci. Appl. **3**, e213 (2014).

[139] TensorFlow Developers, TensorFlow (2023).

[140] R. Collobert and S. Bengio, Links between perceptrons, MLPs and SVMs, in *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, p. 23.

[141] G. N. Hounsfield, Nobel Lecture (1979).

[142] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. **70**, 1895 (1993).

[143] C. Macchiavello and G. M. Palma, Entanglement-enhanced information transmission over a quantum channel with correlated noise, Phys. Rev. A **65**, 050301(R) (2002).

[144] A. D'Arrigo, G. Benenti, G. Falci, and C. Macchiavello, Information transmission over an amplitude damping channel with an arbitrary degree of memory, Phys. Rev. A **92**, 062342 (2015).

[145] A. Rivas, S. F. Huelga, and M. B. Plenio, Quantum non-Markovianity: characterization, quantification and detection, Rep. Prog. Phys. **77**, 094001 (2014).

[146] D. Chruściński, C. Macchiavello, and S. Maniscalco, Detecting non-Markovianity of quantum evolution via spectra of dynamical maps, Phys. Rev. Lett. **118**, 080404 (2017).

[147] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, Phys. Rev. A **77**, 032322 (2008).

[148] D. Greenbaum, Introduction to quantum gate set tomography (2015), arXiv:1509.02921 [quant-ph] .

[149] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).

[150] A. Bendersky, F. Pastawski, and J. P. Paz, Selective and efficient quantum process tomography, Phys. Rev. A **80**, 032116 (2009).

[151] A. Gaikwad, D. Rehal, A. Singh, Arvind, and K. Dorai, Experimental demonstration of selective quantum process tomography on an NMR quantum information processor, Phys. Rev. A **97**, 022311 (2018).

[152] A. Gaikwad, K. Shende, Arvind, and K. Dorai, Implementing efficient selective quantum process tomography of superconducting quantum gates on IBM quantum experience, Sci. Rep. **12**, 3688 (2022).

[153] M. Mohseni and D. A. Lidar, Direct characterization of quantum dynamics, Phys. Rev. Lett. **97**, 170501 (2006).

[154] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).

[155] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White, Ancilla-assisted quantum process tomography, Phys. Rev. Lett. **90**, 193601 (2003).

[156] I. Bongioanni, L. Sansoni, F. Sciarrino, G. Vallone, and P. Mataloni, Experimental quantum process tomography of non-trace-preserving maps, Phys. Rev. A **82**, 042307 (2010).

[157] L. Banchi, E. Grant, A. Rocchetto, and S. Severini, Modelling non-Markovian quantum processes with recurrent neural networks, New J. Phys. **20**, 123030 (2018).

[158] S. Shrapnel, F. Costa, and G. Milburn, Quantum Markovianity as a supervised learning task, Int. J. Quantum Inf. **16**, 1840010 (2018).

[159] S. Martina, S. Gherardini, and F. Caruso, Machine learning classification of non-Markovian noise disturbing quantum dynamics, Phys. Scr. **98**, 035104 (2023).

[160] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, Nat. Rev. Phys. 10.1038/s42254-022-00552-1 (2023).

[161] S. Ahmed, F. Quijandría, and A. F. Kockum, Gradient-descent quantum process tomography by learning kraus operators, Phys. Rev. Lett. **130**, 150402 (2023).

[162] A. R. Morgillo, M. F. Sacchi, and C. Macchiavello, Detecting Markovianity of quantum processes via recurrent neural networks (2024), arXiv:2406.07226 [quant-ph] .

[163] C. J. Wood, J. D. Biamonte, and D. G. Cory, Tensor networks and graphical calculus for open quantum systems, Quantum Inf. Comput. **15**, 759 (2015).

[164] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[165] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).

[166] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, ACM Trans. Quantum Comput. **1**, 1 (2020).

[167] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, Quantum process tomography of a controlled-NOT gate, Phys. Rev. Lett. **93**, 080502 (2004).

[168] J. A. Smolin, J. M. Gambetta, and G. Smith, Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise, Phys. Rev. Lett. **108**, 070502 (2012).

[169] M. S. Byrd and N. Khaneja, Characterization of the positivity of the density matrix in terms of the coherence vector representation, Phys. Rev. A **68**, 062322 (2003).

[170] M. Hamada, A lower bound on the quantum capacity of channels with correlated errors, J. Math. Phys. **43**, 4382 (2002).

[171] C. Macchiavello, G. M. Palma, and S. Virmani, Transition behavior in the channel capacity of two-quibit channels with memory, Phys. Rev. A **69**, 010303(R) (2004).

[172] D. Daems, Entanglement-enhanced transmission of classical information in Pauli channels with memory: Exact solution, Phys. Rev. A **76**, 012310 (2007).

[173] P. Huang, G. He, Y. Lu, and G. Zeng, Quantum capacity of Pauli channels with memory, Phys. Scr. **83**, 015005 (2010).

[174] A. D'Arrigo, G. Benenti, G. Falci, and C. Macchiavello, Classical and quantum capacities of a fully correlated amplitude damping channel, Phys. Rev. A **88**, 042337 (2013).

[175] C. Macchiavello and M. F. Sacchi, Witnessing quantum capacities of correlated channels, Phys. Rev. A **94**, 052333 (2016).

[176] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photon. **5**, 222 (2011).

[177] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (2018) p. 325–338.

[178] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[179] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A **52**, R2493 (1995).

[180] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, Phys. Rev. A **54**, 1098 (1996).

[181] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, Phys. Rev. A **54**, 3824 (1996).

[182] S. J. Devitt, W. J. Munro, and K. Nemoto, Quantum error correction for beginners, Rep. Prog. Phys. **76**, 076001 (2013).

[183] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X **8**, 031027 (2018).

[184] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, J. Phys. Soc. Jpn. **90**, 032001 (2021).

[185] S. N. Filippov, S. Maniscalco, and G. García-Pérez, Scalability of quantum error mitigation techniques: from utility to advantage (2024), arXiv:2403.13542 [quant-ph] .

[186] S. Mangini, L. Maccone, and C. Macchiavello, Qubit noise deconvolution, EPJ Quantum Technol. **9**, 29 (2022).

[187] J. Jiang, K. Wang, and X. Wang, Physical implementability of linear maps and its application in error mitigation, Quantum **5**, 600 (2021).

[188] C. Bény, Quantum deconvolution, Quantum Inf. Process. **17**, 26 (2017).

[189] K. M. R. Audenaert and S. Scheel, On random unitary channels, New J. Phys. **10**, 023011 (2008).

[190] K. Siudzińska, Classical capacity of generalized Pauli channels, J. Phys. A Math. Theor. **53**, 445301 (2020).

[191] Z. Cai, X. Xu, and S. C. Benjamin, Mitigating coherent noise using Pauli conjugation, NPJ Quantum. Inf. **6**, 17 (2020).

[192] S. Khatri and M. M. Wilde, Principles of quantum communication theory: A modern approach (2024), arXiv:2011.04672 [quant-ph] .

[193] L. Lautenbacher, F. de Melo, and N. K. Bernardes, Approximating invertible maps by recovery channels: Optimality and an application to non-Markovian dynamics, Phys. Rev. A **105**, 042421 (2022).

[194] A. Gilyén, S. Lloyd, I. Marvian, Y. Quek, and M. M. Wilde, Quantum algorithm for Petz recovery channels and pretty good measurements, Phys. Rev. Lett. **128**, 220502 (2022).

[195] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photon. **5**, 222–229 (2011).

[196] M. G. A. Paris, Quantum estimation for quantum technology, Int. J. Quantum Inf. **07**, 125 (2009).

[197] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, Rev. Mod. Phys. **89**, 035002 (2017).

[198] C. M. Caves, Quantum-mechanical noise in an interferometer, Phys. Rev. D **23**, 1693 (1981).

[199] P. Kómár, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye, and M. D. Lukin, A quantum network of clocks, Nat. Phys. **10**, 582–587 (2014).

[200] C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, Theor. Comput. Sci. **560**, 7 (2014).

[201] K. Tamaki, H.-K. Lo, C.-H. F. Fung, and B. Qi, Phase encoding schemes for measurement-device-independent quantum key distribution with basis-dependent flaw, Phys. Rev. A **85**, 042307 (2012).

[202] A. N. Boto, P. Kok, D. S. Abrams, S. L. Braunstein, C. P. Williams, and J. P. Dowling, Quantum interferometric optical lithography: Exploiting entanglement to beat the diffraction limit, Phys. Rev. Lett. **85**, 2733 (2000).

[203] M. Gessner, N. Treps, and C. Fabre, Estimation of a parameter encoded in the modal structure of a light beam: a quantum theory, Optica **10**, 996 (2023).

[204] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, Proc. R. Soc. Lond. A. **454**, 339–354 (1998).

[205] M. Hassani, C. Macchiavello, and L. Maccone, Digital quantum estimation, Phys. Rev. Lett. **119**, 200502 (2017).

[206] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Entanglement-free Heisenberg-limited phase estimation, Nature **450**, 393–396 (2007).

[207] L. Pezzé and A. Smerzi, Mach-Zehnder interferometry at the Heisenberg limit with coherent and squeezed-vacuum light, Phys. Rev. Lett. **100**, 073601 (2008).

[208] M. G. Genoni, S. Olivares, and M. G. A. Paris, Optical phase estimation in the presence of phase diffusion, Phys. Rev. Lett. **106**, 153603 (2011).

[209] C. R. Schwarze, D. S. Simon, and A. V. Sergienko, Enhanced-sensitivity interferometry with phase-sensitive unbiased multiports, Phys. Rev. A **107**, 052615 (2023).

[210] N. Wiebe and C. Granade, Efficient bayesian phase estimation, Phys. Rev. Lett. **117**, 010503 (2016).

[211] C. Huerta Alderete, M. H. Gordon, F. Sauvage, A. Sone, A. T. Sornborger, P. J. Coles, and M. Cerezo, Inference-based quantum sensing, Phys. Rev. Lett. **129**, 190501 (2022).

[212] M.-A. Filip, D. M. Ramo, and N. Fitzpatrick, Variational phase estimation with variational fast forwarding, Quantum **8**, 1278 (2024).

[213] J. Sinanan-Singh, G. L. Mintzer, I. L. Chuang, and Y. Liu, Single-shot quantum signal processing interferometry, Quantum **8**, 1427 (2024).

[214] P. Busch, T. Heinonen, and P. Lahti, Heisenberg's uncertainty principle, Phys. Rep. **452**, 155 (2007).

[215] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: Beating the standard quantum limit, Science **306**, 1330 (2004).

[216] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry, π-Corrected Heisenberg limit, Phys. Rev. Lett. **124**, 030501 (2020).

[217] F. Belliardo and V. Giovannetti, Achieving heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error, Phys. Rev. A **102**, 042613 (2020).

[218] https://github.com/simoneroncallo/binary-phase-estimation.

[219] H. Lee, P. Kok, and J. P. Dowling, A quantum Rosetta stone for interferometry, J. Mod. Opt. **49**, 2325–2338 (2002).

[220] M. H. Stone, On one-parameter unitary groups in hilbert space, Ann. Math. **33**, 643 (1932).

[221] M. Schuld, R. Sweke, and J. J. Meyer, Effect of data encoding on the expressive power of variational quantum-machine-learning models, Phys. Rev. A **103**, 032430 (2021).

[222] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, Phys. Rev. Lett. **96**, 010401 (2006).

[223] M. W. Mitchell, Metrology with entangled states, in *Quantum Communications and Quantum Imaging III* (2005) p. 589310.

[224] V. Giovannetti and L. Maccone, Sub-Heisenberg estimation strategies are ineffective, Phys. Rev. Lett. **108**, 210404 (2012).

[225] G. Arfken, G. Arfken, H. Weber, and F. Harris, *Mathematical Methods for Physicists: A Comprehensive Guide* (Elsevier, 2012).

[226] G. W. Leibniz, Conversation of Philarète and Ariste, Following a Conversation of Ariste and Theodore, in *Philosophical Papers and Letters*, edited by L. E. Loemker (Springer, 1989).

[227] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, 1965).

[228] S. Das and W. Struyve, Questioning the adequacy of certain quantum arrival-time distributions, Phys. Rev. A **104**, 042214 (2021).

[229] J. G. Muga, R. S. Mayato, and I. L. Egusquiza, *Time in Quantum Mechanics*, Vol. 1 (Springer, 2008).

[230] G. Muga, A. Ruschhaupt, and A. Campo, *Time in Quantum Mechanics*, Vol. 2 (Springer, 2009).

[231] Y. Aharonov and D. Bohm, Time in the quantum theory and the uncertainty relation for time and energy, Phys. Rev. **122**, 1649 (1961).

[232] N. Grot, C. Rovelli, and R. S. Tate, Time of arrival in quantum mechanics, Phys. Rev. A **54**, 4676 (1996).

[233] E. A. Galapon, F. Delgado, J. G. Muga, and I. L. Egusquiza, Transition from discrete to continuous time-of-arrival distribution for a quantum particle, Phys. Rev. A **72**, 042107 (2005).

[234] J. Kijowski, On the time operator in quantum mechanics and the Heisenberg uncertainty relation for energy and time, Rep. Math. Phys. **6**, 361 (1974).

[235] J. Kijowski, Comment on "Arrival time in quantum mechanics" and "Time of arrival in quantum mechanics", Phys. Rev. A **59**, 897 (1999).

[236] V. Delgado and J. G. Muga, Arrival time in quantum mechanics, Phys. Rev. A **56**, 3425 (1997).

[237] A. Ruschhaupt and R. F. Werner, Quantum mechanics of time, in *The Message of Quantum Science: Attempts Towards a Synthesis*, edited by P. Blanchard and J. Fröhlich (Springer, 2015) Chap. 14.

[238] R. Werner, Screen observables in relativistic and nonrelativistic quantum mechanics, J. Math. Phys. **27**, 793 (1986).

[239] Y. Aharonov, J. Oppenheim, S. Popescu, B. Reznik, and W. G. Unruh, Measurement of time of arrival in quantum mechanics, Phys. Rev. A **57**, 4130 (1998).

[240] T. Jurić and H. Nikolić, Arrival time from the general theory of quantum time distributions, Eur. Phys. J. Plus **137**, 631 (2022).

[241] Y. Aharonov and T. Kaufherr, Quantum frames of reference, Phys. Rev. D **30**, 368 (1984).

[242] Y. Aharonov, S. Popescu, and J. Tollaksen, Each instant of time a new universe, in *Quantum Theory: A Two-Time Success Story* (Springer, 2014) pp. 21–36.

[243] C. Rovelli, Relational quantum mechanics, Int. J. Theor. Phys. **35**, 1637 (1996).

[244] M. Reisenberger and C. Rovelli, Spacetime states and covariant quantum theory, Phys. Rev. D **65**, 125016 (2002).

[245] D. N. Page and W. K. Wootters, Evolution without evolution: Dynamics described by stationary observables, Phys. Rev. D **27**, 2885 (1983).

[246] L. Maccone and K. Sacha, Quantum measurements of time, Phys. Rev. Lett. **124**, 110402 (2020).

[247] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum time, Phys. Rev. D **92**, 045033 (2015).

[248] R. Brunetti, K. Fredenhagen, and M. Hoge, Time in quantum physics: From an external parameter to an intrinsic observable, Found. Phys. **40**, 1368–1378 (2009).

[249] N. Vona and D. Dürr, The role of the probability current for time measurements, in *The Message of Quantum Science: Attempts Towards a Synthesis*, edited by P. Blanchard and J. Fröhlich (Springer, 2015) Chap. 5.

[250] S. Das and D. Dürr, Arrival time distributions of spin-1/2 particles, Sci. Rep. **9**, 2242 (2019).

[251] C. R. Leavens, Time of arrival in quantum and Bohmian mechanics, Phys. Rev. A **58**, 840 (1998).

[252] A. Ananthaswamy, Can we gauge quantum time of flight?, Sci. Am. **326**, 1 (2022).

[253] M. Kozuma, L. Deng, E. W. Hagley, J. Wen, R. Lutwak, K. Helmerson, S. L. Rolston, and W. D. Phillips, Coherent splitting of Bose-Einstein condensed atoms with optically induced bragg diffraction, Phys. Rev. Lett. **82**, 871 (1999).

[254] S. Pandey, H. Mas, G. Drougakis, P. Thekkeppatt, V. Bolpasi, G. Vasilakis, K. Poulios, and W. von Klitzing, Hypersonic Bose–Einstein condensates in accelerator rings, Nature **570**, 205 (2019).

[255] R. Shankar, *Principles of Quantum Mechanics* (Springer, 1994).

[256] C. R. Leavens, Spatial nonlocality of the "standard" arrival-time distribution, Phys. Lett. A **338**, 19 (2005).

[257] C. R. Leavens, On the "standard" quantum mechanical approach to times of arrival, Phys. Lett. A **303**, 154 (2002).

[258] S. Das and M. Nöth, Times of arrival and gauge invariance, Proc. R. Soc. A: Math. Phys. Eng. Sci. **477**, 2250 (2021).

[259] I. L. Egusquiza, J. G. Muga, B. Navarro, and A. Ruschhaupt, Comment on: "On the standard quantum-mechanical approach to times of arrival", Phys. Lett. A **313**, 498 (2003).

[260] C. R. Leavens, Reply to Comment on: "On the 'standard' quantum-mechanical approach to times of arrival" [Phys. Lett. A 313 (2003) 498], Phys. Lett. A **345**, 251 (2005).

[261] J. Kijowski, Arrival time in quantum mechanics (Demonstrated in geometrical order), Acta Phys. Pol. A **143**, S123–S130 (2023).

[262] A. J. Bracken and G. F. Melloy, Probability backflow and a new dimensionless quantum number, J. Phys. A: Math. Theor. **27**, 2197 (1994).

[263] M. Beau and L. Martellini, Quantum delay in the time of arrival of free-falling atoms, Phys. Rev. A **109**, 10.1103/physreva.109.012216 (2024).

[264] M. Beau, M. Barbier, R. Martellini, and L. Martellini, Time-of-arrival distributions for continuous quantum systems (2024), arXiv:2405.02018 [quant-ph] .

[265] K. V. Kuchar, Time and interpretations of quantum gravity, Int. J. Mod. Phys. D **20**, 3 (2011).

[266] J. Leon and L. Maccone, The Pauli objection, Found. Phys. **47**, 1597–1608 (2017).

[267] B. S. DeWitt, Quantum theory of gravity. I. The canonical theory, Phys. Rev. **160**, 1113 (1967).

[268] M. Porrmann, Particle weights and their disintegration I, Commun. Math. Phys. **248**, 269–304 (2004).

[269] P. A. Höhn, A. R. H. Smith, and M. P. E. Lock, Trinity of relational quantum dynamics, Phys. Rev. D **104**, 066001 (2021).

[270] R. Gambini and J. Pullin, The solution to the problem of time in quantum gravity also solves the time of arrival problem in quantum mechanics, New J. Phys. **24**, 053011 (2022).