POLITECNICO DI TORINO
Repository ISTITUZIONALE

Applications in Quantum Hypothesis Testing

(Article begins on next page)

12 June 2023

Doctoral Dissertation
Doctoral Program in Physics (35thcycle)

Applications in Quantum Hypothesis Testing
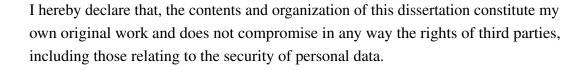
By

Giuseppe Ortolano
******

Supervisor(s):
Dr. Marco Genovese, Supervisor
Dr. Ivano Ruo-Berchera, Co-Supervisor

Reviewers:
Prof. Stefano Olivares, *University of Milan*
Prof. Maria Bondani, *University of Insubria*

Politecnico di Torino
2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

<div align="right">

Giuseppe Ortolano

2023

</div>

I would like to dedicate this work to my family and friends

Abstract

This Thesis work analyzes different protocols in the context of photonic quantum sensing. Quantum sensing in the optical and photonic domain is a developed field of quantum technologies that aims at improving the measurements beyond the performance of conventional techniques by exploiting quantum states of light. Our work finds its theoretical basis in Quantum Hypothesis Testing (QHT), a field that studies the discrimination among a finite set of hypotheses codified on a quantum state, by performing a measurement on it. While a rich field from a theoretical point of view, QHT counts few experimental realizations, with some notable exceptions, such as the ones related to Quantum Illumination. With the aim to partially fill this gap we present, as original results, a more applicative approach and experimental proof-of-principles to some QHT protocols, that could open interesting perspectives for real applications. The first result is the experimental realization of the Quantum reading (QR) protocol. The original QR proposal showed how quantum-correlated optical states of light can enhance the readout of classical digital data, stored in optical memories (an example are DVDs), when compared to classical sensing benchmarks. We have also demonstrated for the first time a quantum advantage in a QHT protocol applied to the monitoring of production processes, namely the identification of a deviation of the distribution of the end-products from a reference, a protocol that we labeled as Quantum Conformance Test (QCT). We finally discuss the more complex problem of pattern recognition. Pattern recognition is the task aimed at sorting images in predetermined classes, an example being the recognition of handwritten digits. The classification is done by classical processing, ranging from simple algorithms to sophisticated machine learning techniques. Regardless of the classification method, the pattern recognition performance is heavily influenced by the reliability of the images to be classified. In our analysis we demonstrate how, using the quantum sensing techniques discussed in the thesis, one can achieve a notable advantage in the recognition task, sometimes with a great amplification

of the pure sensing advantage. We validate our results experimentally showing the scalability of the quantum enhancement with the complexity of the application, not only in theory, but in practical conditions. We show how these results can be achieved using photonic correlated states, that can be routinely produced in laboratories and photon counting (PC) measurements, performed by commercial detectors. All of the protocols discussed show good resistance to experimental imperfections, notably to optical losses, a limiting factor for most of the quantum schemes. This, in conjunction with the relative simplicity of the experimental approach adopted in the realizations, strongly suggests that the protocols discussed are very promising avenues for near-term technological applications.

Contents

List of Figures

Chapter 1

Introduction

Recent years have seen substantial interest in research aimed at using quantum systems to overcome classical limitations. This interest has been mainly motivated by the great potential for revolutionary technological applications. It is, in fact, widely accepted that we are in the middle of a *second quantum revolution*. The first quantum revolution dates back to the middle of the last century. The understanding and deep analysis of the early stages of quantum theory led to ground-breaking technological innovations, such as the transistor, atomic clocks, GPS and lasers. The premises of the second quantum revolution are slightly different. Rather than using quantum theory to better understand the rules governing technological devices, the aim in this new research topic is to control and engineer quantum systems, down to the single particles, and pure quantum features, in order to develop innovative solutions to technological challenges. This newfound interest has led to a race to investments in order to secure the leadership in this promising field. In Europe this resulted in the "Quantum Flagship" with investments ranging in the order of several billions of euros. In its manifesto the flagship identifies four pillars for *quantum technologies*, namely: quantum *computation*, *communication*, *simulations* and *sensing*. The biggest private high-tech companies are now also investing a significant amount of resources in quantum technologies, and small quantum themed start-ups are flourishing all over the world.

Quantum computation is arguably one of the most suggestive and the most known field to the general audience. The first seminal theoretical ideas on quantum computation date back to the early 70s. The power of quantum computation resides

in the fact that classical bits are replaced by quantum bits (qubits), that unlike their classical counterpart can access *superposition* states and be *entangled* with each other, opening up a great amount of computational power. These unique properties allow quantum computers to, in theory, solve problems in minutes that would require, even for classical super computers an excessive amount of time, in some instances in the range of thousands of years or more. In the last decades moderately sized quantum computer have been realized on different physical platforms, the most developed of which are arguably superconducting qubits, used among others by IBM and Google. A big motivation in the push to achieve reliable quantum computing is that Moore's law [4] is approaching its end as the components get smaller and the scales that are being approached are more and more affected by quantum effect inducing noise in classical designs.

Quantum simulations can be seen as the analog counterpart of quantum computation. In practice, quantum simulators use quantum systems whose parameters are experimentally tunable and controllable to simulate the behaviour of another quantum system of interest on which one does not have enough control. They are convenient in many scenarios since they are generally easier to build and require much less control over the single components of the quantum systems when compared with quantum computers. The realization of good quantum simulations can be of huge importance for the development, both theoretical and applicative, of computational chemistry and material sciences, with consequently a big potential for technological innovations in a variety of fields.

Quantum communication [5–7] is very closely connected to the advances achieved in quantum computing. Security in communication is a very important requirement in the modern world both for private and public use. Most classical cryptographic schemes, in use today, are based on computationally hard problems that act as a "lock", relying on the fact that solving those problem without additional information, or a "key", would require classical computers a computational time larger than the expected duration of the data sensitivity. In fact, the main threat to the existing encryption algorithms comes from quantum computing. It has been shown by Peter Shor in 1994 [8] that a specific algorithm running on a quantum computer can easily break all the present cryptographic algorithms based on the factorization in prime numbers of large integers, suggesting that also other computational encryptions may be at risk. A possible solution is performing the encryption on quantum states, in a way that the security is guaranteed directly by pure quantum theory principles, as

opposed to computational complexity, meaning that these quantum protocols are secure from any kind of computational attack, quantum or not, by construction. The research in this latter field has now reached a high Technology Readiness Level (TRL) with devices for secure quantum communication already in commerce. The main challenge in this field is now to find ways to overcome the technical limitation arising when communication is performed over large distances, with the goal to create a global communication network. Quantum cryptography [9] is a sub-field of quantum communication, for which the more general aim is to find the amount of information that can be reliably transmitted over quantum channels.

Finally, the last pillar of quantum technologies is quantum sensing. Quantum sensing encompasses a wide variety of research topics and it is probably one of the most mature fields as well as one of the most promising to deliver near term technology. It is generally considered Quantum sensing [10, 11] any instance in which a quantum object, having quantized energy levels, or quantum coherence is used to measure a physical quantity, with notable examples given by atomic clocks and Nitrogen-Vacancy (NV) centers. Another instance of Quantum sensing is given by the use of quantum entanglement and correlations to improve the precision of measurements beyond classical limits [12]. In particular *photonic quantum sensing* encompasses all the mentioned tasks using quantum states of light. *Quantum metrology* and *quantum hypothesis testing* are notable sub-fields of Quantum sensing. Their main goal is to show an advantage over all possible classical strategies. Quantum metrology [13–15] is concerned with the the estimation of continuous physical parameters encoded in quantum states, for example as a result of the probing of a sample. In this field a practical *quantum advantage* has been shown in a variety of tasks including sub-shot noise microscopy [16, 17], quantum imaging [18–22], super-resolution [23–25], spectroscopy [26, 27], phase measurements and interferometry [28–35], and quantum states have also been used to enhance the performance of one of the most sophisticated instrument for fundamental physics testing, the large scale interferometer used to detect gravitational waves [36]. Quantum hypothesis testing (QHT) [37], on the other hand, is focused on the discrimination among a finite set of hypotheses and will be the main topic of this thesis work. The most basic formulation of QHT is given by the discrimination of two quantum states [38–40], pioneered by Carl Helstrom in his seminal works [37]. Most modern protocols are however formulated in the context of quantum channel discrimination [41, 42], where some object, containing information on the hypothesis is tested by quantum

states and modeled as a quantum channel, i.e. a linear map on the input states. A very well-known protocol in QHT is quantum illumination [43], in which the use of quantum light is shown to enhance the detection of target in a strongly noisy background. This protocol has received significant interest over the years because it may lead to the realization of quantum enhanced radar and lidar. Substantial progress has been made over the original proposal [44–47]. Another notable protocol is Quantum Reading [48], which has been shown to dramatically improve the readout of classical digital memories using entangled bipartite states. While many other theoretical protocols have been proposed in this field, compared to the achievement of quantum metrology, experimental realizations are scarce in QHT, with the remarkable exception of quantum illumination [45, 49, 50]. In this thesis work, in an attempt to partially address this void, we present some progress in this sense. An experimental realization of the quantum reading protocol, in Ref. [1], is exposed in Chapter 3 as an original contribution. We will also present, in chapter 4, the Quantum conformace test, introduced in Ref. [2], a generalization of quantum reading, that can find various applications [2, 3], for example to enhance production process monitoring. Finally, we will present an experimental demonstration of a quantum enhanced sensing protocol for the task of pattern recognition [51].

The thesis is structured as follows:

- In **Chapter 2** we will introduce some important theoretical tools in quantum hypothesis testing, namely the Helstrom bound, the quantum Chernoff bound and the Holevo bound. These tools will be used then in the following chapters to derive some of the main results.

- In **Chapter 3** we will mainly discuss the quantum reading (QR) protocol. We will start by discussing the original formulation of the QR protocol as given in Ref. [48] also discussing the results. We will than present our experimental realization following Ref.[1].

- In **Chapter 4** the Quantum Conformance Test protocol will be reviewed, following our original proposal in Ref. [2]. We will then discuss its experimental implementation. We will conclude this chapter by showing how the formalism introduced by the QCT protocol can be also used to describe the readout of an imperfect digital memory.

- In **Chapter 5** we will show how the sensing techniques introduced in the previous chapter can be used to improve the overall performance of a complicated and practical task, namely pattern recognition.

- In **Chapter 6** we will summarize and discuss the results presented in the previous chapters and give some final remarks.

Chapter 2

Theoretical tools in Quantum Hypothesis Testing

In the field of quantum information, recovering information encoded in a state is a central task. The information can be encoded in a continuous parameter as it is the case in quantum metrology. Another important scenario arises when information is encoded in a discrete collection of quantum states. The most basic formulation of this problem, and the one that finds more practical applications, is the binary case, in which one bit of information is encoded in two possible quantum states. The problem of discriminating two quantum states is very interesting from a theoretical standpoint due to the non-trivial geometry of the Hilbert spaces in which the states are defined. Two general non-orthogonal quantum states cannot be distinguished without ambiguity even if no noise is present, a purely quantum feature. In the classical domain a non-trivial counterpart to quantum state discrimination can be found in the discrimination of probability distributions arising in the setting of stochastic signals. Within classical information, *classical hypothesis testing* is a very well developed field, from which quantum information has drawn plenty of inspiration. The discrimination of two quantum states with the minimum probability of error is the most basic instance of what is generally known as *quantum hypothesis testing*. Another possible approach, apart from the minimum error one, is unambiguous quantum state discrimination in which inconclusive results are allowed in exchange for the certainty of correctly guessing the state when a given outcome occurs. We won't discuss this latter approach in this thesis, but in the next section, we will

briefly mention the result stating that the failure probability in unambiguous state discrimination is at least twice as big as the optimal probability of error.

One of the main contributors to the field of quantum state discrimination/ quantum hypothesis testing is Carl W. Helstrom. Among his broad contribution to the field of quantum information, between the most notable there is the derivation of the *Helstrom Bound* an ubiquitous quantity in quantum information describing the best achievable performance in the discrimination of two arbitrary quantum states, ρ_0 and ρ_1. The next section will be dedicated to the derivation of the Helstrom bound. After this we will discuss another useful tool in quantum hypothesis testing, the *Quantum Chernoff bound*, concerned with the asymptotic behaviour of the probability of error in the discrimination of multi-copy quantum states. Finally, we will discuss another limit, central in quantum communication, the Holevo bound, a bound in the amount of classical information that can be reliably transmitted using quantum states, in the limit of a large number of repeated communications.

2.1 Optimal discrimination of two quantum states: The Helstrom bound

A quantum hypothesis testing problem is formally described as the discrimination of two hypothesis after the measurement on an unknown state, described by the density operator ρ, with the minimum probability of error. Binary state discrimination is the most basic quantum hypothesis testing problem and is given by the hypothesis $\mathcal{H}_0 : \rho = \rho_0$ and $\mathcal{H}_1 : \rho = \rho_1$. The states ρ_0 and ρ_1 are in general mixed and non-orthogonal. Each of the states is prepared with a known prior probability $p(\mathcal{H}_i) := \pi_i$, with the condition $\sum_i \pi_i = 1$. After performing a general Positive Operator Valued Measurement (POVM) on the incoming state ρ the goal is to minimize the probability of incorrectly guessing the hypothesis. Let us start by briefly reviewing POVMs.

A POVM is defined by a set of N measurement operators $\{\Pi_k\}$ each one referring to a different measurement outcome labelled k. These operators are defined such that the probability of having the outcome k when measuring a state ρ is $\text{Tr}(\rho \Pi_k)$. The fact that the probability must be real and non-negative implies that the operators Π_k must be Hermitian positive semi-definite. POVMs are generally required to be complete, i.e. $\sum_k^N \Pi_k = \mathbb{1}$. Note how any incomplete set of measurement operators

$\{\Pi_k\}$, can be completed to form a POVM by adding to the set the operator $\Pi_{N+1} = \mathbb{1} - \sum_k^N \Pi_k$.

Back to the problem of hypothesis testing, if we consider the discrimination of N states, $\{\rho_k\}$ using the POVM $\{\Pi_k\}$ with N elements, where the outcome k is taken as corresponding to the selection of the state ρ_k, we have the probability of error:

$$p_{err} = 1 - p_{suc} = 1 - \sum_k^N \pi_k \mathrm{Tr}(\rho_k \Pi_k) = \sum_k^N \pi_k \sum_{l \neq k} \mathrm{Tr}(\rho_k \Pi_l) \qquad (2.1)$$

Where $p_{suc} = 1 - p_{err}$ is the probability of correctly guessing the hypothesis and we used the completness of the POVM that, in terms of probability, gives the condition $\sum_k^N (\rho \Pi_k) = 1$ for any normalized state ρ. The term $\mathrm{Tr}(\rho_k \Pi_k)$ gives the probability of correctly guessing that the state was prepared in ρ_k, so the total probability of success of the test, p_{suc} is provided by the average of the probability of a correct guess for all the states ρ_k weighted by their prior probability π_k. Equivalently the probability of error is given by the weighted average of the probability of an incorrectly guessing given the state is prepared in ρ_k, given by the term $\sum_{l \neq k} \mathrm{Tr}(\rho_k \Pi_l)$. For the particular case of binary state discrimination Eq.(2.1) reduces to:

$$p_{err} = 1 - \sum_{k=0}^1 \pi_i \mathrm{Tr}(\rho_i \Pi_i) = \pi_0 \mathrm{Tr}(\rho_0 \Pi_1) + \pi_1 \mathrm{Tr}(\rho_1 \Pi_0) \qquad (2.2)$$

In the following we will give a non-variational derivation [40] of the *Helstrom formula* based on the introduction of the Hermitian operator Λ, sometimes called the *Helstrom operator* (or Helstrom matrix):

$$\Lambda = \pi_1 \rho_1 - \pi_0 \rho_0 \qquad (2.3)$$

The probability of error in binary discrimination, given in Eq.(2.2), can be rewritten, in terms of Λ, as:

$$p_{err} = \pi_0 + \mathrm{Tr}(\Lambda \Pi_0) = \pi_1 - \mathrm{Tr}(\Lambda \Pi_1) \qquad (2.4)$$

Introducing the eigenvalues and eigenvectors of the Helstrom operator, defined by $\Lambda = \sum_k \lambda_k |\phi_k\rangle\langle\phi_k|$, we can rewrite Eq.(2.4) as:

$$p_{err} = \pi_0 + \sum_k \lambda_k \langle \phi_k | \Pi_0 | \phi_k \rangle = \pi_1 - \sum_k \lambda_k \langle \phi_k | \Pi_1 | \phi_k \rangle \qquad (2.5)$$

Eq.(2.5) has to be minimized under the conditions $0 \leq \langle \phi_k | \Pi_i | \phi_k \rangle \leq 1$, that follows from the fact that $\langle \phi_k | \Pi_i | \phi_k \rangle = \mathrm{Tr}(|\phi_k\rangle\langle\phi_k|\Pi_i)$ is a probability. The minimization is achieved by imposing $\langle \phi_k | \Pi_0 | \phi_k \rangle = 0$ and $\langle \phi_k | \Pi_1 | \phi_k \rangle = 1$ for all the positive eigenvalues λ_k and $\langle \phi_k | \Pi_0 | \phi_k \rangle = 1$, $\langle \phi_k | \Pi_1 | \phi_k \rangle = 0$ for the negative eigenvalues. The operators Π_0 and Π_1 that fulfill these requirements are:

$$\Pi_1 = \sum_{k \in K^+} |\phi_k\rangle\langle\phi_k|$$

$$\Pi_0 = \sum_{k \in K^-} |\phi_k\rangle\langle\phi_k|$$

where $K^+ = \{k | \lambda_k > 0\}$ and $K^- = \{k | \lambda_k \leq 0\}$ are the sets of indexes k such that the eigenvalues of Λ are positive or non-positive respectively. In other words Π_1 and Π_0 are the projectors onto the positive, Λ^+, and non-positive, Λ^-, parts of the operator Λ respectively. Clearly the indexes with eigenvalues $\lambda_k = 0$ can be assigned to either of the sets without changing the probability of error and the choice of assigning them to the set K^- in this instance was arbitrary and without influence on the final results. Using these definitions we can rewrite the probability of error in Eq.(2.5) as:

$$p_{err} = \pi_0 - \sum_{k \in K^-} |\lambda_k| = \pi_1 - \sum_{k \in K^+} |\lambda_k| \qquad (2.6)$$

Adding the two expressions for the probability of error in the equation above, and using the fact that $\pi_0 + \pi_1 = 1$, we get:

$$p_{err} = \frac{1}{2}(1 - \sum_k |\lambda_k|) \qquad (2.7)$$

that using the definition of Λ in Eq.(2.3)yelds the Helstrom formula for the optimal probability of error in binary state discrimination:

$$p_{err} = \frac{1}{2}(1 - ||\pi_1\rho_1 - \pi_0\rho_0||) \qquad (2.8)$$

where we introduced the norm $||A|| = \mathrm{Tr}(\sqrt{A^\dagger A})$. In the special case of initially pure states, $\rho_i = |\psi_i\rangle\langle\psi_i|$, the Helstrom formula simplifies greatly, since the trace distance $||\pi_1\rho_1 - \pi_0\rho_0||$ can be computed in terms of the overlap $\langle\psi_0|\psi_1\rangle$. For pure states we

have:

$$p_{err} = \frac{1}{2}\left(1 - \sqrt{1 - 4\pi_0\pi_1|\langle\psi_0|\psi_1\rangle|^2}\right) \qquad (2.9)$$

Eq.(2.9) will be used, for instance, in the next chapter to compute an analytical expression for the classical limit in the protocol of quantum reading.

We conclude this section with a brief overview of unambiguous state discrimination and a comparison of its performance with the minimum error discrimination one. In unambiguous state discrimination the goal is to create a scheme in which given outcomes identify quantum states, in a generally non-orthogonal set, with certainty. Since non-orthogonal states cannot be discriminated with certainty, unambiguous discrimination can only be done at the price of allowing some outcome to be inconclusive. The concept is easily visualized in terms of binary discrimination of pure states. Consider the two generally non-orthogonal states $|\psi_0\rangle$ and $|\psi_1\rangle$. Let us consider the projective measurement with elements $\{P_0, P_1\}$, where $P_0 = |\psi_0\rangle\langle\psi_0|$ is the projector onto $|\psi_0\rangle$ and $P_1 = \mathbb{1} - P_0$ is the projector onto its orthogonal subspace. If this measurement is performed and the outcome is P_1 it is clear that the initial state was prepared with certainty in $|\psi_1\rangle$, since $|\psi_0\rangle$ would give always the outcome P_0. On the other hand, if the outcome of the measurement is P_0, the initial state could have been prepared in either $|\psi_0\rangle$ or $|\psi_1\rangle$ and no definitive answer can be given. Similarly, $|\psi_0\rangle$ can be identified with certainty by switching the role of the states in the previous description. The scheme outlined, based on projective measurements is not optimal in general, in terms of minimizing the probability of having an inconclusive result. A better strategy is given by a POVM, of which we skip the details (see e.g. Ref. [40]), mixed with projective measurement in case of very unbalanced priors, whose probability of success (certain result), p_{cer}, in case of equal priors, $\pi_0 = \pi_1 = 1/2$ is given in terms of the overlap of the states as:

$$p_{cer} = 1 - p_f = 1 - |\langle\psi_0|\psi_1\rangle| \qquad (2.10)$$

Where p_f is the probability of the procedure failing. Unambiguous discrimination in the case of mixed state is rarely effective. If two mixed states ρ_0 and ρ_1 have the same support it is not possible to discriminate them with certainty. A necessary condition to be able to unambiguously discriminate two mixed state is that their

supports have a non-zero overlap with the kernel of each other[1]. In other words two mixed states can be distinguished with certainty only if it exist at least a vector, in the Hilbert space in which the states are defined, that can occur in only one of the two states. For mixed states such that unambiguous discrimination can be performed, the following inequality relating the probability of error in minimum discrimination, p_{err}, and the probability of failure in unambiguous discrimination, p_f, holds [40]:

$$p_{err} \leq \frac{1}{2} p_f \qquad (2.11)$$

meaning that even when unambiguous discrimination is possible the probability of failure of the discrimination is at least twice the probability of error in the optimal discrimination.

2.2 Asymptotic efficient discrimination: The Quantum Chernoff Bound

The Helstrom bound, discussed in the previous section, gives the optimal probability of error, p_{err}^H when discriminating two states ρ_0 and ρ_1. Consider now the asymptotic situation in which the states to be discriminated are composed of n copies, i.e. $\rho_i^T = \rho_i^{\otimes n} = \rho_i \otimes ... \otimes \rho_i$. According to the results of the previous section the best probability of error in the discrimination is $p_{err}^H = (1 - ||\pi_1 \rho_1^{\otimes n} - \pi_0 \rho_0^{\otimes n}||)/2$. For large values of n the trace distance in the probability of error, in a lot of cases, becomes tricky to evaluate, even numerically. For this reason, some bounds in this asymptotic scenario are desirable. An upper bound on p_{err}^H is given by the *Quantum Chernoff Bound*. In this section we will start by discussing the classical Chernoff bound[52] and then we will show the derivation of its quantum counterpart.

2.2.1 Classical Chernoff bound

A central task in classical information theory is the discrimination of two probability distributions after a given number n of observations. In Ref. [53] Chernoff derived a series of limits on the asymptotic tail probabilities of observed distributions.

---

[1]The *support* of a state is the subspace in which the state is defined, while its *kernel* it's the subspace orthogonal to the support

From these limits one can derive an informational quantity, commonly known as Chernoff information ξ_{CB}. Notably, given two classical distributions, the probability of error in discriminating them after n observation, $p_{err}^{(n)}$ in the asymptotic regime decreases exponentially at a rate given by the Chernoff information, namely $p_{err}^{(n)} \sim \exp(-n\xi_{CB})$.

Formally, let us consider the classical hypothesis testing problem of discriminating between hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ according to an observation x of the classical random variable X. This could represent, for example, the situation in which one tries to distinguish a given signal from noise fluctuations. Define the prior probabilities as $p(\mathcal{H}_i) := \pi_i$ and the conditional probabilities $p_i(x) := p(x|\mathcal{H}_i)$. In terms of these quantities the total probability of having observation x is $p(x) = \pi_0 p_0(x) + \pi_1 p_1(x)$. The best possible decision rule is the Bayes rule. With this rule the hypothesis i is selected as $i = \arg\max_i p(\mathcal{H}_i|x)$. We can use Bayes theorem to express the posteriori probabilities $p(\mathcal{H}_i|x)$ in terms of the priors and conditional probabilities:

$$p(\mathcal{H}_i|x) = \frac{\pi_i p_i(x)}{p(x)} \tag{2.12}$$

Bayes rules minimizes the probability of error in the decision, p_{err}, by construction. If the Bayes decision rule is used the probability of error is:

$$p_{err} = \int \min\left(p(\mathcal{H}_0|x), p(\mathcal{H}_1|x)\right) p(x) dx =$$
$$= \int \min\left(\pi_0 p_0(x), \pi_1 p_1(x)\right) dx \tag{2.13}$$

The expression in Eq. (2.13) can be upper bounded using the following inequality, true for any $a, b > 0$ [52]:

$$\min(a, b) \le a^\alpha b^{1-\alpha} \qquad\qquad \forall \alpha \in [0, 1] \tag{2.14}$$

That for Eq. (2.13) gives:

$$p_{err} \le \pi_0^\alpha \pi_1^{1-\alpha} \int p_0(x)^\alpha p_1(x)^{1-\alpha} dx \tag{2.15}$$

If we consider a series of n i.i.d. observations $\mathbf{x} = x_1, ..., x_n$ we have $p_i(\mathbf{x}) = \prod_k p_i(x_k)$. The probability of error in the decision after n observations, $p_{err}^{(n)}$, can be upper

bounded as:

$$p_{err}^{(n)} \leq \pi_0^\alpha \pi_1^{1-\alpha} \int \prod_k p_0(x_k)^\alpha p_1(x_k)^{1-\alpha} d\mathbf{x} =$$
$$= \pi_0^\alpha \pi_1^{1-\alpha} \prod_k \int p_0(x_k)^\alpha p_1(x_k)^{1-\alpha} dx_k =$$
$$= \pi_0^\alpha \pi_1^{1-\alpha} \left(\int p_0(x)^\alpha p_1(x)^{1-\alpha} dx \right)^n \qquad (2.16)$$

Taking the logarithm on both sides:

$$\frac{1}{n} \log(p_{err}^{(n)}) \leq \frac{1}{n} \log(\pi_0^\alpha \pi_1^{1-\alpha}) + \log \left(\int p_0(x)^\alpha p_1(x)^{1-\alpha} dx \right) \overset{n\to\infty}{\sim}$$
$$\sim \log \left(\int p_0(x)^\alpha p_1(x)^{1-\alpha} dx \right) \qquad (2.17)$$

The first term on the right-hand side can be neglected as it tends to zero in the asymptotic limit of large n, as long as the priors are not 0. The remaining quantity on the right hand side of Eq. (2.17) depends on the parameter α and the inequality holds true for any $\alpha \in [0,1]$. Taking the minimum of this quantity in this range gives the definition of the *Chernoff information* ξ_{CB}:

$$\xi_{CB} := - \min_{\alpha \in [0,1]} \log \left(\int p_0(x)^\alpha p_1(x)^{1-\alpha} dx \right) \qquad (2.18)$$

The Chernoff information is a measure of similarity of the probability distributions p_0 and p_1. Rewriting Eq. (2.17) in terms of the Chernoff information gives the *classical Chernoff bound*:

$$p_{err}^{(n)} \leq \pi_0^{\alpha^*} \pi_1^{1-\alpha^*} e^{-n\xi_{CB}} \overset{n\to\infty}{\sim} e^{-n\xi_{CB}} \qquad (2.19)$$

where α^* is the value of α that minimizes Eq.(2.18).

2.2.2   Quantum Chernoff bound

The classical problem of discriminating two distributions given a large number of observations has a natural generalization in the quantum domain when considering the discrimination of two states in the limit of measurements performed over a large number of copies. A quantum generalization of the Chernoff bound was first presented in Ref. [54]. Similarly to the classical case, an informational quantity, the

quantum Chernoff information, ξ_{QCB}, related to the distance of the states considered, can be defined such that the probability of error in the discrimination, $p_{err}^{(n)}$, decreases exponentially as the number n of copies increases, at a rate given by the Quantum Chernoff information, $p_{err}^{(n)} \sim \exp(-n\xi_{QCB})$. The quantum Chernoff bound has since found use in the characterization of a variety of quantum hypothesis testing protocols, for example, the quantum reading one, discussed in the next chapter, and it's particularly useful, as already mentioned, in situations in which a direct calculation of the trace distance in computationally heavy and the Helstrom bound is not easily accessible.

Consider the problem of the optimal discrimination of two many-copy quantum states $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$, prepared with prior probabilities π_0 and π_1 respectively. According to the discussion of the previous section, the best probability of error in the discrimination, $p_{err}^{(n)}$ is given by the Helstrom formula, namely in this case:

$$p_{err}^{(n)} = \frac{1 - ||\pi_1 \rho_1^{\otimes n} - \pi_0 \rho_0^{\otimes n}||}{2} \tag{2.20}$$

The Quantum Chernoff bound is concerned with the asymptotic behaviour of $p_{err}^{(n)}$. The probability of error can be upper bounded using the following inequality, on positive operators A, B (for the proof of this inequality see Ref. [54]):

$$\mathrm{Tr}[A + B - |A - B|]/2 \leq \mathrm{Tr}[A^\alpha B^{1-\alpha}] \qquad \forall \alpha \in [0,1] \tag{2.21}$$

Eq.(2.21) plays in the quantum case the same role that Eq.(2.14) played in the classical derivation. The direct substitution $A = \pi_0 \rho_0^{\otimes n}$ and $B = \pi_1 \rho_1^{\otimes n}$ in Eq.(2.21) yields:

$$\frac{1 - ||\pi_1 \rho_1^{\otimes n} - \pi_0 \rho_0^{\otimes n}||}{2} = p_{err}^{(n)} \leq \mathrm{Tr}[(\pi_0 \rho_0^{\otimes n})^\alpha (\pi_1 \rho_1^{\otimes n})^{1-\alpha}] \tag{2.22}$$

Taking the logarithm of both sides leads to:

$$\frac{1}{n} \log\left(p_{err}^{(n)}\right) \leq \frac{1}{n} \log\left(\pi_0^\alpha \pi_1^{1-\alpha}\right) + \log\left(\mathrm{Tr}[\rho_0^\alpha \rho_1^{1-\alpha}]\right) \overset{n \to \infty}{\sim}$$
$$\sim \log\left(\mathrm{Tr}[\rho_0^\alpha \rho_1^{1-\alpha}]\right) \tag{2.23}$$

where once again after taking the limit $n \to \infty$ the quantity depending on the priors vanishes. Similarly to the classical case, the *Quantum Chernoff information* can be

defined taking the minimum of the right-hand side of Eq.(2.23):

$$\xi_{QCB} = - \min_{\alpha \in [0,1]} \log \left(\mathrm{Tr}[\rho_0^\alpha \rho_1^{1-\alpha}] \right) \tag{2.24}$$

It is worth noticing how, if ρ_0 and ρ_1 commute, the definition of ξ_{QCB} coincides with
the classical Chernoff information $\xi_{CB} = - \min_{\alpha \in [0,1]} \log \left(\sum_k p_0(k)^\alpha p_1(k)^{1-\alpha} \right)$,
where p_0 and p_1 are the elements of ρ_0 and ρ_1 in their diagonal form, a desirable
property for a quantum generalization. In terms of the Quantum Chernoff information
we have then the following bound of the probability of error:

$$p_{err}^{(n)} \leq \pi_0^{\alpha^*} \pi_1^{1-\alpha^*} e^{-n\xi_{QCB}} \overset{n \to \infty}{\sim} e^{-n\xi_{QCB}} \tag{2.25}$$

where α^* is the argument minimizing Eq.(2.24). Eq.(2.25) defines the *Quantum
Chernoff Bound*. As already mentioned the quantum Chernoff bound is particular
interest in the asymptotic limit of $n \to \infty$. In this limit, the upper bound in Eq.
(2.25) can be saturated as proven in Ref. [55], where it was shown that the quantum
Chernoff information gives also a lower bound for the logarithm of the probability
of error.

2.3 Distinguishability and information transfer: The Holevo bound and the classical capacity

The Holevo bound [56] is a very useful tool in quantum communication. Its main role
is to upper bound the classical information that can be transmitted using quantum
states. The Holevo quantity χ appearing in the Holevo bound, is also the basis
of another central result in quantum information theory, the Holevo-Schumacher-
Westmoreland (HSW) theorem [57, 58], that gives an expression for the classical
product capacity of a quantum channel. Apart from the already broad application in
quantum communication these tools find applications in other fields such as quantum
hypothesis testing. For instance, in the next chapter the limit of information that can
be extracted by a large digital memory will be derived using the HSW theorem. In
the following we will give a derivation of the Holevo bound and the HSW theorem.

Consider a situation in which a party, that we call Alice, wants to send a classical
message to a receiving party, that we denote as Bob. Alice encodes the message in

an *alphabet* composed of a given number of n symbols. Formally Alice encodes each letter of the message in a classical random variable X, having n possible values, in the collection of quantum states $\{\rho_1, ..., \rho_n\}$ each prepared with probability p_i. Bob then receives the unknown state ρ_i and measures it with a generic POVM to assign a symbol y, associated to the random variable Y, to the signal received. A measure of how well Bob can infer Alice message is given by the *mutual information* $I(X:Y)$ [59]. In particular taking the maximum of the mutual information over all possible POVM defines Bob *accessible information*. In other terms the accessible information gives the number of bits that Bob receives for each bit Alice sent. The *Holevo bound* provides an upper bound on Bob accessible information:

$$I(X:Y) \leq \chi := S(\rho) - \sum_i p_i S(\rho_i) \tag{2.26}$$

$$\rho = \sum_i p_i \rho_i$$

where $S(\rho) = -\mathrm{Tr}(\rho \log \rho)$ is the von Neumann entropy. χ appearing on the right-hand side of the Holevo bound is called the Holevo χ quantity. To prove Eq. (2.26) let us call the quantum system Alice sends to Bob Q. Let us also introduce two auxiliary systems, a preparation one P having basis $\{|x\rangle\}$, $x = 1, .., n$, and a measurement one M having basis $\{|y\rangle\}$, $y = 1, .., n$. Consider the total system initially prepared in the state ρ^{PQM}:

$$\rho^{PQM} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0| \tag{2.27}$$

The measurement system represents the information Bob receives and before the measurement is performed it is set to a "resting" state $|0\rangle\langle 0|$. The role of the system M is to act as a register so that after a measurement is performed with outcome y its state is changed to $|y\rangle$. In this configuration the action of a POVM, having elements $\{\Pi_y\}$ performed by Bob on the system Q can be described in the joint system QM by the quantum operation $\mathscr{E}_M$ defined as follows:

$$\mathscr{E}_M(\rho \otimes |0\rangle\langle 0|) = \sum_y \sqrt{\Pi_y} \rho \sqrt{\Pi_y} \otimes |y\rangle\langle y| \tag{2.28}$$

To prove the Holevo bound we use some properties of the quantum mutual information. The quantum mutual information $\mathscr{I}(A:B)$, of two quantum systems A and B, is defined as:

$$\mathscr{I}(A:B) = S(A) + S(B) - S(A,B) \tag{2.29}$$

where $S(A,B)$ is the entropy of the state of joint system AB. The *quantum mutual information cannot be increased by quantum operations* and the following inequality holds:

$$\mathscr{I}(A:B) \leq \mathscr{I}(A:(B,C)) \tag{2.30}$$

stating the intuitive fact that discarding a system cannot increase mutual information.

For the system PQM defined above we denote with primes the system after the measurement $\mathscr{E}_M$ is performed, i.e. $PQM \overset{\mathscr{E}_M}{\rightarrow} P'Q'M'$. It is easy to verify that the following inequality holds:

$$\mathscr{I}(P':M') \leq \mathscr{I}(P:Q) \tag{2.31}$$

This inequality can be derived using the fact that M is initially in a product state with the other systems and consequently $\mathscr{I}(P:Q) = \mathscr{I}(P:(Q,M))$. Moreover the POVM measurement, $\mathscr{E}_M$, a quantum operation on QM, cannot increase the mutual information according to the property stated above, that gives the inequality $\mathscr{I}(P':(Q',M')) \leq \mathscr{I}(P:(Q,M))$. Finally using Eq. (2.30) we have $\mathscr{I}(P':(Q',M')) \geq \mathscr{I}(P':M')$.

For the right hand side of Eq. (2.31) we have, using Eq. (2.29) and Eq. (2.27):

$$\mathscr{I}(P:Q) = S(P) + S(Q) - S(P,Q) = H(p_x) + S(\rho) - [H(p_x) + \sum_x p_x S(\rho_x)] =$$

$$= S(\rho) - \sum_x p_x S(\rho_x) = \chi \tag{2.32}$$

Where $H(p_x) = -\sum_x p_x \log p_x$ is the classical Shannon entropy. The right-hand side of Eq. (2.31) is in fact the Holevo quantity χ. To conclude the derivation of the Holevo bound, the left hand side must be shown to be $I(X:Y)$. This can be done

with simple algebra. We have in fact:

$$
\begin{aligned}
\rho^{P'M'} = \mathrm{Tr}_{Q'}(\rho^{P'Q'M'}) &= \sum_x p_x |x\rangle\langle x| \otimes \sum_y \mathrm{Tr}(\sqrt{\Pi_y}\rho_x\sqrt{\Pi_y})|y\rangle\langle y| = \\
&= \sum_x \sum_y p_x |x\rangle\langle x| \otimes \mathrm{Tr}(\Pi_y\rho_x)|y\rangle\langle y| \\
&= \sum_x \sum_y p_x p(y|x)|x\rangle\langle x| \otimes |y\rangle\langle y| = \\
&= \sum_x \sum_y p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \qquad (2.33)
\end{aligned}
$$

Using the expression in Eq. (2.33) we have that $\mathscr{I}(P':M') = I(X:Y)$ thus proving the Holevo bound.

Holevo χ quantity appears in another very important result in quantum information, the HWS theorem that we are going to discuss in the following.

In information theory a very important quantity is the *capacity* of a channel. Consider a communication channel that over n uses can *reliably* transmit nR bits of information, where R is called the rate of information transfer. The capacity of the channel is defined as the maximum rate R at which bits of information can be reliably transmitted. In classical information theory, the capacity is a very important quantity when considering noisy communication. A central result for classical communication is the Shannon coding theorem, giving the capacity for discrete memoryless channels, based on the use of *codewords* to better encode information over a large number of uses, $n \to \infty$, of the channel.

In quantum information theory an important problem is to determine the *classical capacity* of a quantum channel. In other words, one wants to determine how much classical information can be sent using quantum states transmitted through a possibly noisy quantum channel $\mathscr{E}$. As in the classical case, the classical capacity of a quantum channel is usually defined in the limit of many repeated uses of the channel, $n \to \infty$. A notable difference between the classical and quantum case is that in principle when considering quantum states over a large number of uses of the channel $\mathscr{E}$ the states sent across different uses can be entangled. A general expression for the classical capacity of a quantum channel, $C(\mathscr{E})$, allowing also entangled signals is not known. The HSW theorem gives, however, an expression for *the product state capacity*, usually denoted as $C^{(1)}(\mathscr{E})$, the capacity that can be achieved only using product states (across different uses of the channel), i.e. states of the form $\rho_1 \otimes ... \otimes \rho_n$. It

has been conjectured [59], but not still proven, that the product state capacity is the same as the capacity, $C(\mathscr{E}) = C^{(1)}(\mathscr{E})$.

Formally define $\chi(\mathscr{E})$ as:

$$\chi(\mathscr{E}) := \max_{\{p_i, \rho_i\}} \left[S\left(\mathscr{E}\left(\sum_i p_i \rho_i \right) \right) - \sum_i p_i S\left(\mathscr{E}(\rho_i) \right) \right] \qquad (2.34)$$

The HSW theorem states that $\chi(\mathscr{E})$ is the product capacity of the channel $\mathscr{E}$:

$$\chi(\mathscr{E}) = C^{(1)}(\mathscr{E}) \qquad (2.35)$$

The proof of the HSW theorem is based on a construction showing that in the limit of a large number of uses there exist an encoding of messages such that the capacity is saturated, however an encoding procedure that can be followed in all cases is not given. A formal proof is long and sometimes technical so we won't report it here (it can be found e.g. in Ref. [59]). As it was the case for the Holevo bound, $\chi(\mathscr{E})$ is a measure of distinguishabilility of the outputs of the channel. The capacity is defined as the maximization over all possible input, being a property only of the channel and not on the states one may use to perform communication. It is interesting to to point out that while one may be tempted to think that the capacity is saturated by states having orthogonal support, this is not always the case [59], as the linear map $\mathscr{E}$ acts not trivially on the geometry of the states and in some cases the capacity is saturated by non orthogonal sets.

Chapter 3

Quantum Reading and its experimental realization

A quantum channel $\mathscr{E}(\rho)$ is a positive linear map acting on a quantum state ρ. Quantum channels are the most general way to describe the evolution of an open quantum system in a given state ρ. In most practical cases quantum channels are also trace-preserving and completely positive, but in general those conditions are not necessary for evolution [59, 60]. Quantum channels play a central role in quantum information, especially in quantum communication. In recent years the problem of *quantum channel discrimination* has received a lot of attention mainly due to the many practical protocols proposed, having potential for technological applications. The most notable of those protocols is arguably Quantum Illumination, proposed by Seth Lloyd in Ref. [43], in which quantum light is used to enhance target detection in high noise setting.

From a theoretical point of view the problem of quantum channel discrimination is particularly interesting because it is a double optimization problem, both over the input state and output measurement. Due to this complexity most of the results in quantum channel discrimination are obtained in the form of upper and lower bounds using the tools described in the previous chapter for state discrimination, such as the Helstrom, Chernoff or Holevo bound.

In this chapter, we will discuss a notable protocol in quantum channel discrimination, *Quantum Reading*, with particular attention on its practical feasibility and experimental implementation. The Quantum Reading protocol is concerned with

the readout of classical information enhanced by the use of quantum resources and its experimental realization. The study and the realization of Quantum reading with feasible photon counting measurement is one of the main result of this thesis. Moreover, in the following chapters we will also discuss a generalization of the Quantum reading protocol, the Quantum Conformance Test, with applications to production process monitoring, as well as the application of quantum the quantum enhanced readout protocols to the practical task of pattern recognition.

3.1   Quantum Reading

The Quantum reading protocol was originally proposed in 2011 in Ref. [48]. Consider a classical optical digital memory composed of a large array of cells. A single memory cell can be seen as a beam splitter encoding a bit of information in one of two, equiprobable, possible values of transmittance (or equivalently reflectance) τ_0 or τ_1. The information is retrieved using a transmitter irradiating a collection of M bosonic signal modes in a state ρ over the memory cell. After the interaction with the cell the M modes are collected by a receiver that performs a generic POVM measurement. The outcome of this measurement is then classically processed to recover the bit of information stored in the cell. The state ρ can be, in general, bipartite. In other words, the procedure can be in general assisted by L idler modes, i.e. modes that do not interact with the cell, but are correlated with the signal ones. A schematic of the procedure is given in Fig. (3.1).

Since a single output of a beam splitter can be modelled by a bosonic pure loss channel, $\mathscr{E}_\tau$, in this formulation the retrieval of information from a memory cell is a quantum channel discrimination problem. Formally, let us store a bit $x = \{0, 1\}$ in a memory cell by means of two equiprobable pure loss channels, $\mathscr{E}_{\tau_0}$ and $\mathscr{E}_{\tau_1}$, with transmittances τ_0 and τ_1. A pure loss channel with transmittance τ corresponds to the following input-output transformation of the field operator $\hat{a} \rightarrow \sqrt{\tau}\hat{a} + i\sqrt{(1-\tau)}\hat{v}$, where $\hat{v}$ describes an environmental vacuum mode. To retrieve the bit, the transmitter irradiates M signal modes over the cell, for a total of μ mean photons, and also sends additional L idler modes directly to the output. The receiver measures the transmitted signal and idler modes, giving as an output a guess y of the classical bit x with an error probability p_{err}. The information recovered is given by the mutual information between the classical bit stored X and the bit recovered Y, $I(X : Y)$, that in terms of

the probability of error p_{err} can be expressed, for equiprobable channels, as [61]:

$$I(X:Y) = 1 - H(p_{err}) \tag{3.1}$$

where $H(p) = -p\log p - (1-p)\log(1-p)$ denotes the binary Shannon entropy. As can be seen from Eq. (3.1) the information recovered is maximized if the probability of error is minimum.

Since we are considering a bosonic state in an infinite Hilbert space, the minimization without energy constrains is trivial, since p_{err} would go to zero as μ approaches infinity. On the other hand, the problem of minimizing p_{err} over all transmitters with fixed mean signal photons μ is difficult to solve. The problem can be solved however if the analysis is restricted to a particular class of input states of interest, namely classical states. According to the quantum optics definition, classical states are all and only the states having a positive P-representation:

$$\rho^{cla} = \int d^{2M}\boldsymbol{\alpha}\, d^{2L}\boldsymbol{\beta}\, P(\boldsymbol{\alpha},\boldsymbol{\beta})\, |\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}| \otimes |\boldsymbol{\beta}\rangle\langle\boldsymbol{\beta}| \tag{3.2}$$

where $P(\boldsymbol{\alpha},\boldsymbol{\beta}) > 0$ is a probability distribution and:

$$|\boldsymbol{\alpha}\rangle = |\alpha_1\rangle \otimes ... \otimes |\alpha_M\rangle$$
$$|\boldsymbol{\beta}\rangle = |\beta_1\rangle \otimes ... \otimes |\beta_L\rangle$$

are M and L modes coherent states in the signal and idler system respectively. In other words classical states are defined as the class of *convex superpositions* of coherent states.

For classical states with fixed signal mean photons μ a lower bound on the probability of error that can be achieved in the discrimination, p_{err}^{cla} is given by:

$$p_{err}^{cla} \geq \mathscr{C}(\mu,\tau_0,\tau_1) := \frac{1 - \sqrt{1 - e^{-\mu(\sqrt{\tau_1}-\sqrt{\tau_0})^2}}}{2}. \tag{3.3}$$

We won't give here a derivation of $\mathscr{C}(\mu,\tau_0,\tau_1)$ since a similar limit in a more general configuration will be derived in Sec.(4.1.2) of the next chapter, and the limit in Eq. (3.3) can be found following the same steps.

The limit in Eq. (3.3) can be saturated without the aid of idler modes, by using a single mode signal coherent state $\rho_{cla} = |\alpha\rangle\langle\alpha|$, with $|\alpha|^2 = \mu$. This can be seen

using the fact that a pure loss channel with transmittance τ, maps a coherent state $|\alpha\rangle$ into another, amplitude damped, coherent state $|\sqrt{\tau}\alpha\rangle$. Once the input state ρ_{in} is fixed, the quantum channel discrimination problem becomes a binary quantum state discrimination one, between the two possible outputs, ρ_0 and ρ_1, for which the lowest error probability is given, as discussed in the previous chapter, by the celebrated Helstrom formula [37, 40]

$$p_{err}^{H} = \frac{1}{2}(1 - ||\pi_0\rho_0 - \pi_1\rho_1||) \tag{3.4}$$

where $|| \cdot ||$ is the trace distance and π_0, π_1 are the prior probabilities that in the case we are considering are fixed to $\pi_i = \frac{1}{2}$.

For coherent input the output of the channel are pure and the trace distance can be expressed in terms of the overlap $\zeta = \langle\sqrt{\tau_0}\alpha|\sqrt{\tau_1}\alpha\rangle = e^{-\frac{\mu}{2}(\sqrt{\tau_0}-\sqrt{\tau_1})^2}$ as:

$$||\pi_0|\sqrt{\tau_0}\alpha\rangle\langle\sqrt{\tau_0}\alpha| - \pi_1|\sqrt{\tau_1}\alpha\rangle\langle\sqrt{\tau_1}\alpha||| = \sqrt{1 - 4\pi_0\pi_1|\zeta|^2}$$

and by substituting this and the condition $\pi_i = \frac{1}{2}$ in Eq. (3.4) we get:

$$p_{err}^{coh} = \frac{1 - \sqrt{1 - |\zeta|^2}}{2} \tag{3.5}$$

that coincides with $\mathscr{C}(\mu, \tau_0, \tau_1)$ proving that a single mode coherent state saturates the bound.

As discussed in Sec(2.1), the optimal probability of error is achieved by projecting into the eigenstates of the Hermitian operator Λ defined in this case as:

$$\Lambda = \frac{1}{2}(|\sqrt{\tau_0}\alpha\rangle\langle\sqrt{\tau_0}\alpha| - |\sqrt{\tau_1}\alpha\rangle\langle\sqrt{\tau_1}\alpha|) \tag{3.6}$$

We can find an orthogonal basis to represent Λ by performing the Gram-Smith orthogonalization on the pair of vectors $(|\sqrt{\tau_0}\alpha\rangle, |\sqrt{\tau_1}\alpha\rangle)$ spanning its support. So we define the basis:

$$|0\rangle = |\sqrt{\tau_0}\alpha\rangle \qquad\qquad |1\rangle = \frac{|\sqrt{\tau_1}\alpha\rangle - \zeta|\sqrt{\tau_0}\alpha\rangle}{\sqrt{1 - |\zeta|^2}} \tag{3.7}$$

The eigenvectors of Λ are:

$$|+\rangle = \frac{1}{\sqrt{2}}\sqrt{1 + \sqrt{1 - \zeta^2}}|0\rangle + \frac{1}{\sqrt{2}}\sqrt{1 - \sqrt{1 - \zeta^2}}|1\rangle \qquad (3.8)$$

$$|-\rangle = \frac{1}{\sqrt{2}}\sqrt{1 - \sqrt{1 - \zeta^2}}|0\rangle - \frac{1}{\sqrt{2}}\sqrt{1 + \sqrt{1 - \zeta^2}}|1\rangle \qquad (3.9)$$

$$(3.10)$$

And substituting the definitions in Eq. (3.7) in Eq. (3.9) leads to:

$$|+\rangle = \sqrt{\frac{1 - \sqrt{1 - \zeta^2}}{2(1 - \zeta^2)}}|\sqrt{\tau_1}\alpha\rangle - \sqrt{\frac{1 + \sqrt{1 - \zeta^2}}{2(1 - \zeta^2)}}|\sqrt{\tau_0}\alpha\rangle \qquad (3.11)$$

$$|-\rangle = \sqrt{\frac{1 + \sqrt{1 - \zeta^2}}{2(1 - \zeta^2)}}|\sqrt{\tau_1}\alpha\rangle - \sqrt{\frac{1 - \sqrt{1 - \zeta^2}}{2(1 - \zeta^2)}}|\sqrt{\tau_0}\alpha\rangle \qquad (3.12)$$

That is in the form of a *pseudo-cat* state. The bit is then recovered, according to the prescription of the previous chapter, by measuring over the projectors $\Pi_0 = |+\rangle\langle+|$ and $\Pi_1 = |-\rangle\langle-| = \mathbb{1}_2 - |+\rangle\langle+|$, $\mathbb{1}_2$ being the identity in the subspace spanned by $\{|0\rangle, |1\rangle\}$. Once the measurement is performed the value τ_0 is selected if the outcome is $+$ and τ_1 is selected otherwise. The result in Eqs.(3.11-3.12) is an original contribution of this Thesis.

Consider now a multi-mode quantum input state in a tensor product of M replicas of a two mode squeezed vacuum (TMSV) states, $|\mathrm{TMSV}\rangle_{S,I}^{\otimes M}$. Remember that a TMSV state in the, Fock basis, takes the form $|\mathrm{TMSV}\rangle_{S,I} = (\cosh r)^{-1}\sum_n(\tanh r)^n|n\rangle_S|n\rangle_I$. Each TMSV state irradiates $\sinh|r|^2 = \mu/M$ mean photons per mode and describes an entangled pair of signal (S) and idler (I) modes, so that we have a total of M signals and corresponding $L = M$ idlers. Globally, μ mean photons are irradiated over the cell. For a sufficiently large number of copies M, it is possible to show that the error probability, p_{err}^{TMSV}, goes below the classical bound $\mathscr{C}$. This can be done using another tool in quantum state discrimination discussed in the previous chapter, the Quantum Chernoff bound. For equal priors we have according to Eq. (2.25) the

following upper bound:

$$p_{err}^{TMSV} \leq \mathscr{Q}(\mu, M, \tau_0, \tau_1) := \frac{1}{2} e^{-M\xi_{QCB}} \tag{3.13}$$

$$\xi_{QCB} = - \min_{\alpha \in [0,1]} \log \left(\text{Tr}[\sigma_0^\alpha \sigma_1^{1-\alpha}] \right)$$

where $\sigma_i = \mathscr{E}_{\tau_i}(|\text{TMSV}\rangle_{S,I}\langle\text{TMSV}|)$. In the asymptotic limit of $M \to \infty$ the upper bound in Eq. (3.13) becomes tight as discussed in the previous chapter. For any signal energy μ such that:

$$\mu > \mu_{th} := \frac{2\ln 2}{2 - \tau_0 - \tau_1 - 2\sqrt{(1-\tau_0)(1-\tau_1)}} \tag{3.14}$$

there is a number of modes $\bar{M}$ such that $\mathscr{Q}(\mu, \bar{M}, \tau_0, \tau_1) \leq \mathscr{C}(\mu, \tau_0, \tau_1)$. Thus, a TMSV transmitter is able to surpass any classical strategy in this regime. In the case of an *ideal memory*, namely a memory having $\tau_0 \leq \tau_1 = 1$, the energy threshold simplifies to $\mu_{th} = 1/2$ and the quantum strategy outperforms the classical one for every $M \geq \bar{M}$.

The quantum performance $\mathscr{Q}(\mu, M, \tau_0, \tau_1)$ can be evaluated more in detail numerically and compared with the classical one. To this end, it is useful to define the informational gain G:

$$G = 1 - H(\mathscr{Q}) - [1 - H(\mathscr{C})], \tag{3.15}$$

that is the difference between the bits recovered by the quantum strategy and the classical one. Note that G defines a lower bound on the actual quantum gain, since $\mathscr{C}$ is a lower bound on the optimal classical probability of error, while $\mathscr{Q}$ is an upper bound on the quantum probability of error with TMSV states.

Ref. [48] with a numerical study analyses the gain as a function of the parameters, in the low photons regime. A quantum advantage is shown in a wide region and it is bigger for higher values of transmittances τ_0 and τ_1. In case of an ideal memory, i.e. a memory having $\tau_1 = 1$, an advantage is already present for a single mode TMSV transmitter, $M = 1$, a case in which there is no guarantee that the QCB is tight. The quantum gain found is nonetheless significant in a wide region. When the number of modes approaches infinity, $M \to \infty$, the magnitude of the gain is sensibly larger and approaches its maximum value, $G \to 1$ in some regions. A gain G approaching 1 means that the quantum strategy retrieves all the information while no information

whatsoever can be recovered by any classical one. The increase in the advantage between the cases $M = 1$ and $M \to \infty$ is due to two different effects. The first one is related to the theoretical approach chosen. The Quantum Chernoff bound is in fact an asymptotic tool and it becomes tighter as the limit $M \to \infty$ is approached. The other one is instead related to the nature of the problem. Spreading the signal energy into a larger number of modes leads, in fact, to an improvement in the overall performance of TMSV transmitters. This effect will find a clear physical interpretation in the following section where explicit *photon counting* measurements will be considered. For this kind of measurements, having a large number of modes means that the measured photon statistics is more peaked leading to better distinguishability of different states.

While the theoretical proposal for the protocol of quantum reading was presented by Ref. [48] in 2011 and the idea has been further explored in a series of papers (e.g., see Refs. [62–65]) an experimental realization has been lacking until recently. A preliminary experiment was performed in Ref. [66] for a perfect fully-unitary variant of the protocol, where zero discrimination error was achieved by using both the outputs of the beam-splitter, modelling the cell. For such an ideal unitary model no entanglement is needed. However, in a realistic scenario, only one output of the cell is available for detection, so the process is clearly non-unitary and must be described, as in the original proposal, by a lossy quantum channel. The *first experimental realization* of the quantum reading protocol as originally proposed in Ref. [48], is in fact an original contribution of this Thesis work and has been reported in Ref. [1] in 2021. In the following we will present the details of this realization.

3.2   Experimental realization of quantum reading with photon counting

The theoretical analysis of the previous section showed that it is possible to obtain a quantum advantage in the protocol of quantum reading. An important question was however left unanswered, i.e. how can this advantage be practically achieved. The answer to this question, and its experimental implementation, are the central results of Ref. [1] where we showed how in this protocol a quantum advantage is obtained by photon-counting measurements combined with a maximum-likelihood
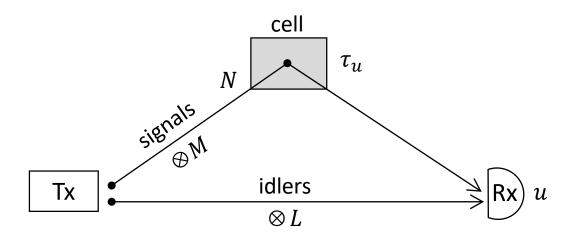
Fig. 3.1 *Scheme of the quantum reading of a memory cell.* Image taken from Ref. [1]. A memory cell encodes a bit u in a pure loss channel with transmissivity τ_u. The cell is read by a transmitter (Tx) irradiating M signal modes and μ mean total photons over the cell, plus L idler modes directly to the output. A generally joint measurement is performed at the receiver (Rx) to decode the bit u up to some error probability p_{err}.

decision,where the input state consists of a collection of TMSV states. The advantage was proven notwithstanding the presence of more than 20% experimental loss, showing a good robustness for the technique proposed, in view of possible real applications.

Consider the single-cell quantum reading scheme described in the previous section. A memory cell stores a classical bit $x = \{0, 1\}$ by means of two equiprobable pure loss channels, $\mathscr{E}_{\tau_0}$ and $\mathscr{E}_{\tau_1}$, with transmittances τ_0 and τ_1. The bit is recovered using a transmitter irradiating M signal (S) modes over the cell, for a total of μ mean photons, and also L additional idler (I) modes directly to the receiver, where a measurement is performed. The outcome of the measurement leads to a guess y of the classical bit x with an error probability p_{err} (see Fig. 3.1). The minimum probability of error achievable by classical states, p_{err}^{cla}, has been given in Eq. (3.3). This probability of error is saturated by highly theoretical measurements, such as the projection on the state in Eq. (3.12), that are hardly feasible in an experimental setting, and to our knowledge it has not even been devised. A much more practical design to retrieve the bit is provided by the *photon-counting (PC)* receiver [1].

3.2.1 The Photon-Counting Receiver

For photon counting measurement performed over the signal and idler modes of a bipartite state ρ, the output is a classical random variable $\mathbf{n} = (n_S, n_I)$, distributed as $p(\mathbf{n}) = \mathrm{Tr}(\rho |n_S, n_I\rangle\langle n_S, n_I|)$, where $|n_k\rangle$ is the eigenstate with eigenvalue n_k of the number operator $\hat{n}_k = \hat{a}_k^\dagger \hat{a}_k$ of the field, $k = S, I$ denotes signal and idler system respectively, and we used the notation $|n_S, n_I\rangle := |n_S\rangle \otimes |n_I\rangle$. The effect of a pure loss channel $\mathscr{E}_\tau$ on the signal mode of a bipartite state is to combine its initial photon number distribution $p_0(\mathbf{n})$ with a binomial distribution $\mathscr{B}(n_S'|n_S, \tau)$ with n_S trials and success probability τ, so that the outcome $\mathbf{n}$ will be distributed according to [67, 68]:

$$p(\mathbf{n}|\tau) = \sum_{m=n_S}^{\infty} p_0(m, n_I)\mathscr{B}(n_S|m, \tau) \qquad (3.16)$$

In Appendix (A) we give a derivation of Eq. (3.16) in the univariate case.

Let us now suppose that $\mathbf{n}$ is the outcome of a photon-counting measurement after a lossy channel with unknown transmissivity τ_x (for $x = 0, 1$). Note how, once the measurement is fixed to photon counting and the initial state (or equivalently in this case its photon number distribution) is fixed as well, the originally quantum problem is reduced to a classical binary hypothesis testing. As already mentioned in the previous chapter, in such problems, the optimal decision is taken using *Bayes rule*, by construction. Using Bayes' theorem, the conditional probability of τ_u, $p(\tau_u|\mathbf{n})$, is given by:

$$p(\tau_x|\mathbf{n}) = \frac{p(\mathbf{n}|\tau_x)p(\tau_x)}{p(\mathbf{n})} = \frac{p(\mathbf{n}|\tau_x)}{p(\mathbf{n}|\tau_0) + p(\mathbf{n}|\tau_1)}, \qquad (3.17)$$

where the last equality follows from the condition of equi-probable channels, $p(\tau_x) = 1/2$. To assign a value y to the recovered bit, Bayes rule prescribes to choose the value $y = 0, 1$ such that $y = \arg\max_x p(\tau_x|\mathbf{n})$. For equal priors, this rule is equivalent to the maximum likelihood decision, $y = \arg\max_x p(\mathbf{n}|\tau_x)$. For any given outcome $\mathbf{n}$, the related error probability will be $p_{err}(\tau_0, \tau_1|\mathbf{n}) = \min_x p(\tau_x|\mathbf{n})$, and the mean error probability of the discrimination is obtained by averaging this value over the outcome distribution $p(\mathbf{n})$:

$$p_{err}(\tau_0, \tau_1) = \sum_{\mathbf{n}} \min_u p(\tau_u|\mathbf{n})p(\mathbf{n})$$
$$= \frac{1}{2}\sum_{\mathbf{n}} \min_u p(\mathbf{n}|\tau_u). \qquad (3.18)$$

The error probability above describes the performance achievable by a photon-counting receiver in the quantum reading protocol pictured in Fig. (3.1) where the transmitter irradiates a generic bipartite state. In general, the formula can be applied to a transmitter with arbitrary M and L by considering a $M + L$ vectorial variable $\mathbf{n}$. Using this analysis we can now evaluate the performance of a given input state with a PC receiver. In particular, we will analyse two cases, the first one is a generic classical transmitter, for which we are able to present an analytical bound on the performance. The second one is the case of quantum multimode TMSV state, that achieves an *absolute quantum advantage* over any classical strategy.

3.2.2 Classical States and PC receiver

To characterize the performance of classical states we use the results of the previous section, showing that the probability of error is fully determined by the photon number distribution of the input state. We remind that a generic multi-mode bipartite classical state is defined as convex superpositions of coherent states, according to Eq. (3.2). In the following, without loss of generality, we will restrict our analysis to single mode classical states. Classical correlation cannot, in fact, improve the discrimination accuracy [48] and spreading the signal energy in more than one mode is not helpful either, as shown in Appendix B. A single mode classical state, ρ_{cla} can be written as:

$$\rho_{cla} = \int d^2\alpha \, p(\alpha) |\alpha\rangle\langle\alpha| \tag{3.19}$$

where $p(\alpha) \geq 0$. In the quantum reading protocol the mean number of signal photons is fixed to a given value μ. For the state ρ_{cla} this condition can be written as:

$$\int d^2\alpha \, p(\alpha) |\alpha|^2 = \mu \tag{3.20}$$

A coherent state $|\alpha\rangle$, with mean photon number $\mu = |\alpha|^2$, has a Poisson photon number distribution, $\mathbf{P}_\mu(n)$, as can be easily seen by direct computation:

$$p_{coh}(n) = |\langle n|\alpha\rangle|^2 = \left| e^{-\frac{|\alpha|^2}{2}} \sum_{m=0}^{\infty} \frac{\alpha^m}{\sqrt{m!}} \langle n|m\rangle \right|^2 =$$

$$= e^{-\mu} \frac{\mu^n}{n!} = \mathbf{P}_\mu(n) \tag{3.21}$$

The photon number variance of a coherent state is then given by its mean number of photons μ. A direct consequence of the definition of classical states on their photon number distribution is that, once the mean number of photons is fixed to μ, their variance is lower bounded by the one of a coherent state. Consider the distribution, $p_{cla}(n)$ of a generic classical state in the form of Eq. (3.19):

$$
\begin{aligned}
p_{cla}(n) = \langle n|\rho_{cla}|n\rangle &= \int d^2\alpha\, p(\alpha)\, |\langle n|\alpha\rangle|^2 \\
&= \int d^2\alpha\, p(\alpha)\, p(n|\alpha)
\end{aligned}
\tag{3.22}
$$

$$
p(n|\alpha) := \mathbf{P}_{|\alpha|^2}(n)
$$

where we highlighted that α parameterizes the distribution on n. The variance $\langle\Delta^2 n\rangle$ of the *compound* distribution can be computed using the following equality:

$$
\langle\Delta^2 n\rangle = \Big\langle \big\langle \Delta^2(n|\alpha)\big\rangle \Big\rangle_\alpha + \Big\langle \Delta^2\langle(n|\alpha)\rangle \Big\rangle_\alpha
\tag{3.23}
$$

Where $\Big\langle \big\langle \Delta^2(n|\alpha)\big\rangle \Big\rangle_\alpha$ is the expectation value over α of the conditioned variance $\langle\Delta^2(n|\alpha)\rangle$ and $\Big\langle \Delta^2\langle(n|\alpha)\rangle \Big\rangle_\alpha$ is the variance over α of the expected value of n conditioned to α, $\langle(n|\alpha)\rangle$. Eq. (3.23) holds for every distribution parameterized in the form of Eq. (3.22). We have then:

$$
\begin{aligned}
\langle\Delta^2 n\rangle = \int d^2\alpha\, p(\alpha)\, |\alpha|^2 + \Big\langle \Delta^2\langle(n|\alpha)\rangle \Big\rangle_\alpha &= \\
= \mu + \Big\langle \Delta^2\langle(n|\alpha)\rangle \Big\rangle_\alpha \geq \mu &= \langle\Delta^2 n\rangle_{coh}
\end{aligned}
\tag{3.24}
$$

that shows how at fixed energy the variance of classical states is minimized by a coherent state. The last inequality follows from the fact that $\Big\langle \Delta^2\langle(n|\alpha)\rangle \Big\rangle_\alpha$ is a variance and hence non-negative.

This result can be used to find the optimal performance in the discrimination. According to Eq. (A.5) after the interaction with the pure loss channel the original photon number distribution is compounded with a binomial one so, using Eq. (3.23), we have for the variance of the output distribution:

$$
\langle\Delta^2 n\rangle_{out} = \tau^2\langle\Delta^2 n\rangle_{in} + \mu\tau(1-\tau)
\tag{3.25}
$$

showing that the input distribution with minimum variance minimizes the variance of output distribution as well. In the discrimination with PC measurements the probability of error is proportional to the overlap of the two photon number output distributions, for τ_0 and τ_1 respectivelly. So the minimum probability of error is achieved by the input states minimizing the variance, i.e. states having Poisson distribution.

By the argument made above the *minimum probability of error* in the quantum reading protocol *with a PC receiver and classical states* can be computed considering an input Poisson photon number distribution $\mathbf{P}_\mu(n)$. This analysis is in agreement with the fact that the Poisson distribution sets the Shot Noise Limit (SNL) [69, 18] an important limit for classical states in estimation problems. A useful property of $\mathbf{P}_\mu(n)$ is that the interaction with a pure loss channel $\mathscr{E}_\tau$ does not change the form of the distribution, in fact:

$$\mathbf{P}_\mu(n) \xrightarrow{\mathscr{E}_\tau} p_{out}(n) = \sum_{m=0}^{\infty} \mathbf{P}_\mu(m)\mathscr{B}(n|m,\tau) =$$
$$= \sum_{m=n}^{\infty} \frac{e^{-\mu}\mu^m}{m!} \binom{m}{n} \tau^n(1-\tau)^{m-n} =$$
$$= \frac{e^{-\mu}(\tau\mu)^n}{n!} \sum_{\alpha=0}^{\infty} \frac{(\mu(1-\tau))^\alpha}{\alpha!} =$$
$$= \frac{e^{-\mu\tau}(\tau\mu)^n}{n!} = \mathbf{P}_{\tau\mu}(n) \tag{3.26}$$

A Poisson distribution describes independent events (photon detections in this case) so, the fact that spreading the energy over M independent modes doesn't change the result has a rather direct physical interpretation in terms of this independence.

Using the invariance property in Eq. (3.26) the possible output distributions after the interaction with the memory cell are $\mathbf{P}_{\tau_0\mu}(n)$ and $\mathbf{P}_{\tau_1\mu}(n)$. To evaluate the probability of error in Eq. (3.18) we can identify a threshold value n_{th} :

$$n_{th} := \frac{\mu(\tau_1 - \tau_0)}{\log(\tau_1/\tau_0)} \tag{3.27}$$
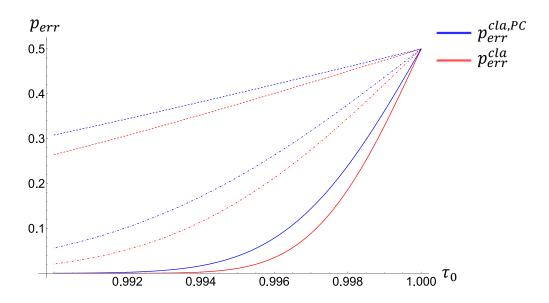
Fig. 3.2 *Comparison between Photon counting and best classical strategies.* The classical probability of error with photon counting $p_{err}^{cla,PC}$ and the absolute classical best performance p_{err}^{cla} are reported as a function of one of the transmittance values τ_0, while $\tau_1 = 1$. The comparison is reported for different mean number of photons, $\mu = 10^4$ (dashed lines), $\mu = 10^5$ (dot-dashed lines) and $\mu = 5 \times 10^5$ (solid lines)

such that for any outcome $n \leq n_{th}$ we have $\mathbf{P}_{\tau_0 \mu}(n) \geq \mathbf{P}_{\tau_1 \mu}(n)$ and vice-versa. For the probability of error we have then:

$$p_{err}^{cla,PC}(\tau_0, \tau_1) = \frac{1}{2} \sum_{n=0}^{n_{th}} \mathbf{P}_{\tau_1 \mu}(n) + \frac{1}{2} \sum_{n=n_{th}}^{\infty} \mathbf{P}_{\tau_0 \mu}(n)$$
$$= \frac{1}{2} \left[1 - \frac{\gamma(\tau_0) - \gamma(\tau_1)}{\lfloor n^{th} \rfloor!} \right] \qquad (3.28)$$
$$\gamma(\tau_1) := \Gamma(\lfloor n_{th} + 1 \rfloor, \tau_0 \mu)$$

Where $\lfloor x \rfloor$ denotes the floor and $\Gamma(x,y)$ is the incomplete gamma function. Eq. (3.28) establishes a lower bound on the error probability that can be achieved by using classical transmitters paired with a PC receiver. Fig. (3.2) shows the comparison between $p_{err}^{cla,PC}$ and the absolute bound p_{err}^{cla} defined in Eq. (3.4) of previous section. As it can be seen the performance offered by the PC receiver is not far from optimality.

3.2.3   TMSV states and PC receiver

Let us now study the photon-counting performance that is achievable by a quantum transmitter based on TMSV states. We consider the transmitter's state $|\text{TMSV}\rangle_{S,I}^{\otimes^M}$, where each signal-idler TMSV state is:

$$|\text{TMSV}\rangle_{S,I} = \sum_n c_n|n\rangle_S|n\rangle_I \tag{3.29}$$

$$|c_n|^2 = \bar{n}^n/(\bar{n}+1)^{n+1}$$

where signal and idler are maximally correlated in the number of photons and locally characterized by a single-mode thermal distribution [70]. In the following we will consider the product state $|\text{TMSV}\rangle_{S,I}^{\otimes^M}$, that still shows a perfect correlation between the total number of photon in the signal and idler system defined as $n_{S/I} = \sum_{m=1}^M n_{S/I}^{(m)}$.

The choice of a collection of M modes is both convenient from a theoretical standpoint, since it gives a better performance, and more natural from an experimental point of view, since typically the integration time of detectors are much larger than the coherence time of the field and, as a result, a large number of temporal modes are collected within each measurement. A more detailed discussion on this can be found in the following sections. The effect of accumulating M modes is that the marginal distributions, of both signal and idler mode separately, are multi-thermal $P_{\mu,M}(n_{S/I})$ [71]:

$$P_{\mu,M}(n) = \binom{\mu+M-1}{\mu} \frac{\langle n\rangle^\mu}{(1+\langle n\rangle)^{\mu+M}} \tag{3.30}$$

where μ is the total mean number of photons and $\langle n\rangle = \mu/M$ is the mean occupation of each separate mode. The variance of this distribution is given by $\langle \Delta^2 n\rangle = \mu(1+\mu/M)$. Keeping the total photon fixed to μ and increasing the number of modes M, i.e. lowering the single mode occupation number $\langle n\rangle$, narrows the distribution and in the limit of $\mu/M \to 0$ the variance goes to $\langle \mu \rangle$ and the distribution tends to the Poisson distribution $\mathbf{P}_\mu(N)$,i.e. for finite values of μ:

$$\lim_{\mu/M\to 0} P_{N,M}(n) = \mathbf{P}_\mu(n) \tag{3.31}$$

In our experimental regime the number of modes is usually very large ($M \sim 10^{13}$, see section on experimental setup), so that the error in approximating the muti-thermal distribution with a Poisson one is negligible. Since the single modes are correlated

by δ functions, their sums must be correlated as well so that the joint distribution of the bivariate variable $\mathbf{n} = \{n_S, n_I\}$ will be:

$$p(\mathbf{n}) = \mathbf{P}_\mu(n)\delta(n - n_S)\delta(n - n_I) \tag{3.32}$$

The effect of an attenuation channel, according to Eq. (3.16), leads to:

$$P(\mathbf{n}|\tau) = \mathbf{P}_\mu(n_I)\mathscr{B}(n_S|n_I, \tau) \tag{3.33}$$

Similarly to the classical case a decision can be taken by identifying a threshold value n_S^{th}:

$$n_S^{th} := \left\{ \frac{\log(\tau_1/\tau_0)}{\log[(1-\tau_0)/(1-\tau_1)]} + 1 \right\}^{-1} n_I \tag{3.34}$$

such that if $n_S < n_S^{th}$, we have $P_{\mu,M}(n_S, n_I|\tau_0) > P_{\mu,M}(n_S, n_I|\tau_1)$ and the value τ_0 is chosen. Otherwise we choose τ_1. Note how the threshold value in Eq. (3.34) does not depend on the form of the initial distribution so it is valid both in the exact muli-thermal case and in the Poisson approximation. We can write the probability of error of the quantum strategy, $p_{err}^{qua,phc}$, as:

$$p_{err}^{qua,phc}(\tau_0, \tau_1) = \frac{1}{2}\sum_{n_I=0}^{\infty}\sum_{n_S=0}^{n_S^{th}} P(\mathbf{n}|\tau_1) + \frac{1}{2}\sum_{n_I=0}^{\infty}\sum_{n_S=n_S^{th}}^{\infty} P(\mathbf{n}|\tau_0) \tag{3.35}$$

To evaluate Eq. (3.35) more easily it is useful to approximate the output distributions in Eq. (3.33) with Gaussian ones, an approximation valid in the limit of $M >> 0$ and when the number of photons is not too small.

Before discussing this approximation more in detail, it is convenient to analyse an important issue in any possible realization of the Quantum reading protocol, the *effect of optical losses* [72–74].

A quantity characterizing every quantum optical experimental setup is the quantum efficiency $0 \leq \eta \leq 1$. This quantity expresses the fraction of photons entering the experimental setup that is actually detected. The total efficiency takes into account photon losses from interaction with the environment and optical components as well as the intrinsic quantum efficiency of the detector used.

It is well known that quantum properties require a good quantum efficiency η to be revealed and the derived advantage over the classical limits is usually very

sensitive the value of η. An analysis of its effect is then necessary to obtain accurate predictions of the experimental outcome. The imperfect experimental setup acts as an amplitude attenuation channel, with transmission η, on the input quantum state. The effect of imperfect quantum efficiency is therefore indistinguishable from the effect that an attenuator, such as the memory storing the value of a bit in its attenuation coefficient as described before. From the standpoint of the photon number distribution, this is stated by the composition property of the binomial distribution characterizing the process:

$$
\begin{aligned}
\sum_{m=n}^{N} \mathscr{B}(m|\mu,\tau)\mathscr{B}(n|m,\eta) &= \sum_{m=n}^{\mu} \binom{\mu}{m}\binom{m}{n}\tau^m(1-\tau)^{\mu-m}\eta^n(1-\eta)^{m-n} = \\
&= \binom{\mu}{n}\sum_{m=n}^{\mu}\binom{\mu-n}{m-n}\tau^m(1-\tau)^{\mu-m}\eta^n(1-\eta)^{m-n} = \\
&= \binom{\mu}{n}(\tau\eta)^n(1-\tau)^{\mu-n}\sum_{\alpha=0}^{\mu-n}\binom{\mu-n}{\alpha}\left(\frac{\tau(1-\eta)}{1-\tau}\right)^{\alpha} = \\
&= \binom{\mu}{n}(\tau\eta)^n(1-\tau)^{\mu-n}\left(1+\frac{\tau(1-\eta)}{1-\tau}\right)^{\mu-n} = \\
&= \binom{\mu}{n}(\tau\eta)^n(1-\tau\eta)^{\mu-n} = \\
&= \sum_{m=n}^{\mu} \mathscr{B}(m|\mu,\eta)\mathscr{B}(n|m,\tau) = \mathscr{B}(n|\mu,\tau\eta) \qquad (3.36)
\end{aligned}
$$

The classical limits found above, both the absolute one in Eq. (3.3) and the PC one in Eq. (3.28), can be computed in case of a *signal system* with an efficiency $\eta_s < 1$ simply performing the substitution:

$$
\tau_\alpha \to \eta_s \tau_\alpha \qquad (3.37)
$$

This is, in fact, a consequence of the indistinguishability of the two attenuation processes, η_S and τ_α, that the state undergoes, and shows how the discrimination of two quantum channels τ_0 and τ_1 with a signal quantum efficiency η_S is equivalent to the discrimination between two channels $\eta_s\tau_0$ and $\eta_s\tau_1$ with perfect quantum efficiency. This argument does not depend on the measurement performed, hence, as pointed out, it is valid also in the case of the absolute bound. Equivalently, the same resut can be obtained by considering the discrimination between the original transmittances τ_0 and τ_1 with attenuated energy $\mu\eta_S$. In other words the effect of

optical losses on the classical performance is equivalent to reduce the energy of the probe state, thus decreasing the accuracy. When quantum-correlated systems are considered, however, aside from the energy reduction, an additional effect induced by losses is the worsening of the correlations, therefore decreasing the advantage that can be obtained. This reduction in correlation is shown in Eq. (3.39-3.40).

Note how in this section we are not discussing the fact that, in case of bipartite correlations, the efficiency in detecting correlated photons, known as *heralding efficiency*, can be lower than the efficiency in detecting the photons in a single arm. This issue will be considered in detail in Chapter (5), when discussing the patter recognition experiment. In the realization of the quantum reading protocol no spatial resolution is needed and this gives the possibility to detect photons in an area much larger than the coherence area of the spatial modes. As a result, experimental misalignments and mode matching are reduced to a minimum, meaning that the heralding efficiency practically coincides with the one of the single arm. This effect will be relevant in the following section regarding Pattern recognition and will be discussed more in depth there.

When the number of collected modes is large, the distribution after the channel can be approximated to a bivariate Gaussian distribution. We have, in fact, a large number of modes whose photon counts are independent identically distributed random variables. Their sum is a random variable with mean value $\bar{\mathbf{n}}$ and covariance matrix Σ. According to the *multivariate central limit theorem* its distribution will converge, in the limit of $M \to \infty$, to a multivariate Gaussian distribution with the same mean vector and covariance matrix:

$$P(\mathbf{n}|\tau) \xrightarrow{M \to \infty} \mathcal{N}(\bar{\mathbf{n}}, \Sigma) \tag{3.38}$$

After the interaction a memory cell with parameter τ and considering also the detection efficiencies, η_S and η_I, on the signal and idler systems respectively, the mean value before detection is $\bar{\mathbf{n}} = \{\tau\eta_S\mu, \eta_I\mu\}$. Since both the losses, η_S and η_I, and the memory cell τ act as pure loss channels the output covariance matrix Σ can then be evaluated using Eq. (3.25) and Eq. (3.16), yielding:

$$\Sigma(\tau) = \begin{pmatrix} \tau^2\eta_S^2\langle\Delta^2 n_S\rangle_{in} + \tau\eta_S\mu(1-\tau\eta_S) & \tau\eta_S\eta_I\langle\Delta n_S\Delta n_I\rangle_{in} \\ \tau\eta_S\eta_I\langle\Delta n_S\Delta n_I\rangle_{in} & \eta_I^2\langle\Delta^2 n_I\rangle_{in} + \eta_I\mu(1-\eta_I) \end{pmatrix} \tag{3.39}$$

where the subscript "in" refers to the quantities at the transmitter (input). The off-diagonal term can be computed from the joint distribution after considering independent losses on the two beams or by statistical considerations [75].

In our case, we have, in view of the initial perfect correlation in photon number between signal and idler, $\langle \Delta^2 n_S \rangle_{in} = \langle \Delta^2 n_I \rangle_{in} = \langle \Delta n_S \Delta n_I \rangle_{in}$. In the Poisson approximation we have $\langle \Delta^2 n_S \rangle_{in} = \langle \Delta^2 n_I \rangle_{in} = \mu$ and for the covariance matrix becomes:

$$\Sigma(\tau) = \begin{pmatrix} \tau \eta_S \mu & \tau \eta_S \eta_I \mu \\ \tau \eta_S \eta_I \mu & \eta_I \mu \end{pmatrix} \tag{3.40}$$

In the Gaussian approximation of Eq. (3.38) the probability of error of the quantum strategy, in Eq. (3.35), becomes:
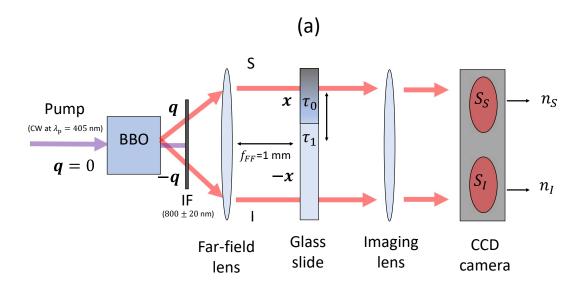
$$p_{err}^{qua,phc} \approx \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dn_I dn_S \min_x \left(\mathcal{N}(\mathbf{n}, \Sigma(\tau_x)) \right) \tag{3.41}$$

Clearly, since the sums over the discrete variables n_S and n_I are being substituted by integrals this approximation is meaningful only if the number of photon detected is $\gg 0$. Note how, while the integrals formally run from $-\infty$, that may seem problematic since clearly photon counts cannot be negative, in the regime of validity of the approximation the probability of having values close to 0 is practically null. In the experimental ranges analysed in the following the mean number of photons is of the order of magnitude of $\mu \sim 10^5$ and the trasmissivity are high so we are always well within the range of validity of this approximation. Eq. (3.35) greatly simplifies the numerical evaluation of the probability of error and will be used in the following analysis.

In Appendix B we show how for independent modes individual photon number measurements on each mode yield no advantage over collective measurement.

3.2.4 Experimental setup

Fig. 3.3(a) reports a scheme of the experimental set-up. The multi-mode state $|\text{TMSV}\rangle_{S,I}^{\otimes M}$ is realized experimentally using the process of spontaneous parametric down conversion (SPDC)[76–78, 67] in a non linear crystal. A $(1\text{cm})^3$ type-II-Beta-Barium-Borate (BBO) non-linear crystal is "pumped" by a CW laser of wavelength $\lambda_p = 405\text{nm}$ and power of 100mW. The down-converted photons are spectrally
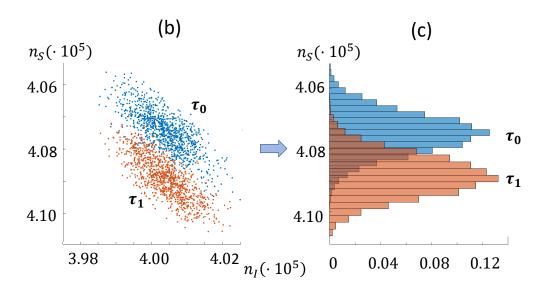
Fig. 3.3 *Experimental setup for quantum reading.* (a) Schematic of the experimental set-up. The multi-mode TMSV source is generated in a BBO crystal. The signal beam passes through the memory cell investigated, whose transmittance can be either τ_0 or τ_1 and is then detected in the S_S region of the CCD camera. The idler beam goes directly to the S_I region of the CCD. n_S and n_I are the total photon counts over S_S and S_I. BBO: Type-II-Beta-Barium-Borate non linear crystal. IF: interferential filter (800 ± 20nm). CCD: charge-coupled device camera. (b) Signal photon counts as a function of the idler ones, 1000 acquisitions. The results regarding $\tau_0 \sim 0.996$, are reported in blue dots, while red dots corresponds to $\tau_1 = 1$. (c) Relative frequency distribution for the signal photon counts n_S, reported in blue for $\tau_0 \sim 0.996$ and in red for $\tau_1 = 1$. The image is taken from Ref. [1]

selected, by an interferential filter (IF) at (800 ± 20)nm, around the degenerate frequency ($\lambda_d = 2\lambda_p = 810$nm). The correlation of the photons is in momentum and it is converted in the spatial correlation in the sample plan with a "far-field" lens in an $f - f$ configuration, meaning that both the crystal and the sample plane, where the correlation are mapped, are positioned at one focal length from the lens (see scheme in Fig. (3.3(a)). The focal length of the lens is $f_{FF} = 1$cm. The sample plane is then imaged to the detection plane by a second, imaging, lens.

The photons are detected by a charge-coupled-device (CCD) camera (Princeton Instrument Pixis 400BR Excelon), working in linear mode, with high quantum efficiency (nominally $> 95\%$ at 810nm) and few $e^-/(\text{Pixel} \cdot \text{Frame})$ of electronic noise. The physical pixels of the camera measure 13μm. In order to lower the acquisition time and increase the read-out signal-to-noise ratio, a 12×12 hardware binning is performed on them, to constitute a detection pixel. Integrating the signal over two spatially correlated detection areas S_S and S_I, for signal and idler respectively, yields the total photon counts n_S and n_I. The temporal modes can be estimated to be $M_t \sim 10^{10}$ and the total number of spatial modes is $M_s \sim 10^3$ (for a more in depth discussion on these estimates see Ref. [79]). Since $N_I \sim 10^5$, the mean occupation number is $N_I/(M_s \cdot M_t) \sim 10^{-8} \ll 1$, meaning that, according to the discussion of the previous section, the marginal distributions are well approximated by Poissonian ones.

The memory cell is implemented inserting in the sample plan (the focal plan of the "far-field" lens) of the first lens a coated glass-slide with a deposition of variable transmission $0.990 < \tau_0 < 1$. One bit of information is stored in the presence ($\tau = \tau_0$) or absence ($\tau = \tau_1 = 1$) of the deposition. To match the optical path of the signal beam, the idler beam passes through the glass as well, but without intercepting the deposition. Fig. 3.3(b) and Fig. 3.3(c) give a visualization of the effect at the base of the quantum enhancement. The joint distributions of n_S and n_I for τ_0 and $\tau_1 = 1$ overlap less, due to their squeezed shape, with respect to the marginal distributions of n_S only, so that the distinguishability is increased. The squeezed shape of the joint distributions, below the Shot Noise limit, of Fig. 3.3(b) is purely due to quantum correlations and cannot be achieved by any classical source. Classical correlated states could in fact result in a distribution with a "sigar" shape as well, but at the cost of an increased marginal variance, resulting in a joint distribution having width along any direction limited by the shot noise set by the fixed energy.

The parameters necessary for the subsequent analysis (N, τ_0, η_1, η_2, electronic noise ν_e) are estimated in a calibration phase. The channels efficiencies are estimated using the absolute calibration method presented in Refs. [80]. To evaluate the error probability in the discrimination between τ_0 and τ_1 two sets of 10000 frames are acquired, one for each value of the transmittance. For each frame, using the values of the parameters estimated in the calibration, $\mathbf{P}_{N,M}(n_S, n_I|\tau_u)$ is computed, from which a value to the bit is assigned using the maximum likelihood rule discussed above. Comparing the value recovered with the true value of the bit, a frequency of error can be estimated that will converge to the probability of error for a large number of frames acquired.

3.2.5   Results

In the following we will report the results of Ref. [1], original contribution of this dissertation, in which the first realization of the quantum reading protocol was given using the PC receiver described above. The quantum advantage is shown by a numerical investigation of the informational quantum gain $G = 1 - H(p_{err}^{qua,phc}) - [1 - H(p_{err}^{cla})]$ where p_{err}^{cla} can be either the theoretical absolute bound of Eq. (3.3) or the PC bound in Eq. (3.28). It turns out that a quantum advantage can be obtained even considering a single TMSV state, i.e. $M = 1$. However, following the discussion of previous section, the spreading of the energy over a large number of modes, $M \to \infty$, gives narrower marginal distributions, that in turn results in an improved discrimination. Thus in the following we will evaluate numerically $p_{err}^{qua,phc}$ in the Gaussian approximation discussed in the previous section, using Eq. (3.41). Fig. 3.4 reports contour plots of the informational gain as a function of the mean number of signal photons μ and one of the values of transmissivity, τ_0, while the other is fixed to $\tau_1 = 1$. In Fig. 3.4(A,B) the prediction are reported for ideal efficiencies $\eta_S = \eta_I = 1$. Note how, in both cases, theory predicts an evident information gain, which may approach the maximum value of 1, meaning that, in certain regions the use of quantum resources allows the full recovery of the stored information, whereas no information could be retrieved by classical means.

Panel **A** shows quantum gain over the classical strategy with PC, while panel **B** shows the advantage over the absolute classical limit. It can be seen how increasing the mean photon number, the maximum of the advantage shifts towards higher values of transmissivity τ_0. Intuitively, this is due to the fact that the gain does not depend
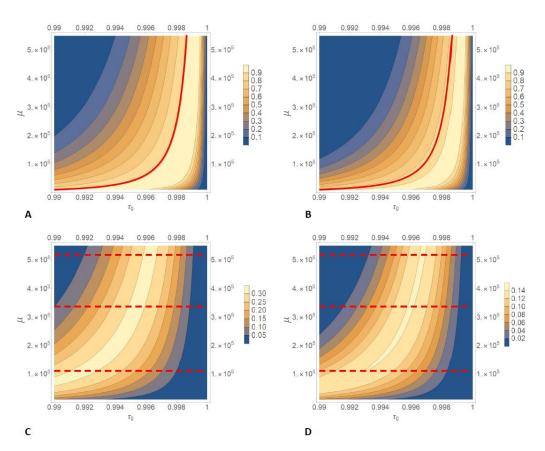
Fig. 3.4 *Theoretical information gain G for a quantum strategy with a PC receiver.* The information gain is computed assuming a TMSV-state transmitter with large number of copies ($M \simeq 10^{13}$) and a receiver based on photon counting. The plots report the gain as a function of the transmittance τ_0 and total mean number of photons μ (higher transmissivity is set to $\tau_1 = 1$).Panel **A** reports the gain over the classical benchmark with PC in Eq. (3.28). In panel **B**, the benchmark is the optimal classical limit in Eq. (3.3). The red curve represent the MED strategy described in the main text, marking the limit after which the channels are classically indistinguishable. In panels **C** and **D**, are reported the cases of imperfect quantum efficiency, $\eta = 0.76$, for both the signal and idler systems. The gain over the photon counting classical bound is shown in panel **C**, and the gain over the optimal classical limit in panel **D**. The dashed lines indicate the regions where experimental data were collected. These experimental data points are reported in Fig. (3.5). These plots are reported from Ref. [1].

only on the quantum performance, but is at its highest when classical strategies start to fail while the quantum performance is still retained. This is well illustrated if one consider the non-optimal mean-energy-discrimination (MED) strategy, i.e measuring the mean photon number that will be either μ or $\tau_0\mu$ (assuming $\tau_1 = 1$). The MED approach starts to fail when the difference in the average photon counts becomes smaller than the noise associated with the Poisson fluctuations, i.e., when $\tau_0 > 1 - \mu^{-1}$. The red line reported in in Fig.(3.4) represents the curve in which this inequality is saturated. Note how, in Fig. 3.4(A), this curve follows the contour lines of the plot, denoting with good accuracy the start of the maximum gain region. Of course, when τ_0 is approaching $\tau_1 = 1$ there is no way to distinguish among the channels, neither classical nor quantum, so the information recovered by any strategy, including the quantum one, must approach zero, and the information gain drops to zero. The competition between these two tendencies determines the maximum of the gain. When comparing with the optimal classical bound in Fig. 3.4(B), the gain regions are in general narrower, and the start of the maximum deviates from the MED curve. However, note that the optimal classical bound in Eq. (3.3) is achieved with a theoretical measurement, not feasible in practice, while the quantum strategy relies on simple photon counting measurement making the high advantage found even more remarkable.

As discussed in the previous section, while the effect of losses on classical states is equivalent to the attenuation of the signal energy, the quantum strategy suffers from an additional worsening of the correlation, resulting in a lower gain. This drop can be seen from Figs. 3.4(C-D), where the scenario with an efficiency $\eta_S = \eta_I = 0.76$ is reported. The maximum gain is reduced to $\simeq 1/3$ or $\simeq 1/6$, depending on the classical benchmark considered. Still, this is a macroscopic amount of information due to the fact that it refers to gain *per cell* and importantly shows how a sizeable quantum advantage can be achieved in a realistic scenario.

Another interesting feature, already present in the ideal case of $\eta = 1$, but even more visible in the plots for the more realistic situation (with $\eta = 0.76$, in Fig. 3.4.D) is the increase of the maximum gain with the energy μ. The maximum value approaches the unity for $\mu \to \infty$ in the the ideal case, but it is bounded by a value ≤ 1, function of η, in the non ideal one. At first this can be surprising, as it is common for quantum advantages to become more evident in a low energy regime. However it must be noticed, that this is not a direct consequence of the increase of energy, but it follows from the trend of the maximum to move towards higher values

of τ_0 as μ increases. As discussed, higher value of τ preserve more of the initial correlation of the initial state and the higher gain reflects that. Moreover the increase comes alongside a narrowing of the region where there is a significant advantage, so that low photon numbers could still be preferable in some cases.

Fig. 3.5 reports the experimental gain G evaluated from the experimental probability of error, p_{err}^{exp}, both with respect to the optimal classical bound, reported in blue, and to the classical PC bound, reported in red. The three panels are obtained for a different number of photons in the signal beam, i.e., $\mu \sim 1.15 \cdot 10^5$, $3.1 \cdot 10^5$ and $5.2 \cdot 10^5$ respectively, corresponding to the sections lines reported in the theoretical plots in Figs.(3.4.C-D). The theoretical error bands in Fig. (3.5) are computed via a numerical simulation of the experiment. The experimental point show a good accordance with the theoretical model, with the majority of the data falling in the confidence region at one standard deviation. All three panels, show a clear quantum advantage in all the region reported. In accordance with theory, the maximum gain is shown to increase with the mean signal energy but at the expenses of a narrowing of the region in which the quantum enhancement can be found, as discussed previously.

Ref. [1] provided the first experimental demonstration of the quantum reading protocol, showing how entanglement is able to boost the retrieval of classical information from an optical memory cell, outperforming any classical strategy for the same number of input photons. It is particularly interesting how this advantage can be obtained with a relatively simple receiver design, based on direct PC measurements, as opposed to other protocols that may require highly-theoretical, joint quantum measurements. Moreover the resilience to losses showed by the protocol makes it a good candidate for possible future application of quantum technologies. The possible applications go beyond the simple memory model to different scenarios based on the discrimination of bosonic loss channels.

In the following section we will discuss another protocol similar to the Quantum reading one, and that can in fact be considered a generalization of it, the Quantum Conformance Test protocol proposed for the first time in Ref. [2] and an original contribution to this Thesis.
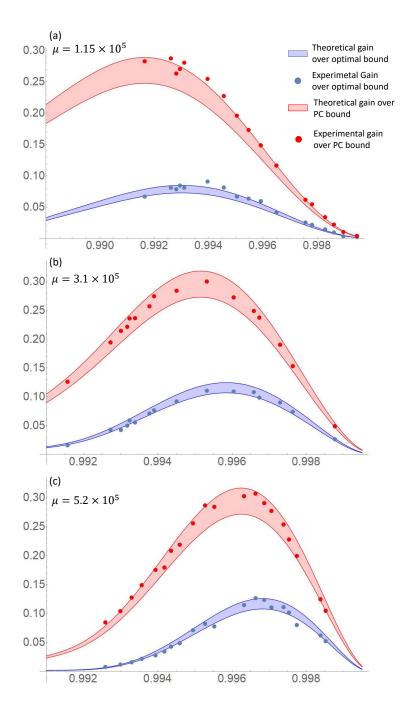
Fig. 3.5 *Experimental gain G of quantum reading*. The experimental gain in bits is reported as a function of the lower transmittance τ_0, while $\tau_1 = 1$. Each of the three panels refer to a different mean photon number μ in the signal beam: (a) $\mu = 1.15 \cdot 10^5$, (b) $3.1 \cdot 10^5$, and (c) $5.2 \cdot 10^5$. Blue data refers to the gain with respect to the classical optimal bound in Eq. (3.3). Red data refers to the gain with respect to the classical photon-counting bound given in Eq. (3.28). The experimental parameters are estimated independently in a calibration step and are the mean signal energy μ, the detection efficiency of signal and idler channel η_S and η_I and the electronic noise ν_e. While the value of μ is intentionally varied in the three panels, the other parameters are kept fixed to: $\eta_S = 0.78$, $\eta_I = 0.77$, and $\nu_e \sim 10^4$. The plots are reported from Ref. [1].

Chapter 4

Quantum Conformance Test and its applications

4.1 Quantum conformance test

In this section, we will discuss the Quantum Conformance Test (QCT) protocol. The QCT protocol was originally proposed recently in Ref. [2]. Its goal is to show that there is an advantage in the use of quantum states in monitoring production processes, modeled as ensembles of end-products having a given probability distribution, within the formalism of quantum information. The same construction used to analyse the QCT protocol, namely the analysis of the discrimination of convex combinations of quantum channels can be used to analyze a variety of different physical situations, e.g. to analyse dissipation in communication channels, and will be also used in the next section to generalize the quantum reading protocol to the readout of imperfect memory cells.

4.1.1 QCT theoretical model

In this subsection, we provide a theoretical formulation of the QCT protocol. Let us define a process $\mathscr{P}_x$ as an ensemble $\{g_x(\theta)\mathscr{E}_\theta\}$, producing a quantum object $\mathscr{E}_\theta$, parameterized by θ. The parameter θ is extracted by a given set $\mathscr{A}$, which can either be discrete or continuous, according to the probability distribution $g_x(\theta)$. The subscript x denotes a binary classical random variable $x = \{0, 1\}$. We suppose
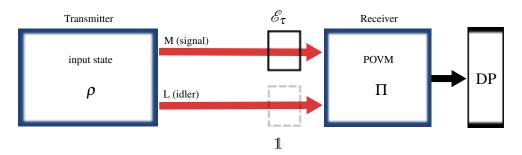
Fig. 4.1 *Quantum conformance test scheme.* A probe state ρ sends M signal modes through a SUT, modelled as a pure loss channel $\mathscr{E}_\tau$. L idler modes are sent directly reach the receiver. $\mathbb{1}$ represents the identity operator. A POVM Π is performed on the output state. A decision, y, is taken using the result of this measurement and data processing (DP), so that the process generating $\mathscr{E}_\tau$ is identified as conform ($y = 0$) or defective ($y = 1$). The scheme is taken from Ref. [2].

that the quantum objects $\mathscr{E}_\theta$ can be produced either by a *reference* process $\mathscr{P}_0$ or a *defective* one $\mathscr{P}_1$, occurring with equal prior probability, i.e. $\pi_x := p(\mathscr{P}_x) = 1/2$. The quantum conformance test consists in irradiating a quantum state ρ over a system under test (SUT) and performing a measurement on ρ after the interaction to determine what process $\mathscr{P}_x$ produced the SUT. A schematic of the QCT protocol is reported in Fig.(4.1). A generic bosonic bipartite state ρ consisting of M signal modes and L idler ones is irradiated over the SUT. After the interaction a generic POVM, Π, having elements $\{\Pi_i\}$ is performed, followed by classical post processing to recover the classical variable $y = \{0,1\}$ giving a guess on the nature x of the producing process $\mathscr{P}_x$. The test is deemed successful if $y = x$, i.e. if the guess is correct. Here, we have to make a difference between the two possible kind of incorrect guesses. In the previous chapter, when analyzing the quantum reading protocol, there was no need to make this distinction as any kind of error would lead to the same result, namely to the same information loss. In the QCT protocol however $x = 0$ and $x = 1$ describe two different physical situations. Hence different errors can have different significance and should be treated differently. The possible scenarios are:

- *False negative*. In this scenario, a SUT produced by a conform process ($x = 0$) is labeled as defective ($y = 1$). In an industrial context, this kind of outcome can be seen as an economic loss for a manufacturer, as a conform process is considered defective, and discarded for no reason. The probability of false negatives will be denoted as p_{10}.

- *False positive*. A SUT produced by a defective process ($x = 1$) is labeled as conform ($y = 0$). This outcome represents a risk since possibly unsafe products are released. The false-positive probability will be referred to as p_{01}.

A direct consequence of the possible asymmetry on the error is that the cost function is not always the probability of error, and vastly depends on the situation at hand. The analysis of false positives and false negatives and the analysis of their tolerance level plays a central role in conformance testing, and will be discussed more in depth later in this section. To evaluate the quantum advantage in a general scenario, however, the total probability of error, p_{err}:

$$p_{err} = \frac{1}{2}\left(p_{01} + p_{10}\right) \tag{4.1}$$

remains a balanced cost function. While the quantitative results may slightly vary in function of the different costs related to different scenarios, the analysis of the probability of error gives a good indication of the advantage that the use of quantum resources can provide.

While the formalism introduced above can be used to analyze any kind of parameterized channel $\mathscr{E}_\theta$, explicit results need the nature of the channel to be specified. Ref. [2] considered the *paradigmatic case of pure loss channels*, and in the following we will report the results obtained in this scenario. We will start by analysing the performance that can be obtained by classical states, in the form of Eq. (3.2), in the QCT protocol, and subsequently we will show that quantum states can guarantee an advantage over any classical strategy.

Consider a Quantum conformance test, performed, according to the schematic of Fig.(4.1), on a SUT that can be modeled by a pure loss channel $\mathscr{E}_\tau$. After the interaction of a bipartite state ρ with the SUT, depending on the production process $\mathscr{P}_x$ the initial state will be mapped into either:

$$\begin{aligned}
\rho_0 &= \mathbb{E}_{\mathscr{P}_0}\left[(\mathscr{E}_\tau \otimes \mathbb{1})\rho\right] \rightarrow \text{reference process} \\
\rho_1 &= \mathbb{E}_{\mathscr{P}_1}\left[(\mathscr{E}_\tau \otimes \mathbb{1})\rho\right] \rightarrow \text{defective process}
\end{aligned} \tag{4.2}$$

where $\mathbb{1}$ denotes the identity on the idler modes, and $\mathbb{E}_{\mathscr{P}}[\cdot]$ the expectation value over the ensemble $\mathscr{P}$. In general, the supports of the output states ρ_0 and ρ_1 will overlap, meaning that the discrimination between the reference and the defective

process after a POVM Π will be affected by a probability of error $p_{err}^{\rho,\Pi}(\mathscr{P}_0,\mathscr{P}_1)$, that is a function of the input state and measurement chosen as well as the processes considered. As anticipated to get a general sense of the performance of different strategies we will consider the minimization of the probability of error over all possible input states and output measurements:

$$p_{err}(\mathscr{P}_0,\mathscr{P}_1) = \min_{\Pi}\left[\min_{\rho}\left[p_{err}^{\rho,\Pi}(\mathscr{P}_0,\mathscr{P}_1)\right]\right] \qquad (4.3)$$

Since we are dealing with bosonic input states, as it was the case in the quantum reading protocol, the minimization in Eq. (4.3) has to be performed under some energy constraint. Note how in this case the lowest possible value of $p_{err}(\mathscr{P}_0,\mathscr{P}_1)$, even for infinite energy, is not always zero rather a finite value given by the overlap of the classical probability distributions, $g_0(\tau)$ and $g_1(\tau)$, characterizing the two processes $\mathscr{P}_0$ and $\mathscr{P}_1$. This is represented in Fig. (4.3). In the following we fix the mean number of signal photons to a given value μ. Physically this condition makes sense as one would like to limit the energy irradiated over the SUTs, that in some case of interest may be sensitive to the testing. In the following subsection we will give a bound on the performance of classical states assuming an optimal measurement.

4.1.2   Absolute bound for classical states

As introduced in the previous chapter, classical states are defined as the states having a positive P-representation [71] and can be represented in the form of Eq. (3.2). Consider a classical input state ρ^{cla} with fixed mean signal energy μ:

$$\int d^{2M}\boldsymbol{\alpha}\, d^{2L}\boldsymbol{\beta}\, P(\boldsymbol{\alpha},\boldsymbol{\beta})\,|\boldsymbol{\alpha}|^2 = \mu \qquad (4.4)$$

The output states, after interaction with the SUT, are defined by Eq. (4.2) and are denoted here as ρ_0^{cla} and ρ_1^{cla}. Let us formally define the best classical performance, p_{err}^{cla}, as:

$$p_{err}^{cla} = \min_{\rho_{cla}}\left[\min_{\Pi}\left[p_{err}^{\rho,\Pi}(\mathscr{P}_0,\mathscr{P}_1)\right]\right] \qquad (4.5)$$

where the outer minimum is over the subset of classical states. Fixing the input state, the minimum probability of error for the QCT protocol that can be obtained using

classical states is equal to the minimum probability of error in the discrimination of ρ_0^{cla} and ρ_1^{cla}. The best performance in this task is given once again by the Helstrom formula:

$$p_{err}(\rho_0^{cla}, \rho_1^{cla}) := \min_{\Pi} \left[p_{err}^{\rho_{cla}, \Pi}(\mathscr{P}_0, \mathscr{P}_1) \right] = \frac{1}{2} \left(1 - D\left(\rho_0^{cla}, \rho_1^{cla} \right) \right) \qquad (4.6)$$

where we used $D(\rho_0, \rho_1) = ||\rho_0 - \rho_1||/2$ to denote the trace distance. In view of Eq. (4.6) the quantity in Eq.(4.5) can be lower bounded by upper bounding the trace distance, $D\left(\rho_0^{cla}, \rho_1^{cla} \right)$, for a generic classical input state ρ_{cla}. Let us consider first a scenario in which only the defective process $\mathscr{P}_1$ is distributed according to some arbitrary distribution $g_1(\tau)$, while the reference process $\mathscr{P}_0$ is strongly peaked, i.e. it can be assumed that all the ensemble is composed of a single value of the transmittance parameter τ_0. In this situation, for any classical input state ρ_{cla}, we can use the strong convexity of the trace distance to write:

$$D(\rho_0^{cla}, \rho_1^{cla}) \leq \int d^{2M}\boldsymbol{\alpha}\, d^{2L}\boldsymbol{\beta}\, P(\boldsymbol{\alpha}, \boldsymbol{\beta}) \times$$
$$\times D(\mathscr{E}_{\tau_0}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|) \otimes |\boldsymbol{\beta}\rangle\langle\boldsymbol{\beta}|, \mathbb{E}_{\mathscr{P}_1}[\mathscr{E}_{\tau}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|)] \otimes |\boldsymbol{\beta}\rangle\langle\boldsymbol{\beta}|)$$
$$= \int d^{2M}\boldsymbol{\alpha}\, P(\boldsymbol{\alpha}) D(\mathscr{E}_{\tau_0}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|), \mathbb{E}_{\mathscr{P}_1}[\mathscr{E}_{\tau}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|)]), \qquad (4.7)$$

where the expectation value $\mathbb{E}_{\mathscr{P}_1}[\cdot]$ is taken over the distribution $g_1(\tau)$ and the last equality follows from the fact that the unaffected reference system doesn't change the trace distance. Using once again the convexity of the trace distance we have:

$$D(\mathscr{E}_{\tau_0}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|), \mathbb{E}_{\mathscr{P}_1}[\mathscr{E}_{\tau}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|)])$$
$$\leq \mathbb{E}_{\mathscr{P}_1}[D(\mathscr{E}_{\tau_0}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|), \mathscr{E}_{\tau}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|))]$$
$$= \mathbb{E}_{\mathscr{P}_1}[D(|\sqrt{\tau_0}\boldsymbol{\alpha}\rangle\langle\sqrt{\tau_0}\boldsymbol{\alpha}|, |\sqrt{\tau}\boldsymbol{\alpha}\rangle\langle\sqrt{\tau}\boldsymbol{\alpha}|)] \qquad (4.8)$$

To derive this expression, we used the fact that pure loss channels map coherent states into coherent states, i.e. $|\alpha\rangle \xrightarrow{\mathscr{E}_{\tau}} |\sqrt{\tau}\alpha\rangle$. The distance between two pure states can be explicitly calculated as:

$$D\left(|\sqrt{\tau_0}\boldsymbol{\alpha}\rangle, |\sqrt{\tau}\boldsymbol{\alpha}\rangle \right) = \sqrt{1 - F(|\sqrt{\tau_0}\boldsymbol{\alpha}\rangle, |\sqrt{\tau}\boldsymbol{\alpha}\rangle)^2}$$
$$= \sqrt{1 - e^{-|\boldsymbol{\alpha}|^2(\sqrt{\tau_0} - \sqrt{\tau})^2}} \qquad (4.9)$$

with $F(\cdot,\cdot)$ is the fidelity between two quantum states, that for pure state is the square modulus of their overlap. Thus we can write:

$$D(\rho_0,\rho_1) \le \mathbb{E}_{\mathscr{P}_1}\left[\int d^{2M}\boldsymbol{\alpha}\, P(\boldsymbol{\alpha})\sqrt{1-e^{-|\boldsymbol{\alpha}|^2(\sqrt{\tau_0}-\sqrt{\tau})^2}}\,\right] \qquad (4.10)$$

Given the condition in Eq. (4.4) on the mean number of photons on the signal system μ we have (for proof see Supplementary Materials of Ref. [48]):

$$\int d^{2M}\boldsymbol{\alpha}\, P(\boldsymbol{\alpha})\sqrt{1-e^{-|\boldsymbol{\alpha}|^2(\sqrt{\tau_0}-\sqrt{\tau})^2}}$$
$$\le \sqrt{1-e^{-\mu(\sqrt{\tau_0}-\sqrt{\tau})^2}} \qquad (4.11)$$

for any probability distribution $P(\boldsymbol{\alpha})$. Using the above inequality in Eq. (4.10) gives:

$$D(\rho_0,\rho_1) \le \mathbb{E}_{\mathscr{P}_1}\left[\sqrt{1-e^{-\mu(\sqrt{\tau_0}-\sqrt{\tau})^2}}\,\right] \qquad (4.12)$$

That yields the following lower bound for the performance of classical states in the QCT protocol when one of the processes is strongly peaked:

$$p_{err}^{cla} \ge \frac{1-\mathbb{E}_{\mathscr{P}_1}\left[\sqrt{1-e^{-\mu(\sqrt{\tau_0}-\sqrt{\tau})^2}}\,\right]}{2} \qquad (4.13)$$

Now if we relax the condition of strongly peaked reference process, i.e. we let $\mathscr{P}_0$ be distributed as $g_0(\tau_0)$, we can use the convexity on the trace distance on both arguments to have in place of Eq. (B.3) the following inequality:

$$D(\mathbb{E}_{\mathscr{P}_0}[\mathscr{E}_\tau(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|)],\mathbb{E}_{\mathscr{P}_1}[\mathscr{E}_\tau(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|)]) \le$$
$$\mathbb{E}_{\mathscr{P}_0}\mathbb{E}_{\mathscr{P}_1}[D(\mathscr{E}_{\tau_0}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|),\mathscr{E}_{\tau_1}(|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|))] =$$
$$= \mathbb{E}_{\mathscr{P}_0}\mathbb{E}_{\mathscr{P}_1}[D(|\sqrt{\tau_0}\boldsymbol{\alpha}\rangle\langle\sqrt{\tau_0}\boldsymbol{\alpha}|,|\sqrt{\tau_1}\boldsymbol{\alpha}\rangle\langle\sqrt{\tau_1}\boldsymbol{\alpha}|)] \qquad (4.14)$$

Repeating the previous steps we can then lower bound p_{err}^{cla} for two arbitrary processes as:

$$p_{err}^{cla} \ge \mathscr{C} := \frac{1-\mathbb{E}_{\mathscr{P}_0}\mathbb{E}_{\mathscr{P}_1}\left[\sqrt{1-e^{-\mu(\sqrt{\tau_0}-\sqrt{\tau_1})^2}}\,\right]}{2}, \qquad (4.15)$$

The quantity $\mathscr{C}$ establishes a lower bound for the discrimination error probability when considering classical input states and an optimal measurement strategy. Note

how this bound is not tight, which means it may not be reached by any classical receiver.

4.1.3   Classical states and PC receiver

In the previous section we derived $\mathscr{C}$, the limit on the performance that can be achieved with an optimal classical strategy, i.e. optimal classical input states and an unspecified optimal receiver. Since this bound is not tight, it is not guaranteed that it can be reached. Even in the case in which a strategy could be found that saturates it, its implementation may be of difficult practical realization. For this reason in this section we consider the case in which classical states are paired with the PC receiver described in the previous chapter. The analysis of the best classical performance achievable with the PC receiver gives a second classical benchmark, whose performance can be experimentally validated.

Analogously to the quantum reading protocol we have that also in the QCT case for classical input states the use of idler modes does not improve the performance. Let us remind from Chapter (3) that, for classical states, the Poisson photon number distribution has the minimum variance allowed at a fixed mean value. This means that the best performance is achieved using signal states having a Poisson photon number distribution. Fig. (4.3) offers a visualization of this. Indeed, the error probability p_{err} is proportional to the overlap of the measurement outcome distributions which, in the case of a PC measurement, are the photon number distributions. Consequently an input state with a narrower photon number distribution will lead to better discrimination performances.

The total probability of error, p_{err}, of the procedure can be evaluated as the average over the two possible processes, $\mathscr{P}_0$ and $\mathscr{P}_1$, of the conditioned probabilities of error:

$$p_{err} = \frac{1}{2}(p_{01} + p_{10}) \tag{4.16}$$

where according to the formalism introduced above p_{ij} is the probability that the discrimination procedure will give the outcome i conditioned to the SUT being generated by the process $\mathscr{P}_j$. Since with a PC receiver the best classical performance is achieved by input states having a Poisson photon number distribution, $\mathbf{P}_\mu$, we evaluate Eq. (4.16) in this case. In order to do so, we exploit the result reported in Eq. (3.26) of Chapter (3), the effect of a pure loss channel $\mathscr{E}_\tau$ on the photon number

distribution, $\mathbf{P}_\mu(n)$, of a state Poisson distributed with parameter μ, is simply a rescaling of the mean photon number:

$$\mathbf{P}_\mu(n) \xrightarrow{\mathscr{E}_\tau} \mathbf{P}_{\tau\mu}(n) \tag{4.17}$$

A SUT generated by the process $\mathscr{P}_x$ has a transmittance τ distributed with known probability density function $g_x(\tau)$. The conditioned photon number distribution, $p(n|\mathscr{P}_x)$, can then be evaluated by taking the expected value over $g_x(\tau)$ of the distribution conditioned on the specific τ:

$$h^{(x)}(n) := p(n|\mathscr{P}_x) = \langle p(n|\tau)\rangle_g = \int_0^1 \mathbf{P}_{\tau\mu}(n)g_x(\tau)d\tau \tag{4.18}$$

The probability of error can be expressed in terms of the distribution in Eq. (4.18) using the fact that:

$$p_{ij} = \sum_{\{n|y=i\}} p(n|\mathscr{P}_j) \tag{4.19}$$

where the sum is over all values of n such that the outcome of the decision is $y = i$. The integral in Eq. (4.18) defines a compound distribution. We have then:

$$\langle n\rangle_h = \left\langle \langle (n|\tau)\rangle_{\mathbf{P}_{\tau\mu}} \right\rangle_g = \mu\langle \tau\rangle_g \tag{4.20}$$

$$\langle \Delta^2 n\rangle_h = \left\langle \langle \Delta^2(n|\tau)\rangle_{\mathbf{P}_{\tau\mu}} \right\rangle_g + \left\langle \Delta^2\langle (n|\tau)\rangle_{\mathbf{P}_{\tau\mu}} \right\rangle_g =$$
$$= \mu\langle \tau\rangle_g + \mu^2\langle \Delta^2\tau\rangle_g \tag{4.21}$$

To proceed in the analysis of the probability of error let us start by considering the case in which both the processes distributions, $g_x(\tau)$, have Gaussian form with mean $\bar{\tau}_x$ and variance σ_x^2, i.e. $g_x(\tau) = \mathbf{G}_x(\tau)$. According to Eq. (4.20-4.21), we have for the final photon number distribution, $h_{\mathbf{G}_x}(n)$:

$$N_x := \langle n\rangle_h = \mu\bar{\tau}_x \tag{4.22}$$

$$s_x^2 := \langle \Delta^2 n\rangle_h = \mu\bar{\tau}_x + \mu^2\sigma_x^2 \tag{4.23}$$

If μ is large enough then the distribution $h_G(n)$ will converge to a Gaussian one itself:

$$h_{\mathbf{G}_x} \sim \mathbf{G}_{N_x, s_x^2}(n) \qquad\qquad \mu >> 0 \tag{4.24}$$

where $G_{N_x, s_x^2}(n)$ denotes a Gaussian distribution on the outcome n having mean N_x and variance s_x^2. In this approximation we can evaluate the probability of error according to Eq. (4.16) and (4.19) as:

$$p_{err} = \frac{1}{2} \left(\sum_{\{n|y=1\}} p(n|\mathscr{P}_0) + \sum_{\{n|y=0\}} p(n|\mathscr{P}_1) \right)$$
$$\approx \frac{1}{2} \left(\int_{\{n|y=1\}} \mathbf{G}_{N_0, s_0^2}(n) + \int_{\{n|y=0\}} \mathbf{G}_{N_1, s_1^2}(n) \right) \tag{4.25}$$

when used once again the assumption $\mu >> 0$ in Eq. (4.24) to approximate the the sums with integrals over n.

To determine the regions of integration, we impose the condition $\mathbf{G}_{N_0, s_0^2}(n_{th}) = \mathbf{G}_{N_1, s_1^2}(n_{th})$. This condition defines up to two threshold values, n_{th}, for n, whose expression if $s_0 \neq s_1$ can be computed to be:

$$n_{th}^{(-)} = \frac{N_0 s_1^2 - N_1 s_0^2 - \kappa}{(s_1 - s_0)(s_1 + s_0)}$$
$$n_{th}^{(+)} = \frac{N_0 s_1^2 - N_1 s_0^2 + \kappa}{(s_1 - s_0)(s_1 + s_0)}$$
$$\kappa := \sqrt{s_0^2 s_1^2 \left(2(s_1 - s_0)(s_1 + s_0)(\log s_1 - \log s_0) + (N_1 - N_0)^2 \right)}$$

and, if $s_0 = s_1$:
$$n_{th}^{(0)} = \frac{N_0 + N_1}{2} \tag{4.26}$$

Consider the first case, i.e. $s_0 \neq s_1$. The case $s_0 = s_1$ can be treated in a similar way with a few adjustment so we won't report it. We assume, without loss of generality,

$$s_1 > s_0,$$

we have $n_{th}^{(+)} > n_{th}^{(-)}$, and the value $y = 1$ will be selected if $n > n_{th}^{(+)}$ or $n < n_{th}^{(-)}$, so that:

$$p_{err} \approx \frac{1}{2} \left(\int_{-\infty}^{n_{th}^{(-)}} \mathbf{G}_{N_0, s_0^2}(n) + \int_{n_{th}^{(-)}}^{n_{th}^{(+)}} \mathbf{G}_{N_1, s_1^2}(n) + \int_{n_{th}^{(+)}}^{\infty} \mathbf{G}_{N_0, s_0^2}(n) \right) \tag{4.27}$$

We note how the first integral should run from 0 rather than $-\infty$ but since we are considering $\mu >> 0$ we can approximate the exact range of integration with the latter ($n_{th}^{(-)}$ can take also negative values but in that case the integral will not contribute to p_{err}). Let us define:

$$E^x(n_{th}) = \text{erf}\left[\frac{n_{th} - N_x}{\sqrt{2}s_x}\right] \tag{4.28}$$

Where $\text{erf}(x)$ is the error function. We have:

$$p_{err} \approx \frac{1}{2}\left(1 - q_G\right) \tag{4.29}$$

$$q_G := \frac{1}{2}\left(E^0(n_{th}^{(+)}) + E^1(n_{th}^{(-)}) - E^0(n_{th}^{(-)}) - E^1(n_{th}^{(+)})\right) \tag{4.30}$$

To simplify the expression above, we note how, if $s_0 \approx s_1$ one of the threshold values will be far from the center of both distributions so that it can be neglected and the solution can be considered unique. In this approximation we have:

$$q_G \approx \begin{cases} q_G^{(+)} & \text{if } \bar{\tau}_0 \leq \bar{\tau}_1 \\ q_G^{(-)} & \text{if } \bar{\tau}_0 > \bar{\tau}_1 \end{cases} \tag{4.31}$$

where:

$$q_G^{(+)} := \frac{1}{2}\left(E^0(n_{th}^{(+)}) - E^1(n_{th}^{(+)})\right) \tag{4.32}$$

$$q_G^{(-)} := \frac{1}{2}\left(E^1(n_{th}^{(-)}) - E^0(n_{th}^{(-)})\right) \tag{4.33}$$

Eq. (4.29-4.33) provide an approximation for the best performance of classical states in the QCT protocol when the processes considered are Gaussian distributed.

Another interesting case to analyze is the one of *uniformly distributed* processes. Consider a uniform distribution having mean $\bar{\tau}$ and half width δ, $\mathbf{U}_{\bar{\tau},\delta}(\tau)$ the integral

in Eq. (4.18) can be evaluated analytically. We have:

$$
\begin{aligned}
h_{\mathbf{U}}(n) = p(n|\mathscr{P}_x) &= \frac{1}{2\delta} \int_{\bar{\tau}-\delta}^{\bar{\tau}+\delta} \mathbf{P}_{\tau\mu}(n)d\tau = \\
&= \frac{\Gamma(n+1,\mu(\bar{\tau}-\delta)) - \Gamma(n+1,\mu(\bar{\tau}+\delta))}{2\delta\mu n!}
\end{aligned}
\tag{4.34}
$$

where $\Gamma(a,b)$ is the incomplete gamma function, and for brevity we omit the dependence on x of h and the parameters of the distribution. Using Eq. (4.20)-(4.21) we have:

$$
N_x = \langle n \rangle_h = \bar{\tau}_x \mu
\tag{4.35}
$$

$$
s_x^2 = \langle \Delta^2 n \rangle_h = \bar{\tau}_x \mu + \frac{1}{3}\delta_x^2 \mu^2
\tag{4.36}
$$

If as in previous section, we assume, $\mu \gg 0$, the distribution in Eq. (4.34) has two noteworthy extreme cases:

1. if $s_x > \mu\delta_x$, $h_U(n)$ can be approximated by a Gaussian distribution, $h_U(n) \sim G_{N_x,s_x^2}(n)$.

2. if $s_x \ll \mu\delta_x$, $h_U(n)$ can be approximated with a uniform one, $h_U(n) \sim \mathbf{U}_{N_x,\mu\delta_x}(\tau)$.

These approximations in the two regimes are shown in Fig. (4.2).

In the first case, since the final distribution is a Gaussian one, the results above can be used in the evaluation.

The previous analyses are given as examples for some notable distributions. In a more general scenario the probability of error depends on both the distributions at hand, g_0 and g_1. While sometimes it will be possible to find either exact or approximate analytical solutions for the probability of error, as it was the case for the Gaussian and uniform cases considered above, when this is not possible the probability of error can be evaluated numerically. Let us define the generic best performance of classical states paired with PC measurements as:

$$
\mathscr{C}^{PC} := p_{err}^{cla,PC} = \frac{1}{2}\left(1 - q_g\right)
\tag{4.37}
$$

where it is understood that the form of q_g is determined by both the distributions g_x.
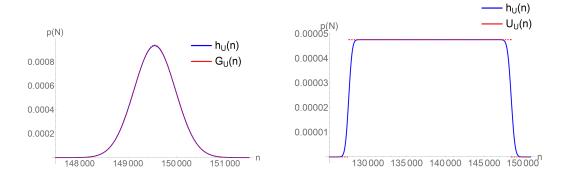
Fig. 4.2 *Limiting cases for the distribution $h_x(n)$.* In figure are shown the two limiting cases for the distribution $h_U(n)$, in case of uniform process distribution, as described in the main text. The panel on the left represents case (1), $s_x > \mu\delta$. The one on the right case (2), $s_x << \mu\delta$. The image is taken from the Supplementary materials of Ref. [2].

In the next section we will analyse the performance TMSV input states.

4.1.4   TMSV states and PC receiver

In the following we will provide an overview of the quantum strategy that we used to surpass the best classical performance $\mathscr{C}$, thus proving a quantum advantage. The input state used ρ, is the same used in the Quantum reading protocol and discussed in the previous chapter, namely a collection of M replicas of a two-mode squeezed vacuum (TMSV) state. The receiver used is the PC one.

According to the analysis of Sec(3.2.1) the probability of error is found using Bayes theorem that for the QCT protocol yields:

$$p(\mathscr{P}_x|\boldsymbol{n}) = \frac{p(\boldsymbol{n}|\mathscr{P}_x)\,p(\mathscr{P}_x)}{p(\boldsymbol{n})} = \frac{p(\boldsymbol{n}|\mathscr{P}_x)}{p(\boldsymbol{n}|\mathscr{P}_0) + p(\boldsymbol{n}|\mathscr{P}_1)} \qquad (4.38)$$

where we used the fact that the defective and reference processes are equiprobable, $p(\mathscr{P}_0) = p(\mathscr{P}_1) = 1/2$, and $p(\boldsymbol{n}|\mathscr{P}_x) = \langle n_S, n_I | \rho_x | n_S, n_I \rangle$. If we consider a single value of transmittance τ from Sec.(3.2.1) we can write:

$$p(\boldsymbol{n}|\tau) = \sum_{m=n_S}^{\infty} P_{\mu,M}(m, n_I)\,\mathscr{B}(n_S|m, \tau). \qquad (4.39)$$
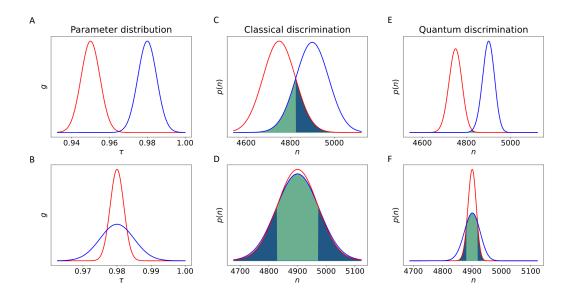
Fig. 4.3 *Visualization of error probabilities in the QCT protocol.* Panels **A-B** show the distributions of the reference process (red), $\mathscr{P}_0$, and defective one (blue), $\mathscr{P}_1$. The two rows present the two cases of big and very small initial overlap. Panels **C-D** show the resulting photon number distribution $p(n)$ when classical state is used as a probe. The overlaps between the two distributions, highlighted in green and blue, the two colors distinguishing the cases where the reference or the defective process are most likely, are visualizations of the conditional error probabilities, at a given value of τ, while their weighted sum gives an visualization of the total error probability. The advantage offered by quantum correlation can be seen in Panels **E-F** that display the case in which quantum probes are used. The plots are taken from Ref. [2].

and then making use the linearity of quantum operations the distribution conditioned to the process $\mathscr{P}_x$ becomes:

$$p\left(\boldsymbol{n}|\mathscr{P}_x\right) = \int_{\mathscr{A}_x} p\left(\boldsymbol{n}|\tau\right) g_x\left(\tau\right) d\tau. \tag{4.40}$$

Naturally all the approximation that we made in the previous chapter on the distribution $p\left(\boldsymbol{n}|\tau\right)$ are still valid as long as we remain in the regime of high number of modes $M >> 0$ and not too low signal energy μ. These conditions are met in all the range of theoretical and experimental exploration in the following, so that we can evaluate Eq. (4.40) using Eq. (3.38) and according to Eq. (3.35) the probability of error is:

$$\mathscr{Q} := p_{err}^{qua,PC} = \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dn_I dn_S \min_x \left(p\left(\boldsymbol{n}|\mathscr{P}_x\right) \right), \tag{4.41}$$

$$p\left(\boldsymbol{n}|\mathscr{P}_x\right) = \int_{\mathscr{A}_x} \mathscr{N}\left(\mathbf{n}, \Sigma(\tau_x)\right) g_x\left(\tau\right) d\tau.$$

where $\mathscr{N}\left(\mathbf{n}, \Sigma(\tau_x)\right)$ is a multivariate normal distribution on the photon number having mean $\bar{\mathbf{n}}$ and covariance matrix given by Eq. (3.40). We conclude this section by giving a visualization of the advantage provided by quantum correlated states. In panels **A-B** of Fig. (4.3) two possible scenarios for the initial distributions of conform and defective processes, $g_0(\tau)$ and $g_1(\tau)$, are shown. In panel **A**, two distributions whose overlap is negligible are considered, whereas, in panel **B**, the parameter distributions are significantly overlapping. In panels **C-F**, the photon number probability distributions, from which the outcome y is determined, is shown both in the case of classical (**C-D**) and quantum (**E-F**) input states. The initial process distributions are convoluted with the noise emerging from the measurement, associated with each probe state. One can observe that the noise present in the state which probes the SUT translates into an error in the discrimination. In both cases considered, it is clear how the effect of the quantum correlation on the marginal measurement distribution is a dramatic narrowing leading to a significant reduction in the overlap and thus on the probability of error.

4.1.5 Non-ideal efficiency

Similarly to the QR protocol discussed in the previous chapter, the most relevant experimental effects are losses due to imperfect detectors and optical elements.
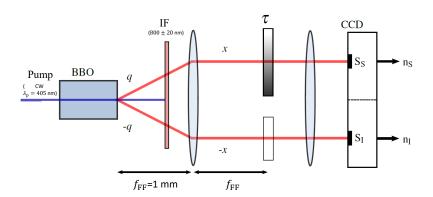
Fig. 4.4 *Schematic of the experimental setup for QCT*. The experimental scheme is analogous to the one used for the Quantum reading. The multi-mode TMSV state is generated pumping a BBO crystal with a laser at 405 nm. The correlation in momentum is converted into correlation in position in the sample plane using a lens of focal f_{FF}. The sample is imaged on the CCD camera using a second lens. The signal beam passes through the sample of transmittance τ and is then detected in the S_S region of the CCD, while the idler beam goes directly to S_I, without interacting with the sample. The idler optical path is matched with the sample's one using a non-absorbing glass. n_S and n_I are collected integrating the signals over the two detection regions. The scheme is taken from Ref. [2]

These effects can be encompassed by a single coefficient, $\eta \leq 1$, the efficiency of the channel. According to the discussion of Sec.(3.2.3) the effect of the efficiency can be evaluated using the fact that the action of a loss η and the one of the SUT, commute. For classical probes, the effect of η is to reduce the energy available for the discrimination and in the classical case the bounds found can be evaluated accounting for the inefficiency by performing the substitution $\mu \rightarrow \eta\mu$ in the formulas referring to the ideal efficiency case. In the quantum case an additional effect is the reduction of correlation between signal and idler. Due to this reduction in correlation, in general it is expected that the quantum advantage will be reduced as η becomes lower. This is accounted for in the covariance matrix Σ, appearing in Eq. (4.41), whose expression is given in Eq. (3.39).

In the following sections, we will discuss the experimental setup used to realize the QCT protocol and the data analysis.

4.1.6   Experimental setup

The experimental setup is analogous to the one used for the Quantum Reading realization described in Sec.(3.2.4) of the previous chapter. The experiment is based on spontaneous parametric down-conversion, generating a multi-mode TMSV state in a 1 cm^3 type-II β-barium borate (BBO) crystal pumped by a continuous-wave laser at $\lambda_p = 405$ nm, delivering a power of 100 mW. The correlation in momentum of the down-converted photons is mapped into spatial correlations using a lens in f-f configuration. The number of correlated photons in the signal and idler branch is measured by integrating the signal in two selected region of a CCD camera operated in the conventional mode. The number of counts of the camera is indeed proportional to the number of photons. The proportionality coefficient, named electronic gain, is calibrated by a specific procedure. A schematic is given in Fig. (4.4). The SUT is an absorption sample consisting of a coated glass plate, which presents different transmittance regions, realized with depositions of variable density. To match the optical paths of both beams a blank coated glass is inserted in the idler beam's path. While the experimental parameters explored are different from the one used in the QR experiment the number of spatial-temporal modes collected is still very large ($\sim 10^{13}$) and the mean occupation number per mode is very small ($\sim 10^{-8}$) meaning that the condition required for the approximation made in the sections above are always met in the experimental regime explored. The detection efficiencies η_S and η_I are estimated, along with the other experimental parameters, in a preliminary characterization by exploiting the correlations of the SPDC process [79, 76]. Since the final objective of the QCT protocol is the discrimination of two distribution we collected a large number of measurement, $N_{\mathscr{D}_\tau} = 2 \cdot 10^4$, for different values of transmittance τ, thus creating the experimental dataset that we used to mimic arbitrary distributions with the procedure described in the next section.

4.1.7   Data analysis

In this section we will give an overview on how the data needed to evaluate the experimental probabilities of error was collected and analyzed. A more technical and detailed discussion on the data analysis is reported in Appendix C.

   According to Eq. (4.1) the total probability of error is composed of the two conditional contributions p_{10} and p_{01}. Experimentally this contributions are estimated

separately. As it will be discussed in the following sections, to reduce the number of parameters, the reference process, $\mathscr{P}_0$ will be considered strongly peaked around a value τ_0. Under this condition a dataset, $\mathscr{D}$, conditioned to the process being $\mathscr{P}_0$ can be created simply by taking $N_{\mathscr{D}}$ measurements with a SUT having transmittance τ_0. The outcome of the measurement can then be processed according to the discussion of the previous section to assign a label, either 0 or 1. Since the dataset is generated by $\mathscr{P}_0$ the frequency of error $f_{\mathscr{D}}(\mathscr{P}_1|\mathscr{P}_0)$ can be estimated simply by counting the number of instances in which the process was mislabeled as 1 over the total number of measurements. The frequency $f_{\mathscr{D}}(\mathscr{P}_1|\mathscr{P}_0)$ will then converge to the conditional probability of error p_{10} for a very large number of measurements.

The defective process, $\mathscr{P}_1$, will be distributed according to some arbitrary distribution $g_1(\tau)$. In order to create a dataset $\mathscr{D}$, conditioned to the process being $\mathscr{P}_1$ the measurements must be taken on SUTs having transmittance distributed according to $f(\tau) \sim g_1(\tau)$, i.e. we want the experimental distribution of the transmittance values probed to be as close as possible to the theoretical one of $\mathscr{P}_1$. While this could be done *a priori* it is more convenient, as we argue in Appendix C, to take measurements uniformly across a wide enough range of transmittance values and acting *a posteriori*, with a procedure of statistical weighting. This procedure consist in discarding measurements on certain values of τ in order to reach the condition $f(\tau) \sim g_1(\tau)$, and it is described in detail in Appendix C. The advantages of using the statistical weighting, rather than restraining the values of τ *a priori*, is that it mitigates the lack of experimental control over the exact value of τ and the fact that the same experimental data could be rearranged to approximate different target distributions. Once the dataset conditioned to $\mathscr{P}_1$ is generated, the conditional probability of error p_{01} can be once again estimated from the frequency of error $f_{\mathscr{D}}(\mathscr{P}_0|\mathscr{P}_1)$, that will converge to it in the limit of a large number of measurements.

4.1.8 Theoretical Results

For any strategy considered the performance of the QCT is a function of the mean photon number μ, as well as the form of the distributions of the processes considered, $g_0(\tau)$ and $g_1(\tau)$. As already mentioned, Fig. (4.3) provides a visualization of the discrimination problem. The measurement strategy considered at the receiver is photon counting. In panel (**A**) the distributions of the processes to be discriminated are shown, $g_0(\tau)$ in red and $g_1(\tau)$ in blue. In the first row, the two distributions
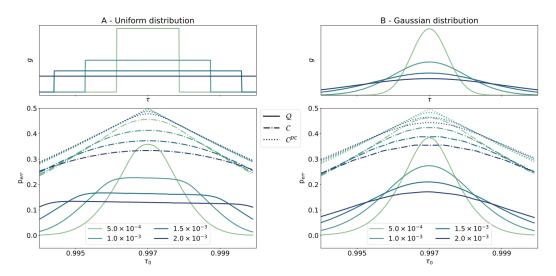
Fig. 4.5 *Theoretical error probabilities for uniform and Gaussian distributions.* In both panels, the mean number of signal photons is fixed to $\mu = 10^5$ and the mean value of the defective process is $\bar{\tau}_1 = 0.997$. The reference process distribution is considered strongly peaked, $g_0 \approx \delta (\tau - \bar{\tau}_0)$. The plots show the error probabilities as functions of the mean value of the reference process $\bar{\tau}_0$, for different variances of the defective process σ^2, reported in the legend, and for different discrimination strategies. The quantum strategy is reported in solid line, the classical strategy with PC is dotted and the optimal classical strategy is dash-dotted. In Panel A the defect distribution, $g_1(\tau)$, is uniform, while in panel B it is Gaussian. The distributions $g_1(\tau)$ for different values of variance are reported over the plots. The image is from Ref. [2].

barely overlap. Nonetheless, when a probe with finite energy is used to perform the discrimination, the intrinsic quantum noise of the state results in a significant amplification of the overlap on the photon number distribution, i.e. the measurement outcome distributions. For the best classical input states this is shown in panel (**C**), where the two possible photon number distributions are plotted, after the interaction with the SUT. The overlap area, highlighted in green, offers a visual representation of the error probability, in particular p_{err} is equal to half the highlighted area. The overlap cannot be reduced using classical states without increasing the signal energy. On the other hand, quantum correlations, at the same signal energy, offer a significant improvement, as already pointed out. Panel (**E**) shows the marginal photon number distribution for the quantum strategy described in Sec(4.1.4), conditioned to a given measurement outcome for the idler system. The overlap between the two distributions is dramatically reduced, leading to much more efficient discrimination. It is indeed in the quantum photon number correlations that the nature of the advantage resides. The effect is to greatly reduce the photon number noise. The second row depicts a situation in which the overlap of the parameter's distributions' is significant. In panel (**B**), the distributions have different variances but the same mean value. The classical discrimination strategy, pictured in panel (**D**), performs very poorly in this case, with the two processes being almost indistinguishable. Panel (**F**) demonstrates, once again, the quantum strategy can be used to recover some distinguishability. Note how the best performances (shown in Fig. (4.3-**E** and 4.3-**F**) are achieved in the limit of large M so that $\mu/M \ll 1$, the regime of validity of the analysis above.

In order to reduce the number of considered parameters, from this point on, we assume the reference process $\mathscr{P}_0$ to be strongly peaked around τ_0, i.e. we consider processes such that $g_0(\tau) \approx \delta(\tau - \bar{\tau}_0)$, where $\delta(\tau)$ is the Dirac delta distribution. For the defective process distribution $g_1(\tau)$ we consider two different noteworthy cases: the Gaussian distribution, $g_1(\tau) = \mathbf{G}_1(\tau)$, and the uniform one, $g_1(\tau) = \mathbf{U}_1(\tau)$. The choice of considering a Gaussian probability distribution for the transmittance τ is justified by the wide range of physical phenomenons it can describe. In the Gaussian case, the processes are fully characterized by their mean value that, following the notation introduced in Sec.(4.1.3), we denote $\bar{\tau}$, and their variance σ^2. On the other hand, the uniform distribution, $\mathbf{U}(\tau)$ is well suited to describe situations in which there is a complete lack of knowledge of the nature of the process, whose range can be limited by the physical constraints of the apparatus. Once more, the processes are characterized by two parameters only: their mean $\bar{\tau}$ and half-width δ. To assure

a fair comparison of the error probabilities for Gaussian and uniform distributions, their parameters are chosen such that the resulting variances are the same. Using the fact that for a uniform distribution $\langle \Delta^2 \tau \rangle_{\mathbf{U}} = \delta^2/3$ the condition to be imposed is $\delta = \sqrt{3}\sigma$.

In Fig. (4.5) we report the performance of each of the strategies described above: the classical one with optimal measurement, $\mathscr{C}$, discussed in Sec(4.1.2), the classical one with photon counting, $\mathscr{C}^{pc}$ discussed in Sec(4.1.3) and the quantum strategy $\mathscr{Q}$ of Sec.(4.1.4). The error probability is plotted as a function of the reference transmittance, τ_0. The number of signal photons is fixed to $\mu = 10^5$, and as the mean value of the transmittance generated by the defective process to $\bar{\tau}_1 = 0.997$. Panel (**A**) reports the case of a uniformly distributed defective process, while panel (**B**) considers Gaussian distributed one. The results are shown for different values of the variance (denoted by different colors). In the range of parameters studied, the quantum strategy (solid lines) outperforms both the classical lower bound (dashed lines) and the strategy employing a classical probe and PC (dotted lines). Note that, when $\tau_0 = \bar{\tau}_1$ both classical and quantum strategies have their maximum error probability. Indeed, in this configuration, the overlap between the initial distributions g_x is maximum and this initial indistinguishability, intrinsic to the problem, cannot be reduced whatever the readout strategy adopted, either classical or quantum. However, the quantum strategy adds less measurement noise with respect to the classical approach, maintaining a certain discrimination capability, while the classical approach leads to the saturation of the error probability, $p_{err} \sim 0.5$.

4.1.9   Experimental results

The presence of a quantum advantage for the QCT protocol is validated experimentally by implementing the quantum strategy of Sec.(4.1.4) with the setup described in Sec(4.1.6). The classical strategy with PC is implemented with the same setup using the fact that the marginal distributions of signal and idler system are very close to Poissonian ones, meaning that they can achieve the best performance. In other words the classical strategy is implemented using only the measurement outcomes on the signal system to make the final decision. In our experimental demonstration the efficiency on signal and idler branch was estimated to be $\eta_S \approx \eta_I = 0.78$, meaning that the advantage obtained in a real scenario is less than the one yielded by theoretical prediction with ideal efficiency presented in Sec(4.1.8).
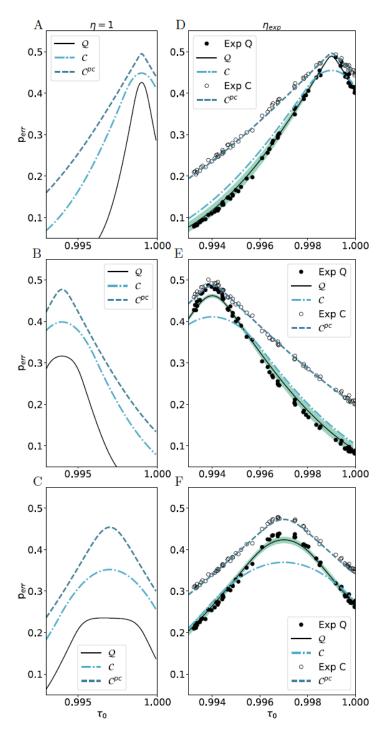
Fig. 4.6 *Experimental results for the QCT protocol.* The experimental and theoretical error probabilities are plotted as a function of the reference parameter τ_0, for different values of the defective process' mean value $\bar{\tau}$ and half-width δ $A : \{\delta = 0.001, \bar{\tau} = 0.999\}$, $B : \{\delta = 0.002, \bar{\tau} = 0.994\}$ and $C : \{\delta = 0.003, \bar{\tau} = 0.997\}$. On the left-hand are shown the theoretical curves for ideal detection efficiency ($\eta_S = \eta_I = 1$). In the right column are reported the experimental error probabilities obtained using the quantum PC strategy, black dots, and the classical PC one, circles, along with the theoretical curves in both cases and the theoretical absolute classical bound, $\mathscr{C}$. For the quantum theoretical error probability are reported green shaded areas representing the confidence interval at one standard deviation. The experimental efficiencies of the channels are $\eta_S \simeq \eta_I \approx 0.76$, while the mean number of signal photons is $\mu \approx 10^5$. Image from Ref. [2].

The experimental results are shown in Fig. (4.6). The three rows in Fig. (4.6) report the error probability for three different setting of the defective distributions distributed uniformly: panels A, D with $\{\delta = 0.001, \bar{\tau} = 0.999\}$, panels B,E with $\{\delta = 0.002, \bar{\tau} = 0.994\}$, panels C,F with $\{\delta = 0.003, \bar{\tau} = 0.997\}$. The reference process is assumed to be well approximated by a delta function centered in $\bar{\tau}_0 = \tau_0$, as it was already the case in the previous section. The experimental points are reported for the classical and quantum strategy with photon counting together with the theoretical prediction for the quantum strategy $\mathscr{Q}$ and the classical performance with PC,$\mathscr{C}^{PC}$, as well as the classical optimal bound $\mathscr{C}$. The results are displayed in Fig. 4.6 (D-E-F). The theoretical results with ideal efficiency, $\eta_S = \eta_I = 1$, are reported as a comparison in Fig. 4.6 (A-B-C). The theoretical curves drawn using the parameters estimated for each acquisition, the mean numbers of photons, the efficiencies η_S and η_I of the signal and idler channels (both set to 1 in panels (A-B-C)), as well as the electronic noise ν of the camera. Data at different values of τ were collected using a sample with a varying spatial transmittance. From this experimental dataset different defect distributions can be realized following the procedure outlined in Sec(4.1.7).

Most points fall within the confidence interval for all three regimes considered. The quantum strategy always brings an advantage with respect to the classical one based on PC, even in the case of degraded detection efficiency of the real experiment. However, in some regions, namely when τ_0 start to fall within the range of the defect distribution, the optimal classical bound $\mathscr{C}$ on the error probability can become smaller than the quantum strategy error probability. This point can be put into perspective considering Fig. 4.6 (A-B-C), which are constructed with the same experimental parameters but unit detection efficiency. In this case, the quantum strategy overcomes any classical one without exception in the region shown. We note that, as expected, while the classical error probabilities are only slightly modified by the change in efficiency η, the quantum one improves significantly in the case of $\eta = 1$. This effect stems once again from the fact that, spurious losses η reduce the photon number correlations between signal and idler channels, an effect relevant only to the quantum strategy.
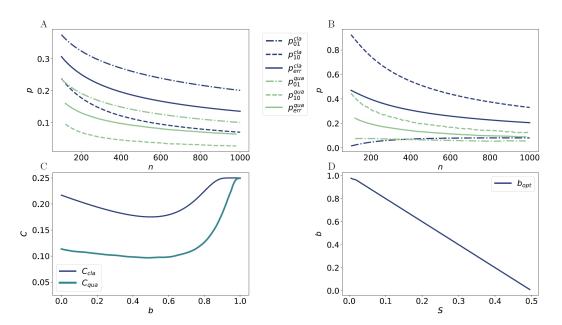
Fig. 4.7 *Cost analysis in QCT*. The dependence on the signal photon number n_S of the total error probability (p_{err}), false positive (p_{01}) and negative (p_{10}) are shown in Panel **A** in case of post processing minimizing p_{err}. The reference process distribution is considered strongly peaked around the value $\tau_0 = 0.8$. The defective process distribution is chosen uniform with mean $\bar{\tau} = 0.9$ and half-width $\delta = 0.09$. In panel **B**, the parameter are the same, but a biased maximum likelihood post-processing is used with bias coefficient $b = 0.6$. In panel **C**, the number of photons is fixed to $\mu = 500$ and it is analysed the dependence if the cost C on b. All the other parameters are the same as the previous panels. Panel **D** reports the optimum value of b as a function of S. The plots are taken from Ref. [2].

4.1.10   Cost function analysis

As anticipated in Sec.(4.1.1) the average probability of error is not always the best *cost function*. In some situations, it can be required to impose a constraint in either one of the conditional probabilities p_{01} or p_{10}, instead of their sum. These two types of error represent, in fact, different scenarios and the minimization of one can be deemed more important than the other and it is then natural that the cost function has to be weighted accordingly. The notion of *cost* quantifies the fact that in the mislabelling of a process, the false-positives and false-negatives may not be equivalent for the operator making the decision. One type of error may be more "costly" than the other. To extend our analysis to this more general scenario we introduce the coefficient $0 < S < 1$, and define the total cost C of the conformance test as

$$C = Sp_{10} + (1 - S)p_{01}. \tag{4.42}$$

The cost C replaces the total probability of error as the figure of merit to be minimized for the evaluation of the QCT. The minimization of the cost will also minimize the total probability of error, p_{err}, if $S = 1/2$. On the other hand if $S \neq 1/2$, minimizing C will not, in general, yield the minimum total error probability.

In the following, it will be shown how the readout strategies presented in the previous sections, consisting of PC and a maximum likelihood decision, both for classical and quantum probes, can be modified to minimize a generic cost C. Following the formalism introduced in the previous sections, the general term for the probabilities p_{ij} is:

$$
\begin{aligned}
p_{01} &= \sum_{\mathbf{n}} p(\mathbf{n}|\mathscr{P}_1)\Theta[p(\mathbf{n}|\mathscr{P}_0) - p(\mathbf{n}|\mathscr{P}_1)] \\
p_{10} &= \sum_{\mathbf{n}} p(\mathbf{n}|\mathscr{P}_0)\Theta[p(\mathbf{n}|\mathscr{P}_1) - p(\mathbf{n}|\mathscr{P}_0)]
\end{aligned}
\tag{4.43}
$$

where Θ is the Heaviside step function, equal to 1 if its argument is positive and 0 otherwise.

In Fig. (4.7-**A**) the false-positive and false-negative probabilities are plotted as functions of the mean number of signal photons for both the classical and quantum PC strategies, in the case where p_{err} is the optimized quantity. In both cases, classical

(red lines) and quantum (blue lines), the false-negative probability tends to be smaller than the false-positive one, meaning that the procedure is more likely to select the reference process rather than the defective one. This unbalance is determined by the fact that the analysis is performed on a reference process having a distribution strongly peaked around τ_0, while, the defective process is distributed uniformly. In terms of photon counts probability, this results in the fact that, when there is a significant overlap, the most peaked process, $\mathscr{P}_0$, is chosen, while the defective one is selected in a range that is larger, but also less likely to be measured. This results in the procedure being biased towards $\mathscr{P}_0$. This can also be seen in Fig. (4.8) where are reported the experimental conditional probabilities of error, along with the theoretical curves as a function of the reference process transmittance τ_0. The experimental points for the false positive and negative of the classical and quantum strategies with a PC receiver, can be evaluated using the data analysis procedure, described in Sec.(4.1.7) and Appendix C, from the same dataset used to evaluate the total probability of error. The preference toward $\mathscr{P}_0$ is more evident when there is a strong overlap between reference and defective process, i.e. as $\bar{\tau}_0$ approaches $\bar{\tau}_1$ (set to $\bar{\tau}_1 = 0.999$ in Figure). In this region in fact very little information is available for the discrimination an the algorithm selects according to the discussed preference, resulting in the peak in distance between p_{10} and p_{01}.

The post-processing used up until now, i.e the decision rule selecting y when $p\left(\boldsymbol{n}|\mathscr{P}_y\right) \geq p\left(\boldsymbol{n}|\mathscr{P}_{1-y}\right)$, minimizes the probability of error by construction. To minimize the cost C, a different post processing is needed. This new post-processing can be found by modifying the decision rule used up until now, to make it more likely that one specific process is selected according to some parameters. A new decision rule can be defined such that y is selected when:

$$B^{(y)}p\left(\boldsymbol{n}|\mathscr{P}_y\right) \geq B^{(1-y)}p\left(\boldsymbol{n}|\mathscr{P}_{1-y}\right) \tag{4.44}$$

where we have introduced:

$$B^{(0)} = \frac{1-b}{2}$$
$$B^{(1)} = \frac{1+b}{2} \tag{4.45}$$

with $b \in [-1,1]$. Hereinafter, we will refer to this new post-processing as "biased maximum likelihood". If a positive b is selected, the defective process will be chosen
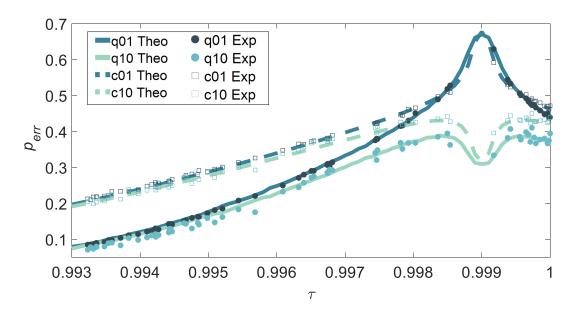
Fig. 4.8 *Theoretical and experimental conditional probabilities of error for the quantum and classical strategies with PC receiver.* The experimental conditional probabilities of error, for unbiased post-processing, are reported along with the theoretical ones for a uniform defect distribution with $\bar{\tau} = 0.999$ and $\delta = 0.001$. The mean number of signal photons is $\mu \approx 10^5$ and the channels efficiencies are $\eta_S \simeq \eta_I \approx 0.76$. The plot is reported from Ref. [2]

more often, while a negative b makes it more likely that reference one is selected. In other words, the value of b can be varied to shift p_{01} and p_{10}, reducing the cost function C at the cost of the increase of the total error probability. This can be seen in Fig. (4.7.**B**) where p_{01} and p_{10} have been moved, with respect to the balanced case of Panel **A**, and the total error probability slightly increased, for both the classical and quantum strategies. The quantum advantage found when minimizing the error probability is well preserved with the biased maximum likelihood decision. The optimization of C is performed by varying the coefficient b. As an example, consider the case in which $S = 1/4$, meaning that each false positive is considered three times as costly as a false negative. In this situation, a positive value for b is required to reduce the number of times the reference process $\mathscr{P}_0$ is selected overall, thus reducing the false positive probability p_{01}. In Fig. (4.7-**C**) it is reported the dependence of C on $b^{(0)}$. Both in the quantum and classical cases, as expected, there is a single optimum value for b, that minimizes the cost. Fig. (4.7-**D**) shows the dependence of the optimum value of b on S, with respect to the cost C.

4.1.11 Discussion

In the first section of this chapter, we discussed the QCT, a protocol addressing the conformance test task, exploiting quantum probe states to show an advantage over classical measurements. In particular we considered test where the systems under test, representing possible outputs of the unknown process to be monitored, are be modelled by pure loss channels.

In Sec.(4.1.2) we found a lower bound on the error probability that can be achieved in the discrimination task, using any classical source paired with an optimal measurement. In Sec.(4.1.4) we discussed a strategy based on quantum states, namely TMSV states, in conjunction with a simple receiver consisting of a PC measurement and a maximum likelihood decision. It was shown how this quantum strategy is able to surpass the performances of any classical strategy. The quantum enhancement is due to the high degree of correlation (entanglement) of TMSV states, whose nature is fundamentally quantum. To give an additional experimental benchmark, we also analysed, in Sec.(4.1.3) the case of a classical source paired with photon-counting measurements at the receiver.

The quantum advantage was shown to persist for a wide range of parameters. The advantage exists even in presence of losses due to possible experimental imperfections, and the results were validated with an experimental realization of the protocol. An experimental advantage was showed in a realistic scenario where experimental losses amounted to more than 20%, highlighting the robustness of the protocol.

The states used to achieve these results are relatively easy to produce and the receiver design is of simple implementation as well, making the results obtained particularly significant in view of possible practical applications of the protocol with present technology. It is remarkable that an advantage was found notwithstanding the fact that the bound on the performance of classical states, $\mathscr{C}$, is not tight, meaning that the actual quantum advantage could effectively be higher. Moreover in the next section we will show how the proposed quantum strategy, relying on an independent measurement over individual systems is able to surpass the performance of classical states even when joint measurements over a collection of systems are allowed for the latter. The QCT protocol discussed in this chapter could find applications in the foreseeable future in significant problems concerning the monitoring of production process of any object probed with quantum states. As an example, the results showed

on loss channel QCT can be used to enhance the accuracy in the identification of issues in concentration and composition of chemicals production.

4.2   Quantum readout of an imperfect memory

The formalism introduced above with the QCT protocol is suited to analyse a variety of problems beside the production process monitoring proposed in the original formulation. In the following we will apply this formalism to the readout of an *imperfect digital memory*. This problem was first studied in Ref. [3], an original contribution of this Thesis work. An imperfect memory is defined as a memory composed by cells encoding binary information in values of transmittance that are not known *exactly* but are extracted from two well characterized probability distributions. In other words rather then encoding the information in two real values of transmittance τ_0 and τ_1, as it was the case for the ideal memory cells in the quantum reading protocol discussed in Chapter 3, an imperfect memory encodes the information in two classical random variables T_0 and T_1. This description is more suitable to describe real scenarios. In fact the encoding of classical data in a physical support can be done up to some level of accuracy due to errors and the imperfection of the writing process. Moreover, some degradation of the storage data can happen over time because of physical or chemical instability of the support. Any realistic read-out strategy should take into account this uncertainty and aim to minimize its effect.

In the following we will give a formal statement of the readout problem with imperfect memories and compare the performance that can be obtained by classical and quantum strategies. The analysis will be extended beyond the single cell readout by considering a more general classical benchmark on a multicell memory where collective measurements of the probed cells are allowed. It will be shown that local, cell-by cell, quantum readout remains anyway better in recovering the stored information even when compared to a global classical benchmark. A multicell memory can be seen as a large block of cells for which the information is stored, according to some classical encoding, in classical codewords expressed by a set of cells configurations. In the limit of very large memories (infinite number of cells) the maximum amount of information retrieved can be found with a constrained (at fixed energy) optimization of the Holevo bound [56, 58, 57], discussed in Sec.(2.3) of

Chapter (2). A solution of this optimization for classical input states and imperfect encoding will be shown.

4.2.1 Memory cell model and local readout strategies

Consider an optical memory composed of D cells, each storing one bit of information in two possible values of transmittance τ_0 and τ_1. As always the readout of each cell is carried out using a transmitter emitting a bipartite optical probe in a state ρ, consisting of M optical modes in the signal system, interacting with the cell while L modes of the idler system are sent directly to a receiver, where a joint POVM is performed on signal and idler modes. A decision on the value, $y = 0, 1$, of the recovered bit is taken after classical post processing of the measurement result. An ideal memory, like the one described in the Quantum reading protocol (Chapter 3), stores the information in one of two real transmittance parameters, τ_0 or τ_1 of each cell, modeled as a pure loss quantum channel $\mathscr{E}_\tau$. Similarly a single cell of an *imperfect optical memory* stores one bit of information in one of two possible random values of transmittance, T_i, with $i = 0, 1$, each having known probability distribution $g_i(\tau_i)$, and prepared with the same prior probability $p_0 = p_1 = 1/2$. In other words the problem of information retrieval from an imperfect memory cell is a quantum channel discrimination problem, between convex combinations of pure loss channels $\mathscr{E}_\tau$. Hence the problem is formally the same as the QCT and the results found in the previous section can be used. While in the QCT protocol the most used figure of merit was either the total probability of error or some weighted one, in the context of memory reading, the information recovered by the procedure is a more fitting figure of merit. Remember that given the total probability of error p_{err} the information recovered I is:

$$I = 1 - H(p_{err}) \tag{4.46}$$

where $H(p) = -p \log_2 p - (1-p) \log_2(1-p)$ is the binary Shannon entropy. In the following, we will compare the performance of the *local* quantum read-out strategy based on TMSV input states, where local means that each cell is probed and measured separately, with respect to three different classical benchmarks: the two local ones discussed in the QCT protocol, namely the absolute classical bound and the classical bound with PC measurements and finally a newly introduced global optimal classical strategy. In the last one, an array of D cells is probed by a tensor

product state $\rho^{\otimes D}$, and the receiver is allowed to perform a collective measurement across the memory.

To this end let us introduce the information recovered by each of these strategies:

- *Optimal classical strategy.* The optimal classical strategy refers the performance of classical states (defined in Eq. (3.2)) paired with an optimal measurement. For this scenario the quantity $\mathscr{C}$ in Eq. (4.15) establishes a lower bound on the probability of error. For more details see Sec(4.1.2). We point out that $\mathscr{C}$ may not be a tight bound, so it may be not possible to find a strategy that saturates it. From $\mathscr{C}$ we can define an absolute upper bound on the information that can be recovered by classical states, with local readout, as:

$$I_{\mathscr{C}}^{abs} := 1 - H(\mathscr{C}) \tag{4.47}$$

- *PC classical strategy.* The second strategy consist in classical states paired with the PC receiver described in Sec.(3.2.1). This strategy is described in Sec.(4.1.3) and the best performance achievable is reached by input states having Poisson photon number distribution and expressed by the quantity by $\mathscr{C}^{PC}$ in Eq. (4.37). Consequently we define the information recovered:

$$I_{\mathscr{C}}^{PC} := 1 - H(\mathscr{C}^{PC}) \tag{4.48}$$

- *Quantum strategy.* The final local strategy considered consists of a collection of M copies of TMSV states paired with the PC receiver. The performance in this scenario, discussed in Sec (4.1.4), is given by the quantity $\mathscr{Q}$ in Eq. (4.41). The information recovered in this case is:

$$I_{\mathscr{Q}}^{PC} := 1 - H(\mathscr{Q}) \tag{4.49}$$

In the following section we will introduce a new bound on the performance that can be achieved by classical states when encoding and collective measurement over a very large array of cells are allowed. We call this bound the *global classical limit*.

4.2.2 Global classical limit

A real memory is constituted by a very large array of D cells. In informational terms, every single cell can be seen as encoding the binary random variable X in an ensemble $\{p_i, \rho_i\}$, $i = 0, 1$. A direct encoding (one bit per cell) may not be the more efficient storing strategy if the array of length D can be probed in parallel and a joint measurement on all the outputs is allowed. In general it is most efficient to use *codewords* to store information onto the array of cells. The strategies that we will consider for the retrieval are such that each cell is probed by a copy of a state ρ, meaning that the total probing state is of the form $\rho^{\otimes D}$. The information is recovered by a joint global POVM measurement at the output. The multicell storage/retrieval of information can be seen as a transfer of information through the memory, characterized in terms of quantum channels. Consequently the maximum rate of information, I^D, that can reliably stored and retrieved per single channel use (cell of the memory) is upper bounded [62] by the Holevo quantity χ, introduced in Sec.(2.3) :

$$I^D \leq \chi(\rho) := S(\rho) - \sum_i p_i S(\rho_i) \tag{4.50}$$

where $\rho = \sum_i p_i \rho_i$. In this case $i = 0, 1$, $\rho_i = \mathbb{E}_{g_i}[(\mathscr{E}_{\tau_i} \otimes mathds1)\rho]$ are the states after the interaction with the memory cell and p_i is the probability for the cell to be prepared with either one of the values of T_i, that in our case is $p_i = 1/2$. In the case of $D \to \infty$, that usually is a good approximation since digital memories are usually very large, the Holevo-Schumacher-Westmoreland (HSW) theorem, discussed in Sec.(2.3), assures that a POVM exist such that the bound is saturated. The optimization of the quantity in Eq. (4.50) over any state ρ at fixed mean signal energy μ provides an ultimate limit on the rate of information per bit that can be stored and retrieved accurately. However this optimization is not easy to perform on the full space of states. The maximization becomes significantly easier if it is restricted to the class of classical input states. Let us define the maximum global classical information, χ_{cla}, as the maximization of χ over all classical states:

$$\chi_{cla} := \max_{\rho_{cla}} \chi(\rho) \tag{4.51}$$

To find χ^{cla} we first find the Holevo quantity for a single mode coherent state and, following the procedure of a similar proof in Ref. [62], given in the context of an ideal memory, then we show that the quantity found is equal to the classical capacity.

Consider an input single mode coherent state $|\alpha\rangle\langle\alpha|$ for the readout procedure. The constraint on the number of photons (see Eq(4.4)) gives the condition $|\alpha|^2 = \mu$. Let us denote the Holevo quantity for this state as χ_α. To compute χ_α we use the fact that pure loss channels map coherent states into coherent states, $\mathscr{E}_\tau(|\alpha\rangle\langle\alpha|) = |\sqrt{\tau}\alpha\rangle\langle\sqrt{\tau}\alpha|$, giving for the output states after the interaction with the memory cell:

$$\rho_0^\alpha = \mathbb{E}_{g_0}[|\sqrt{\tau}\alpha\rangle\langle\sqrt{\tau}\alpha|] = \int d\tau g_0(\tau)|\sqrt{\tau}\alpha\rangle\langle\sqrt{\tau}\alpha|$$

$$\rho_1^\alpha = \mathbb{E}_{g_1}[|\sqrt{\tau}\alpha\rangle\langle\sqrt{\tau}\alpha|] = \int d\tau g_1(\tau)|\sqrt{\tau}\alpha\rangle\langle\sqrt{\tau}\alpha| \qquad (4.52)$$

Since there are no idler modes we omitted them from the notation. We have then:

$$\chi_\alpha = S(p_0\rho_0^\alpha + p_1\rho_1^\alpha) - \sum_i p_i S(\rho_i^\alpha) \qquad (4.53)$$

The computation of the entropy is complicated by the fact that the distributions g_i are in general continuous ones. To overcome this problem a uniform discretization of the distribution can be performed to some appropriate dimension k. In this approximation the integrals over τ in Eq. (4.52) are substituted by finite sums, $\int d\tau \rightarrow \sum_i^k$. With this substitution it is easy to see how each entropy term in Eq. (4.53) can be rewritten as some convex combination of a finite subset of maximum dimension $2k$ of coherent states. The calculation of χ_α than can be reduced to the calculations of terms in the form:

$$S\left(\sum_{i=1}^{K} q_i|\sqrt{\tau_i}\alpha\rangle\langle\sqrt{\tau_i}\alpha|\right) \qquad (4.54)$$

where q_i are suitable probability coefficients depending on the initial distributions g_i and the term considered. For the first term on the right hand side of Eq. (4.53) the coefficients q_i will depend also on the probabilities p_i, that in this discretization represent the initial distributions $g_x(\tau)$, and the sum will in general be on $K = 2k$ terms. For the two entropies in the summation, the sum will go over $K = k$ terms.

The set $\{|\sqrt{\tau_i}\alpha\rangle\}$ is a non-orthogonal basis of a K-dimensional Hilbert space. The entropy can be computed using the fact that for any state $\rho = \sum_{i=1}^{K} q_i|\sqrt{\tau_i}\alpha\rangle\langle\sqrt{\tau_i}\alpha|$, we have:

$$S(\rho) = S(QG) \qquad (4.55)$$

where $Q = \text{Diag}[q_i]$ and $G(i,j) = G(j,i) = \langle\sqrt{\tau_i}\alpha|\sqrt{\tau_j}\alpha\rangle$ are the elements of the Gram matrix G. We can write, in fact [81]:

$$S(\rho) = -\frac{\partial}{\partial n}\text{Tr}(\rho^n)\Big|_{n=1} \tag{4.56}$$

And

$$
\begin{aligned}
\text{Tr}(\rho^n) &= \sum_{i_1,\dots,i_n=1}^{K} q_{i_1}\cdots q_{i_n}\langle\sqrt{\tau_{i_1}}\alpha|\sqrt{\tau_{i_1}}\alpha\rangle\cdots\langle\sqrt{\tau_{i_n}}\alpha|\sqrt{\tau_{i_1}}\alpha\rangle = \\
&= \sum_{i_2,\dots,i_n=1}^{K} q_{i_2}\cdots q_{i_n}\langle\sqrt{\tau_{i_2}}\alpha|\sqrt{\tau_{i_3}}\alpha\rangle\cdots\langle\sqrt{\tau_{i_{n-1}}}\alpha|\sqrt{\tau_{i_n}}\alpha\rangle GQG(i_n,i_2) = \\
&= \sum_{i_n=1}^{K} q_{i_n}G(QG)^{n-1}(i_n,i_n) = \text{Tr}[(QG)^n]
\end{aligned}
\tag{4.57}
$$

And combining the results of Eq. (4.56) and Eq. (4.57) yields the equality reported in Eq. (4.55).

Eq. (4.55) allows to skip the orthogonalization of the subspace, which would represent a computational heavy task for large values of k, so it is useful even when a closed form may not be available for a given probability distribution and for arbitrary k. A numerical evaluation of χ_α can be performed using Eq. (4.55-4.56). The Holevo quantity χ_α for the continuous set is recovered as the limit for $k \to \infty$ of $\chi_\alpha(k)$. In the following section we will show how χ_α coincides with the classical capacity χ_{cla}.

The fact that the classical capacity can be computed using a single-mode coherent transmitter can be proved using the argument of Ref. [62] that we report in the following.

Consider the class $\mathscr{P}$ of pure coherent transmitters. The class formed by convex superpositions of elements of $\mathscr{P}$ constitutes the class of classical states. It can be proved [62], using the convexity on ρ of the Holevo information χ, that if the capacity of the class $\mathscr{P}$, $\chi_\mathscr{P}$, is concave in the number of photons μ, it must be larger or equal to the capacity of the whole class of classical states:

$$\chi_\mathscr{P} \geq \chi_{cla} \tag{4.58}$$
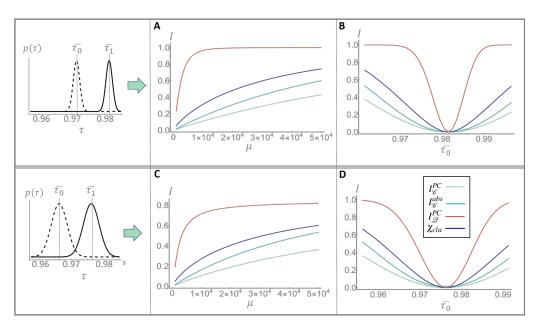
Fig. 4.9 *Comparison of information recovery.* The information recovered by the classical strategies and the quantum one is compared for different parameters of the transmittance distributions $g_0(\tau_0)$ and $g_1(\tau_1)$, both assumed gaussian. In the first row, the mean value of g_1 is fixed to $\bar{\tau}_1 = 0.982$ and the standard deviations of the distributions to $\sigma_1 = \sigma_0 = 0.001$. In panel **A**, the mean value of g_0 is fixed to $\bar{\tau}_0 = 0.972$ and the information is shown as a function of the mean number of signal photons μ. In panel **B** the mean photon number is fixed to $\mu = 10^4$ and the information is reported as a function of $\bar{\tau}_0$. In the second row, the parameters of the transmittance distributions are changed in order to increase their overlap, with $\bar{\tau}_1 = 0.976$, $\bar{\tau}_0 = 0.966$ and $\sigma_1 = \sigma_0 = 0.0025$. In panel **D** $\mu = 10^4$. The plots are reported from Ref. [3].

$\chi_{\mathscr{P}}$ is in fact concave in this case and the concavity can be checked numerically, so Eq. (4.58) holds. On the other hand given that $\mathscr{P}$ is a subclass of the classical states, the capacity of $\mathscr{P}$ cannot be larger than the one of classical states, so that Eq. (4.58) must be an equality. Then the *capacity of classical states is saturated by pure coherent states*.

To complete the proof, one must simply show that a single coherent mode saturates the capacity of $\mathscr{P}$. This can be done [62] using the fact that, acting with unitary transformations on the output states one of any pure classical input, one can rearrange the signal photons in a single mode obtaining the same result as a single mode input [62]. Given that the Holevo quantity is calculated on the output states and that the von-Neumann entropy does not change under unitary transformation this means that single mode coherent states can achieve the best performance in $\mathscr{P}$, i.e. $\chi_{\alpha} = \chi_{\mathscr{P}} = \chi_{cla}$, that concludes the proof.

4.2.3   Results

To analyse the readout performance of the readout strategies outlined, in the following the distributions $g_i(\tau_i)$ of the random variables T_i will be assumed to be Gaussian and with the mean value denoted as $\bar{\tau}_i$ and the standard deviation as σ_i. Fig. (4.9) reports the information recovered per cell, for two possible configurations of the transmittance's distributions. In the first row it is shown a situation in which the overlap between g_0 and g_1 is negligible. In this case, although the transmittance values are uncertain, in principle the value of the bit is codified in an almost unambiguous way, and an ideal readout would be able to almost always recover the bit correctly. However, quantum fluctuations introduce further noise, which reduces the final distinguishability, and at fixed signal energy the information that can be recovered will be limited. Panel **A** shows the information recovered as a function of the mean number of signal photons μ. The information recovered increases as the signal energy is increased, up until it saturates at the maximum amount of information for a single binary cell, i.e. one bit. In the range shown in the figure the saturation is only visible for the quantum strategy, $I_{\mathscr{Q}}^{PC}$, reported in red, that reaches it earlier than any of the classical ones, even earlier than the global capacity bound χ. This shows how with the use of quantum resources a reliable recovery of the information can be reached with significantly less energy than otherwise needed. In Panel **B** the mean number of signal photons is fixed, along with all other parameters,
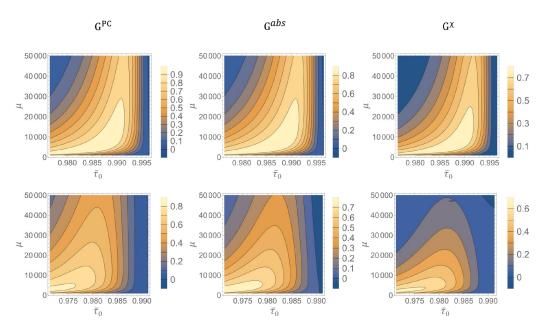
Fig. 4.10 *Quantum Gain in the readout.* The quantum gain G is reported as a function of the mean number of signal photons μ and the mean value of one of the encoding transmittance distributions, $\bar{\tau}_0$. In the upper row, the parameters are fixed to $\sigma_0 = \sigma_1 = 0.001$ and $\bar{\tau}_1 = 0.997$. Starting from the left, G^{PC} is the quantum gain over to the performance of classical states an photon counting receiver, G^{abs} over the absolute bound of classical states considering local readout, and G^{χ} over the bound on the classical performance with global measurements. The lower row reports the same figures of merit but the parameters are changed to $\sigma_0 = \sigma_1 = 0.0025$ and $\bar{\tau}_1 = 0.991$. Plots from Ref. [3].

and it is shown the dependence of the information recovered on the value $\bar{\tau}_0$. The recovered information is higher when $\bar{\tau}_0$ is far from the fixed mean value $\bar{\tau}_1$, as it can be expected, and starts decreasing when the two values get too close. However, the strong degree of correlation of the source in the quantum case, reflects in a much narrower low-informative region, the deep in Fig. 4.9**B**, due to the reduction in quantum fluctuations. In the second row of Fig. 4.9 it is reported the case of a more significant overlapping of the initial distributions, due to their increased standard deviations, showed in the bottom-left box. The dependence on the mean photon number μ is showed in panel **C** while in panel **D** it is showed the dependence on the transmittance $\bar{\tau}_0$. The information saturates at a value smaller than 1 bit (specifically 0.8) due to the initial overlapping of the distributions. In view of the initial ambiguous encoding, in fact, not even a perfect measurement could recover full information. Panel **D** displays an effect similar to panel **B**, with a drop in information recovered as $\bar{\tau}_0$ approaches the fixed value of $\bar{\tau}_1$. The low-informative region is widened w.r.t the case of non-overlapping distributions. In this scenario, where there is a greater part of indistinguishability not due to fluctuations, the quantum strategy still offers a significant improvement over the classical ones.

Fig. 4.10 show contour plots of the quantum gain that as in Chapter (2) is defined as the difference:

$$G = I_{\mathcal{Q}}^{PC} - I_{cla} \qquad (4.59)$$

where I_{cla}, can represent each one of the three classical bounds defined in Eq.s (4.47), (4.48) and (4.50). G^{PC}, G^{abs} and G^{χ} is the quantum gain w.r.t $I_{\mathcal{C}}^{PC}$, $I_{\mathcal{C}}^{abs}$ and χ_{cla} respectively. The plots in Fig. 4.10 report the gain as a function of the mean number of photons μ and of the transmittance $\bar{\tau}_0$. The other parameters are fixed. In the first row the standard deviation of both distributions is fixed to $\sigma_0 = \sigma_1 = 0.001$, a value small enough to reduce the initial overlapping. The quantum gain in all three cases is significant in most of the region of analysis. The quantum strategy, realized with PC measurement, performs very well against the same measurement strategy realized with classical probes, with the gain G^{PC} reaching values above 0.9 bits. Such a high gain means that in this region using a quantum probe would allow to recover almost all the information in region whereas the same detection strategy with classical states would almost completely fail. A similar result is found for the gain over the optimal local classical bound, G^{abs}, although the values of the gain are slightly lower, with a maximum of around 0.8 bits. It is remarkable that the gain G^{χ} over the classical global limit, representing the bound on the information recovered per cell after a
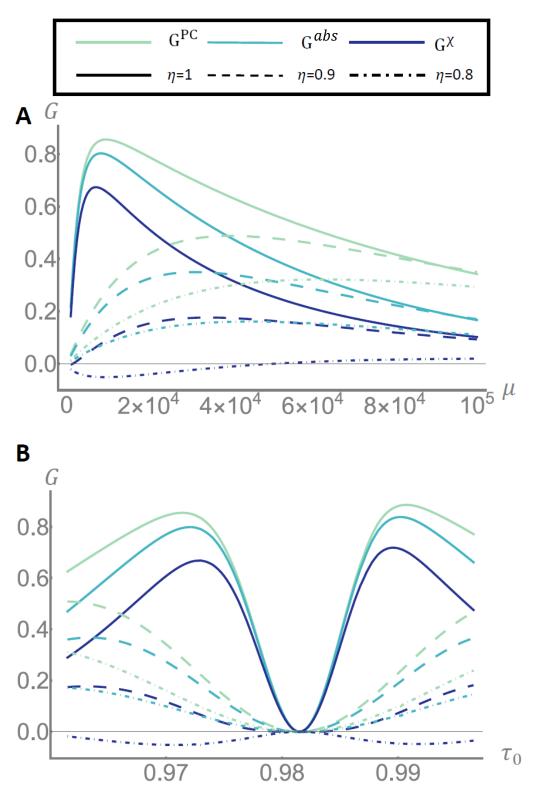
Fig. 4.11 *Efficiency analysis* The quantum gain is showed for different values of overall efficiency η. Panel **A** shows the dependency of the gain on the main number of signal photons μ. The case of ideal efficiency $\eta = 1$ is reported in solid line, while the cases $\eta = 0.9$ an $\eta = 0.8$ are reported in dashed and dot dashed lines respectively. The parameters of the distributions g_0 and g_1 are $\bar{\tau}_0 = 0.972$, $\bar{\tau}_1 = 0.982$ and $\sigma_0 = \sigma_1 = 0.001$. In panel **B** the mean number of signal photons is fixed to $\mu = 10^4$ and the dependency on the mean transmittance $\bar{\tau}_0$ is showed. Plots from Ref. [3]

global measurement, is significantly higher than zero in a wide region and peaks at values higher than 0.7 bits per cell. These results shows how, even if not trivial classical encoding over large memories is used, quantum correlation still offer an improvement. For small mean number of photons, where quantum fluctuations are more relevant, the range of transmittance showing a significant advantage is wider as it can be seen in all three panels. In the second row the effect of increasing the standard deviation to $\sigma_0 = \sigma_1 = 0.0025$ is shown. This increase leads to a large overlapping between the initial distributions in the region explored. The effect is, in general, a reduction of the gain due to the initial ambiguity of the encoding that reduces the value of accessible information a recovery to less than one bit and consequently the space for quantum advantage.

As extensively discussed in the previous chapter for the QR protocol, and in the previous section for the QCT one, optical losses play a central role in experimental realizations. We remind that losses are accounted by the term $1 - \eta$, with η being the efficiency $0 \leq \eta \leq 1$ and that while for the classical case the losses are equivalent to an effective reduction of the probing energy, in the quantum case losses also have a hindering effect on the correlations thus reducing the advantage that can be reached. The effect of losses on the gain is reported in Fig. 4.11. The gain for the distribution case of non-overlapping initial distribution for different values of the efficiency η. In panel **A** it is displayed the gain as a function of the mean number of signal photons μ. An advantage over the classical local bounds is preserved up to $\eta = 0.8$ (20% of losses), although the magnitude of the gain is reduced. The advantage w.r.t. the classical capacity bound is preserved for higher values of efficiency, $\eta = 0.9$, but it is lost at $\eta = 0.9$. We point out that the figure reported refers to gain per cell of information, so even a small fraction of gained information could result in a sizeable improvement over very large memories. Panel **B** reports the gain as a function of the mean transmittance $\bar{\tau}_0$. Along with the overall reduction of the gain, due to losses it can be observed a widening of the low-informative region in the $\bar{\tau}_0$ range. Finally, in Fig. 4.12 we compare the performance of the retrieval strategy proposed in this section with the one using only the two central values of transmittance $\bar{\tau}_0$ and $\bar{\tau}_1$, ignoring the distributions characterizing the memory. The information $I_{\mathscr{C}}^{PC}$ recovered using classical states and photon counting is compared with the information $I_{\mathscr{C}}^{PC,MV}$ recovered with the same resources but using only the mean values of the distributions. As expected, the characterization of the memory,
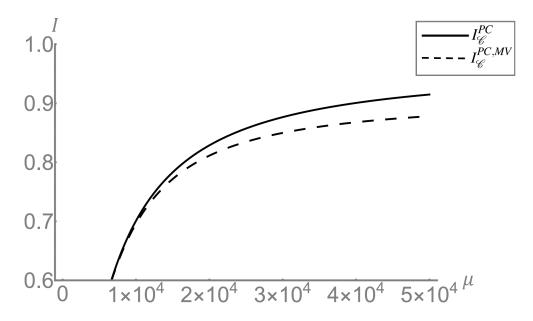
Fig. 4.12 *Comparison of post-processing.* In the plot is shown the information recovered with two maximum likelihood post processing, one $I_{\mathscr{C}}^{PC}$, using full information on the distributions of T_0, and T_1, as discussed in the main text, and the other $I_{\mathscr{C}}^{PC,MV}$ using only the mean values of the distributions $\bar{\tau}_0$ and $\bar{\tau}_1$. The parameters are fixed to $\bar{\tau}_0 = 0.925$, $\sigma_0 = 0.005$, $\bar{\tau}_1 = 0.965$ and $\sigma_1 = 0.01$. Figure form Ref. [3].

i.e. the knowledge of the distribution of the physical parameter used for the encoding, and its use in the decision algorithm bring an advantage in the readout.

4.2.4   Discussion

In this second section of the chapter, the effect of an imperfect characterization of a memory cell on its readout performance was analysed. It was considered the case of an optical memory storing a bit of information in two values of transmittance T_0 and T_1 that are not known with arbitrary precision. In other words T_0 and T_1 are classical random variables for which we assumed Gaussian distributions. For the readout task a quantum readout strategy, consisting of TMSV states transmitter and a PC receiver was compared with three classical informational limits, namely the classical optimal bound for a single-cell readout, the best classical performance achievable with a PC receiver, and the classical capacity limit allowing collective measurements over a memory composed of a large number of cells. It was shown how the local quantum strategy reaches a notable advantage in terms of bit recovered per cell over

all of them. Remarkably this is true even when considering the global classical strategy, as anticipated at the end of the Sec.(4.1). To study the performance in view of possible applications, we considered the effect of optical losses in this scenario as well, showing the robustness of the protocol to this effect, with the advantage preserved for losses around 20%, the value achieved in experiments in the similar realizations of the other protocols discussed. We also demonstrated how taking properly into account the parameter distributions, when available, in the decision algorithm allows, in general, to optimize the readout performance. Albeit the work discussed in this section is mainly focused on a model of digital memory, the results can be applied to different scenarios involving convex superposition of loss channels.

Chapter 5

Quantum enhanced pattern recognition

In this chapter we will discuss an application to the enhanced quantum sensing techniques, discussed in the previous chapters, to a complex task of practical interest, namely pattern recognition. We will start by giving a formal definition of the patter recognition problem, highlighting its relation with the single-pixel readout, and then showing how the enhancement offered by a quantum readout translates to a dramatic improvement in the pattern classification performance, both theoretically and experimentally. We will discuss two different experimental setups. In the first method the images are acquired with a scanning procedure, i.e. each spatially distributed pixel is read independently at different times. In the second one all the image is probed at once in a single shot, which may be preferable in applicative scenarios. While the difference between the two configurations is mainly technical the single shot acquisition introduces a not trivial trade-off between spatial resolution and quantum enhancement that we will comment.

5.1 The patter recognition problem

A binary pattern is defined as an array of $d \times d$ binary pixels. Each pixel encodes one bit of information. Let us consider a physical realization of the pixel that, as in Chapter 3, physically stores the information in two possible values of transmittance, τ_0 and τ_1. A pattern recognition task is defined as the classification of patterns into

predetermined categories, an example being the classification of handwritten numerical digits. Since we are considering patterns stored in a physical support the first step to be performed is a sensing procedure to efficiently extract the information. The sensing should be designed to minimize the pattern classification error probability, p_{err}^C. As in the previous chapters, the general sensing procedure can be described by an input state ρ, in general bipartite to allow ancillary assistance, interacting with the object encoding the information, and an output POVM measurement Λ followed by classical post-processing, as shown in Fig. (5.1.A). In the following, the measurement Λ considered we will be restricted to be only local, single-pixel, ones.

After the sensing the outcome is processed classically. For local measurements this means that a binary value is assigned to each pixel with a certain probability of error p_{err}^{px}. Then the pattern, represented by a certain configuration of the cell's values is classified by a given classical algorithm. The bit-flip probability p_{err}^{px} is related to the pattern classification error p_{err}^C by a generally non-linear map $\mathcal{M}_C$ defined by:

$$p_{err}^C = \mathcal{M}_C(p_{err}^{px}) \tag{5.1}$$

clearly the map $\mathcal{M}_C$ will depend on the classification problem at hand, i.e. the images space in which the patterns are defined and the classes considered, and it is reasonable to assume it to be a *monotonic* map for any effective classification algorithm. Given the monotonicity of $\mathcal{M}_C$ the minimization of p_{err}^C is equivalent to minimizing p_{err}^{px}, over all possible input state and POVM measurements.

We consider once again sensing performed by bosonic states at fixed energy, so the *single-pixel readout* problem is the same as what discussed in Chapter 3. The readout strategies considered here are:

- **Optimal classical strategy**: Classical input states ρ_{cla}, i.e. convex superposition of coherent states, are paired with a theoretical optimal measurement Λ. This strategy was analysed in Sec.(3.1), and its performance is given in Eq. (3.3). However, the measurement needed to saturate this theoretical bound appears non-feasible experimentally.

- **PC classical strategy**: An experimentally feasible classical strategy based on photon counting (PC) measurements at the receiver, with performance close to optimality. This second strategy is discussed in Sec.(3.2.2) and its best performance is reported in Eq. (3.28).

- **PC Quantum strategy**: The quantum strategy is realized using a product of Two Mode Squeezed Vacuum (TMSV) States, $^{\otimes M}|\text{TMSV}\rangle$, as input and a PC measurement. Its performance is analysed in Sec.(3.2.3).

We point out that the cells' size and dimension of the array d^2, can be a native characteristic of the physical support but also they can be practically determined by resolution properties of the sensing apparatus, for example the pixel structure of a camera or the resolution of an optical system, so that, in practice, τ_0 and τ_1 are intended as the values corresponding to each readout pixel. As anticipated we experimentally realized the quantum enhanced pattern recognition in two configurations, a scanning-like and a parallel, single-shot one. These experimental realizations are discussed in the following section.

5.2   Experimental Pattern recognition

The two experimental schemes are pictured in Fig. 5.1 (B-C). The state $\rho_{qua} = {}^{\otimes M}|\text{TMSV}\rangle_{S,I}$ is produced by travelling wave Spontaneous Parametric Down Conversion (SPDC) of type II in a non-liner crystal pumped by a continuous laser. Due to the conservation of the transverse momentum $\mathbf{q}$, in particular $\mathbf{q}_S = -\mathbf{q}_I$, signal and idler photons are emitted along correlated directions (spatial modes) which are mapped into symmetric pixels of a CCD camera. As was the case in the previous Chapters, in experimental conditions, the exposure time of an image acquisition is much longer than the coherence time of the process, resulting in many temporal modes (on the order of 10^{11}) for each acquisition, giving a total number of photons collected, μ in the order of a few thousands. The conditions $M \gg 1$ and $\mu \ll 1$ which maximize the quantum advantage are clearly fulfilled and, at the same time, so that the single-pixel photon statistics is practically indistinguishable from the Poisson one. For this reason, the best classical performance with photon counting can be experimentally reached by considering only the signal beam in this configuration, according to the results of Sec.(3.2.2) stating that the classical limit in Eq. (3.28) is reached by any Poisson distributed input states. The experimental setup is the same described in the previous chapters and a more detailed description can be found in Sec.(3.2.4) and Sec. (4.1.6).
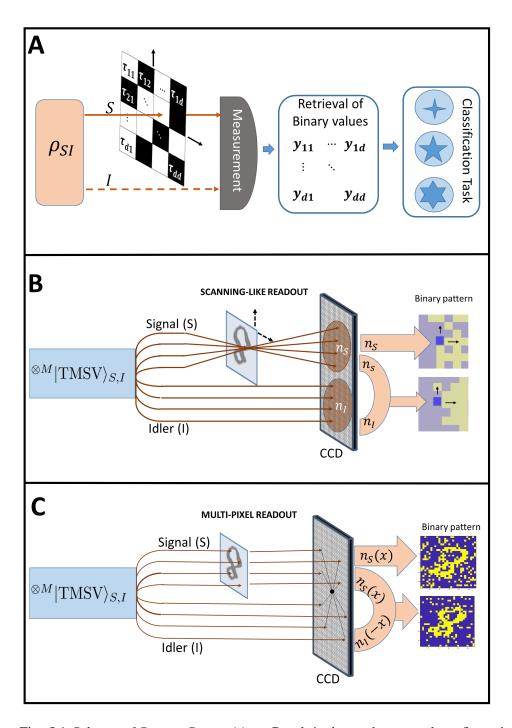
Fig. 5.1 *Scheme of Pattern Recognition*. Panel **A** shows the general configuration of a pattern recognition problem. An image composed of binary pixels, having one of two possible values of transmittance, is probed by a bipartite state ρ_{SI}. The signal (S) system interacts with the image. The idler (I) system is directly sent to the measurement apparatus where a joint measurement is performed with the signal. The result of the measurement are processed classically to classify the patterns. Panel **B** show a possible experimental configuration, called scanning-like, in which each pixel of the image is read separately and the total image is "scanned" with sequential single-pixel measurements. Panel **C** shows a different configuration in which the image is read in a single shot and pairwise correlation of the different modes of the initial state are used to enhance the classification.

Traveling wave SPDC is, in general, spatially broadband. Hence there are many, in the order of several thousands, pairwise correlated spatial modes [82–84] available for spatially resolved multi-pixel quantum readout. However, this introduces a trade-off between the spatial resolution, in terms of the number of spatial pixels available in a single-shot measurement, and the magnitude of reduction of the pixel error probability due to quantum correlation. This point is discussed in Sec.(5.2.2).

In the next section we will describe the "scanning-like readout" experimental configuration, that yields the best quantum enhancement. In section (5.2.2) we will present the "multi-pixel readout" that allows spatially resolved pattern acquisitions within a single shot measurement, a scenario closer to real applications.

5.2.1 Scanning-like readout

The scanning-like scenario mimics a readout in which the physical binary pattern encoded by the transmittance τ_0 and τ_1 is scanned point-by-point. This configuration is shown in Fig. (5.1-A). First, a large number of independent measurements of two well characterized transmittances τ_0 and τ_1 is taken. A bit value ($y = 0, 1$) for each measurement is assigned according to Bayes rule following the procedure discussed in Sec.(3.2.1). Here, the photon-counts registered into a large region of the CCD array, collecting a large number of signal and idler spatial modes respectively, are summed up to mimic a single-pixel. This allows to get the best possible correlation between the signal and idler number of photons, n_S and n_I. The result of this procedure are two sets of binary data $A_0 = \{y_k^{(0)}\}_{k=1,..,K}$ and $A_1 = \{y_l^{(1)}\}_{l=1,..,L}$ for the known values of the transmittance τ_0 and τ_1 respectively, where $L + K \gg d^2$. This same procedure is performed both for the quantum strategy and the classical one, where only the signal beam is used.

From those sets the average experimental pixel error probabilities can be evaluated. The results are reported in (5.2.**A**). The plot shows the probability of error as a function of one of the transmittances τ_0, while the other value is fixed to $\tau_1 = 1$, along with all the other parameters. The experimental points are reported both for the classical (red) and quantum (blue) strategies realized with PC measurements, along with their theoretical curves, and compared with the theoretical bound on the classical performance (green), defined by Eq. (3.3). In general the error probability, for all strategies, increases as the two transmittances become closer and closer, as
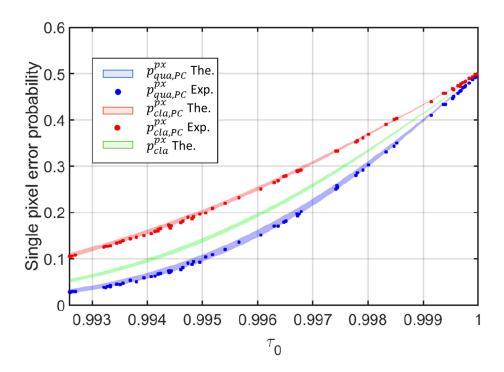
Fig. 5.2 *Single-pixel readout error probabilities.* The plot reports the error probability in the readout of the single-pixel as a function of one of the transmittances, τ_0, while the other is kept fixed at $\tau_1 = 1$. The estimated mean number of photons per pixel is $\mu = (1.45 \pm 0.05) \cdot 10^5$, the efficiencies are $\eta_S = 0.795 \pm 0.01$ and $\eta_I = 0.79 \pm 0.005$, and the electronic noise is $v \approx 6 \cdot 10^3$. The probability of error for the quantum strategy is reported in blue, the classical strategy with photon counting in red and the best classical strategy in green.

expected. However, through the range of parameters showed, the strategy employing quantum states sensibly outperforms the classical one paired with the suboptimal PC measurement and clearly surpasses also the theoretical absolute classical bound.

The data from the set A_0 and A_1 can be virtually rearranged to reproduce the outcomes of a sequential spatial scanning of any given $d \times d$ binary pattern. For this reason we call this configuration "scanning-like readout". Note that, for an accurate evaluation of the pattern recognition task, which is the goal of the present analysis, the used approach is highly preferable with respect to the real scanning of spatial samples. In fact, the last one would requires the physical realization and scanning of a large number of different spatial sample, which is hardly feasible and actually not necessary.
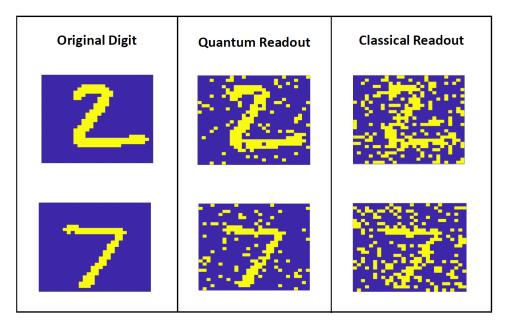
Fig. 5.3 *Readout of handwritten digits with different strategies.* Two examples of binary pattern of dimension 28x28 with quantum and classical (Photon counting) readout noise.

The pattern recognition performance was evaluated using the MINST handwritten digit dataset [85], containing a training set composed of 60 thousand samples of handwritten numerical digits and a test set of 10 thousand ones. An experimental test set for the pattern recognition evaluation is assembled by reproducing all the test patterns by picking experimental data from the sets A_0 and A_1. Some examples of digit contained in the dataset, and a comparison of the effect of the noise in either a classical and a quantum binary readout are shown in Fig. (5.3).

The noisy digits produced by the sensing were classified using a *k-Nearest Neighbor* (k-NN) classifier. A k-NN classifier compares the pattern tested with all the patterns in its training set. A metric is decided to evaluate the distance between the tested pattern and the training ones. In the case of binary patterns the most natural metric is the Hamming distance, defined as the number of discording pixels for two patterns. According to the metric chosen, the k closest patterns in the training set to the tested pattern are selected and the class of the test pattern is assigned as the more common among the k closest training patterns. The most intuitive example is given by the case $k = 1$ when the algorithm simply assigns the class of the closest pattern in the training set to the tested one. The classification error P^C_{k-NN} is reported in Fig. (5.4.**A**) as a function of the transmittance τ_0. Note how the performance of the (k-NN) classifier is related in a highly non-linear way to the single-pixel
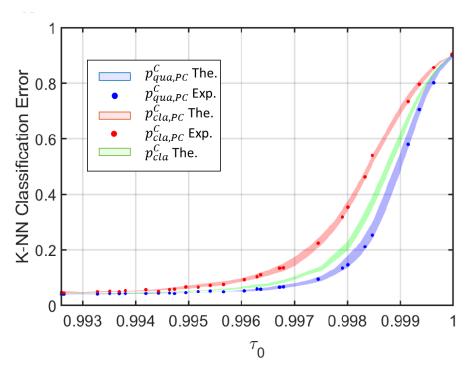
Fig. 5.4 *Classification error with the K-NN algorithm.* The plots show the classification error in the handwritten digit recognition as a function of one of the encoding transmittance value, τ_0, while $\tau_1 = 1$. The points reported refer to a classification performed with a k-NN algorithm. Experimental points are reported for the quantum (blue) and classical (red) photon counting readout along with the theoretical predictions. The green band reports a simulation of the performance obtained with a pixel readout at the absolute classical limit.

error probability. For smaller τ_0, roughly corresponding to a classical pixel error probability below 0.3 (see Fig. 5.2.**A**), the classification is very robust to noise. In this range, in fact, the classification error is very small for all the strategies and, consequently, the quantum advantage is also rather small. For higher transmittance τ_0, however, the algorithm becomes very sensitive to the pixel readout noise level and here the advantage provided by the quantum strategy is amplified. For instance, for $\tau_0 = 0.9983$, a modest pixel error probability advantage, from $p_{cla,PC}^{px} = 0.39$ to $p_{qua,PC}^{px} = 0.33$, causes the classification error to be reduced from $P_{k-NN}^{C} \approx 46\%$ to $P_{k-NN}^{C} \approx 21\%$.

It is natural to wonder if the amplification of the quantum advantage observed for the k-NN classifier is a property of this specific algorithm or is a more general behavior characterizing the pattern classification task. To partially address this question, we performed the same analysis with another classification algorithm that

uses a more complex and refined Machine Learning approach: a Convolutional Neural Network (CNN).

While for the k-NN algorithm a noiseless training set was the best option this is not the case for the CNN. If no noise is introduced during the training stage, thus not allowing the CNN to "adapt" to a noisy classification, the algorithm could perform poorly when presented with very noisy test images. This is confirmed by the numerical investigation, showing that, within statistical fluctuations, the best classification performance is achieved when the algorithm is trained with training patterns having similar levels of noise as the test ones. Fig. (5.5) reports the results for CNN classification performance.

By comparing Fig. (5.5) and Fig. (5.4), it appears that the behaviour of CNN and k-NN classifiers is qualitatively very similar. In particular, the results confirm that the advantage in sensing provided by quantum resource, in the single-pixel readout, is preserved and even amplified when considering further complex non-liner post-processing task, such as pattern recognition.

5.2.2   Parallel multi-pixel readout

The scanning-like readout of the previous section represents a meaningful proof-of-principle, necessary for a clean and faithful comparison with the theory. In this section we consider the case of a multi-pixel parallel readout. This configuration is depicted in Fig. 5.1 B. The pattern is acquired in single-shot, as it would be convenient in various realistic scenarios. The TMSV state produced by SPDC is spatially multimode characterized by a Gaussian intensity cross-correlation with finite coherence length l_c, which can be much smaller than the beam size, so that a spatially parallel readout can be performed using a single source. Pairwise correlated modes are detected by symmetric pairs of pixels in the CCD chip as represented in Fig. 5.1 B, so that thousand pixel pairs are in principle non-classical correlated at the same time. However, to efficiently collect the correlated modes, the size of the pixel l_D must be larger than the coherence length. This poses a limit in the number of independent pixel available for a parallel readout. It is convenient to introduce a "collection efficiency" η_c, representing the probability that, given a photon detected in a pixel namely of the signal arm, its twin arrives in the appointed pixel in the idler arm. Let us denote the single-arm efficiencies η_S and η_I, for the signal and
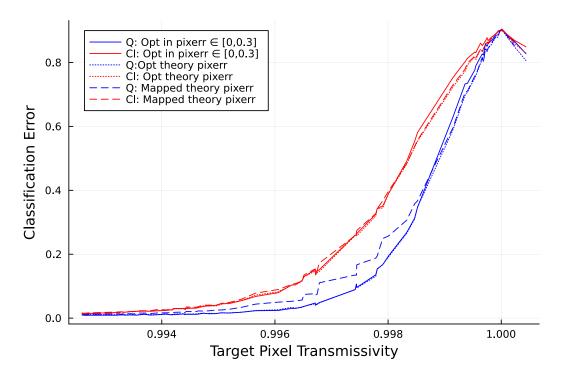
Fig. 5.5 *Classification error with the CNN*. The classification error as a function of τ_0 ($\tau_1 = 1$) is reported for a CNN. Simulated points are reported for the quantum (blue) and classical (red) photon counting readout. The classification performance in this case depends on the level of noise of the training set. The best classification is shown in the solid line, while the performance obtained by training with the theoretical predicted level of noise is reported in dashed lines.

the idler arm respectively. In this way, the probability to detect a pair of correlated photon in two selected pixels is the product $\eta_c \eta_S \eta_I$. The photon statistics and the correlation are modified by the non-unitary efficiency. In particular, the covariance of the detected photon numbers n_S and n_I, is:

$$\langle \Delta n_S \Delta n_I \rangle = \eta_c \eta_S \eta_I \langle \Delta n_S \Delta n_I \rangle_{in} \qquad (5.2)$$

where $\langle \Delta n_S \Delta n_I \rangle_{in}$ is the initial covariance of the source. More in depth discussion on this topic is presented, for example, in Ref.s [75, 86]. The collection efficiency η_c increases as the dimension of the detection pixel becomes much larger than the coherence area of the source. In the single cell-readout configuration of Chapter (3-4) or the 'scanning-like readout' presented in the last section, it is possible to collect a large amount of coherence areas in a single detection pixel, so that $\eta_c \to 1$, whereas when spatial resolution is required, η_c will typically be smaller than 1. The consequence of this reduced correlation efficiency is a trade-off between the spatial resolution and the amount of the quantum advantage in the single-pixel readout, since the classical strategies do not use correlations to reach the best performance, while the TMSV quantum one does. The enhancement of the quantum signature as the size of the detection pixel increases can be seen by plotting the Noise reduction Factor (NRF), orange curve in Fig. (5.6). The NRF, defined as NRF $= \langle \Delta^2 (n_S - n_I) \rangle / \langle n_S + n_I \rangle$ [82, 75], is an indicator of non-classicality. Assuming balanced channels efficiency $\eta_S = \eta_I = \eta_0$, can be written in terms of η_c as NRF $= 1 - \eta_0 \eta_c$ [75]. The effective resolution was reduced by performing an averaging filter, of appropriate size, in order to preserve the number of pixels in the final image.

Fig. 5.6 A shows the experimental pixel error probabilities for both classical an quantum PC readout in function of the resolution, namely the pixel size. On the one side, the increase of the pixel size has the direct consequence of raising the number of signal photons per pixel, that explains the dropping of classical error probability. On the other side, quantum error probability decreases faster than the classical analogous, because the detectable quantum correlation increases while loosing spatial resolution.

Here, the comparison of the experimental results with a theoretical bounds is not straightforward and we will avoid it. In fact, the binary pattern are physically fabricated with certain characteristic spatial features. When the image is binned at a lower resolution, comparable with the spatial features, for a pixel that falls across
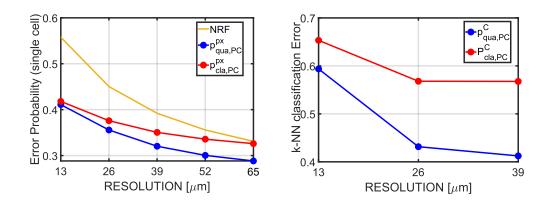
Fig. 5.6 *Multi-Pixel readout performance as a function of the spatial resolution.* In the left panel the single-pixel error is shown, as well as the NRF (see main text). The mean number of photons per pixel, at full resolution is $\mu = 1380 \pm 20$, while the total channel efficiencies of signal and idler branches are $\eta_S = 0.795 \pm 0.01$ and $\eta_I = 0.815 \pm 0.01$. In the right panel it is reported the error in the pattern classification with a k-NN algorithm.

the edge between τ_0 and τ_1 the binary model is no longer valid. Such border effects becomes not negligible as the resolution decreases and they are very complicated to be accounted for in a theoretical model. In fact the number of pixels on the border, the distribution of their effective average transmittance depends on the characteristic of the set of patterns.

We evaluated the pattern recognition performance once again on the MINST handwritten digit dataset using a k-NN algorithm for the classification. While the test dataset is composed of 10 thousand patterns, producing and acquiring 10 thousand samples is a not feasible and unnecessary. In fact, the error in the classification can be broken down in two contributions. The first contribution is intrinsic to the classification algorithm while the second one comes from the noise in the readout. In our case, with a pixel readout error larger than 0.3, the classification error is clearly dominated by the noise in the sensing. For reference, the classification error with noisy readout is in the range of $40-60\%$ while the intrinsic classification error without noise is of the order of 4%. So, we assume that the classification error can be well studied printing only a small subset of pattern from the original MINST set, namely 10, and performing on each of them one thousand different acquisitions with independent noise realizations, forming a total dataset of 10 thousands noisy test images. The samples, containing digits from 0 to 9, were realized on a AR-cooated glass slide combining laser lithography, sputtering deposition of titanium in high

vacuum and lift-off technique. The transmittance of the deposition is estimated as $\tau_0 = 0.987 \pm 0.003$.

The resulting classification error is reported in Fig. 5.6B as a function of the spatial resolution. The results are reported up to a resolution of 39 μm which is equivalent to an effective rescaling of a factor 3 with respect to the original size of the dataset (28x28 pixels). They confirm the main results of the of the previous section, in sensing scenario dominated by quantum noise, a relatively small quantum enhancement in the single-pixel probability, namely from 0.33 to 0.29, is converted in a substantial improvement, around 20%, in the classification.

5.3   Discussion

In this section, we demonstrated experimentally how quantum-enhanced sensing offers a significant advantage in the task of pattern recognition. As a test bed we have considered the problem of classification of handwritten digits. A quantum readout strategy based on TMSV states and photon counting measurement brings a relevant advantage in the classification errors with respect to any classical sensing.

The experiment was realized in two different configurations. The first one, a "scanning-like readout", was performed with the aim to reach the maximum quantum advantage at the price of a higher acquisition time, due to the fact that a scanning of the sample is required. In this case, the quantum strategy was able to clearly outperform the optimal classical bound.

In the second configuration a single-shot multipixel readout of the pattern was considered. In this case it was shown a large advantage with respect to the classical bound that uses photon counting, an interesting result in view of possible real applications.

To show that the quantum advantage is independent of complexity of the classification algorithm the same data was processed both with a k-NN classification, a simple but robust strategy, and a more computationally complex Convolutional neural network algorithm. In both cases the results obtained were very similar, the quantum advantage in the sensing is maintained and often amplified in the pattern recognition task. The results presented here are particularly promising for short terms application in biological spatial pattern recognition. Nonetheless, they also

pave the way to experimental studies of quantum-enhanced pattern recognition in the spectral and temporal domain.

Chapter 6

Conclusions

In this dissertation we analysed different protocols in the field of quantum hypothesis testing (QHT). This field, while very rich from a theoretical standpoint it's still at a stage in which experimental demonstrations are somewhat lacking, with some exceptions like the achievements in the specific case of quantum illumination. With the aim of fill the gap, the original results presented in the previous chapters showed experimentally an *absolute* quantum advantage in QHT protocols that may have significant importance for applications. In particular we focused on the quantum reading (QR) of classical information encoded in a digital memory, as well as quantum conformance test (QCT) of production process and pattern recognition tasks. Here by absolute we mean an experimental advantage obtained over theoretical ultimate bounds on the classical performances, such as the ones reported in Eq.(3.3) and Eq.(4.15). These results are, in fact even more noteworthy when considering the rarity of practical demonstrations in this particular field. Moreover we want to point out that even though the results were indeed proof-of-principle, the technique used are relatively simple when compared to what is required for other quantum technologies or even some sophisticated protocols in classical communication. In all the instances reported, an advantage was shown using TMSV states, that nowadays are routinely produced in quantum optics laboratories, and well developed correlated photon counting techniques. The robustness shown by these techniques to detection noise, and particularly optical losses, suggest that even though these realization are really recent, having all been published in the last couple of years, possible real applications may not be so far.

The first original contribution that we discussed, in chapter 3, was the experimental realization of the QR protocol. This protocol has been proposed in 2011 but an experimental demonstration was lacking. In our realization we implemented a very simple receiver design consisting in photon counting measurements followed by a maximum likelihood decision. As it turns out this simple design, paired with TMSV input states is enough to reach a convincing advantage in the protocol, close to the theoretical predictions of the original proposal, at least considering ideal detectors. Moreover this design showed a very good resilience to optical losses, with a significant portion of the advantage preserved in experimental conditions in which the total efficiency was slightly lower than 80%. We stress once again the importance of this point, as optical losses are often the dominant deviating effect between ideal performances and practical one.

The QCT protocol, introduced in Chapter 4, describes the discrimination of two production processes, namely discovering possible deviations of a process with respect to a reference. Each process outcome is represented by a classical ensamble of quantum channels and thus the QCT can be seen as a generalization of the QR protocol, from the discrimination of pure loss channels, to the discrimination between convex combinations of channels. For this protocol we found an analytical expression for a lower bound on the probability of error, characterizing the performance of classical states. With the same quantum readout strategy of the QR, we demonstrated theoretically and experimentally also for the QCT an advantage with respect to the classical counterparts. Additionally a deep analysis of the false positive and false-negative types of discrimination errors has been done, in view of their practical importance in the context of QCT. Moreover, we showed how the formalism and results obtained in the QCT protocol can be used to model a different problem, namely a digital memory with *imperfect* writing. In this context we demonstrated how the quantum strategy with direct single cell encoding and readout is not only able to beat the classical strategy in the same scenario, but remarkably also a more general, *global* classical strategy. A global strategy consists of codeword encoding over a large amount of cells, and a global output readout. In this scenario we were able to find a limit to the performance by formulating the problem in terms of mutual information over repeated uses, which is accounted by the Holevo bound.

Finally in Chapter 5 we showed that the quantum enhancement offered by quantum states to the sensing can dramatically improve the performance in a complex task such as pattern recognition. It turns out that the relation between the pixel error

and pattern classification error is not a linear function. As a exemplary case, we considered the classification of handwritten digits. When evaluating the classification performance, even an apparently modest quantum reduction of the single pixel error results in a significant improvement in the classification. The quantitative advantage obtained clearly depends on the classification algorithm used, and for this reason we considered two paradigmatic examples: a k-nearest neighbour classifier and a convolutional neural network. The first is one of the most simple classification algorithm avaliable but resilient to input noise, while the second one is far more complex. The overall results were very similar for both algorithms. Our results strongly suggest that in this task the quantum advantage is non only preserved but even amplified in some regions. This is indeed a very interesting result in view of possible application of quantum techniques to complex tasks.

In conclusion in this work we tried to approach the field of quantum hypothesis testing/ quantum channel discrimination in a more application oriented way. In the fast developing field of quantum technologies an analysis of the practical feasibility of the various protocols is indeed very important, but still missing in some cases. The protocols we discussed were for the most part concerned with pure loss channels a very well studied topic. In particular, in quantum metrology it has been known for many years that photon counting measurements yield near optimal results, in the estimation of the loss parameter. Nonetheless, as far as we know, we were the first to analyze their performance in the discrete case. Maybe unsurprisingly we found that, as in the continuous counterpart, photon counting measurements perform very well. The analysis of the practical attainability will play an increasingly important role in quantum hypothesis testing as the fields approaches near technological maturity, and we think that drawing inspiration from fields such as quantum metrology or the also very closely related quantum communication could be very beneficial, with the opposite being of course true as well. The experimental realization of protocols in QHT is of huge interest in view of the relevance and variety of their possible applications. For the discrimination of bosonic losses these applications range from the enhancement of classical data readout, to conformance testing, e.g. for chemical concentrations and compositions, spectroscopy, biological pattern recognition, unknown objects ranging and position finding and many others.

References

[1] G. Ortolano, E. Losero, S. Pirandola, M. Genovese and I. Ruo-Berchera. Experimental quantum reading with photon counting. *Science Advances*, 7(4):eabc7796, 2021.

[2] G. Ortolano, P. Boucher, I.P. Degiovanni, E. Losero, M. Genovese and I. Ruo-Berchera. Quantum conformance test. *Science Advances*, 7(52):eabm3093, 2021.

[3] Giuseppe Ortolano and Ivano Ruo-Berchera. Quantum readout of imperfect classical data. *Sensors*, 22(6), 2022.

[4] G. Moore. Cramming more components onto integrated circuits. *Electronics*, Volume 38(8), April 1965.

[5] N. Gisin. Quantum communications. In *Advanced Photonics (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*, page SpM3G.5, 2018.

[6] G. Chesi, S. Olivares and M. G. A. Paris. Squeezing-enhanced phase-shift-keyed binary communication in noisy channels. *Phys. Rev. A*, 97:032315, 2018.

[7] A. Allevi and M. Bondani. Novel scheme for secure data transmission based on mesoscopic twin beams and photon-number-resolving detectors. *Scientific Reports*, 12(1):15621, 2022.

[8] P.W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994.

[9] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. Shamsul Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi and P. Wallden. Advances in quantum cryptography. *Adv. Opt. Photon.*, 12(4):1012–1236, 2020.

[10] C. L. Degen, F. Reinhard and P. Cappellaro. Quantum sensing. *Reviews of Modern Physics*, 89:035002, 2017.

[11] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook and S. Lloyd. Advances in photonic quantum sensing. *Nature Photonics*, 12(12):724–733, 2018.

[12] A. S. Clark, Maria Chekhova, J. C. F. Matthews, J. G. Rarity and R. F. Oulton. Special topic: Quantum sensing with correlated light sources. *Applied Physics Letters*, 118(6):060401, 2021.

[13] V. Giovannetti, S. Lloyd and L. Maccone. Advances in quantum metrology. *Nature Photonics*, 5:222 EP –, 2011. Review Article.

[14] M. A. Taylor and W. P. Bowen. Quantum metrology and its application in biology. *Physics Reports*, 615:1 – 59, 2016.

[15] E. Polino, M. Valeri, N. Spagnolo and F. Sciarrino. Photonic quantum metrology. *AVS Quantum Science*, 2(2):024703, 2020.

[16] N. Samantaray, I. Ruo-Berchera, A. Meda and M. Genovese. Realization of the first sub-shot-noise wide field microscope. *Light: Science & Applications*, 6:e17005 EP –, 2017.

[17] J. Sabines-Chesterking, A. R. McMillan, P. A. Moreau, S. K. Joshi, S. Knauer, E. Johnston, J. G. Rarity and J. C. F. Matthews. Twin-beam sub-shot-noise raster-scanning microscope. *Optics Express*, 27(21):30810–30818, 2019.

[18] G. Brida, M. Genovese and I. Ruo Berchera. Experimental realization of sub-shot-noise quantum imaging. *Nature Photonics*, 4:227–230, 2010.

[19] C. Lupo and S. Pirandola. Ultimate precision bound of quantum and subwavelength imaging. *Phys. Rev. Lett.*, 117:190802, 2016.

[20] M. Genovese. Real applications of quantum imaging. *Journal of Optics*, 18(7):073002, 2016.

[21] I. Ruo-Berchera and I. P. Degiovanni. Quantum imaging with sub-poissonian light: challenges and perspectives in optical metrology. *Metrologia*, 56(2):024001, 2019.

[22] C. Lupo, Z. Huang and P. Kok. Quantum limits to incoherent imaging are achieved by linear interferometry. *Phys. Rev. Lett.*, 124:080503, 2020.

[23] M. Tsang, R. Nair and X.-M. Lu. Quantum theory of superresolution for two incoherent optical point sources. *Phys. Rev. X*, 6:031033, 2016.

[24] Z. Huang and C. Lupo. Quantum hypothesis testing for exoplanet detection. *Phys. Rev. Lett.*, 127:130502, 2021.

[25] U. Zanforlin, C. Lupo, P. W. R. Connolly, P. Kok, G. S. Buller and Z. Huang. Optical quantum super-resolution imaging and hypothesis testing. *Nature Communications*, 13(1):5373, 2022.

[26] H. Shi, Z. Zhang, S. Pirandola and Q. Zhuang. Entanglement-assisted absorption spectroscopy. *Phys. Rev. Lett.*, 125:180502, 2020.

[27] Z. Huang, C. Schwab and C. Lupo. Ultimate limits of exoplanet spectroscopy: A quantum approach. *Phys. Rev. A*, 107:022409, 2023.

[28] S. Olivares and M. G A Paris. Bayesian estimation in homodyne interferometry. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 42(5):055506, 2009.

[29] 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.

[30] R. Demkowicz-Dobrzanski, M. Jarzyna and J. Kolodynski. Quantum limits in optical interferometry. *Progess in Optics*, 60:345, 2015.

[31] C. Sparaciari, S. Olivares and M. G. A. Paris. Gaussian-state interferometry with passive and active elements. *Phys. Rev. A*, 93:023810, 2016.

[32] M. V. Chekhova and Z. Y. Ou. Nonlinear interferometers in quantum optics. *Adv. Opt. Photon.*, 8(1):104–155, 2016.

[33] R. Schnabel. Squeezed states of light and their applications in laser interferometers. *Physics Reports*, 684:1 – 51, 2017.

[34] C. Schäfermeier, M. Ježek, L. S. Madsen, T. Gehring and U. L. Andersen. Deterministic phase measurements exhibiting super-sensitivity and super-resolution. *Optica*, 5(1):60–64, 2018.

[35] G. Ortolano, I. Ruo-Berchera and Enrico Predazzi. Quantum enhanced imaging of nonuniform refractive profiles. *International Journal of Quantum Information*, 17(08):1941010, 2019.

[36] J. Aasi et al. Enhanced sensitivity of the ligo gravitational wave detector by using squeezed states of light. *Nature Photonics*, 7(8):613–619, Aug 2013.

[37] C. Helstrom. *Quantum detection and estimation theory*. Academic Press, New York, 1976.

[38] A. Chefles and S. M. Barnett. Quantum state separation, unambiguous discrimination and exact cloning. *Journal of Physics A: Mathematical and General*, 31(50):10097–10103, 1998.

[39] A. Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, 2000.

[40] J. A. Bergou, U. Herzog and M. Hillery. *11 Discrimination of Quantum States*, pages 417–465. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[41] A. W. Harrow, A. Hassidim, D. W. Leung and J. Watrous. Adaptive versus nonadaptive strategies for quantum channel discrimination. *Physical Review A*, 81:032339, 2010.

[42] S. Pirandola, R. Laurenza, C. Lupo and J. L. Pereira. Fundamental limits to quantum channel discrimination. *npj Quantum Information*, 5(1):50, 2019.

[43] S. Lloyd. Enhanced sensitivity of photodetection via quantum illumination. *Science*, 321(5895):1463–1465, 2008.

[44] S.-H. Tan, B. I. Erkmen, V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, S. Pirandola and J. H. Shapiro. Quantum illumination with gaussian states. *Physical Review Letter*, 101:253601, 2008.

[45] E. D. Lopaeva, I. Ruo Berchera, I. P. Degiovanni, S. Olivares, G. Brida and M. Genovese. Experimental realization of quantum illumination. *Physical Review Letter*, 110:153603, 2013.

[46] Y. Zhang, D. England, A. Nomerotski, P. Svihra, S. Ferrante, P. Hockett and B. Sussman. Multidimensional quantum-enhanced target detection via spectrotemporal-correlation measurements. *Phys. Rev. A*, 101:053808, 2020.

[47] T. Gregory, P.-A. Moreau, E. Toninelli and M. J. Padgett. Imaging through noise with quantum illumination. *Science Advances*, 6(6):eaay2652, 2020.

[48] S. Pirandola. Quantum reading of a classical digital memory. *Physical Review Letter*, 106:090504, 2011.

[49] Z. Zhang, S. Mouradian, F. N. C. Wong and J. H. Shapiro. Entanglement-enhanced sensing in a lossy and noisy environment. *Physical Review Letter*, 114:110506, 2015.

[50] G. Sorelli, N. Treps, F. Grosshans and F. Boust. Detecting a target with quantum entanglement. *IEEE Aerospace and Electronic Systems Magazine*, 37(5):68–90, 2022.

[51] L. Banchi, Q. Zhuang and S. Pirandola. Quantum-enhanced barcode decoding and pattern recognition. *Phys. Rev. Appl.*, 14:064026, 2020.

[52] F. Nielsen. An information-geometric characterization of chernoff information. *IEEE Signal Processing Letters*, 20(3):269–272, 2013.

[53] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[54] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, Ll. Masanes, A. Acin and F. Verstraete. Discriminating states: The quantum chernoff bound. *Phys. Rev. Lett.*, 98:160501, 2007.

[55] M. Nussbaum and A. Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *The Annals of Statistics*, 37(2):1040 – 1057, 2009.

[56] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems Inform. Transmission*, 9(3):177–183, 1973.

[57] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland and W. K. Wootters. Classical information capacity of a quantum channel. *Phys. Rev. A*, 54:1869–1876, 1996.

[58] A.S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998.

[59] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, USA, 10th edition, 2011.

[60] A. Shaji and E.C.G. Sudarshan. Who's afraid of not completely positive maps? *Physics Letters A*, 341(1):48–54, 2005.

[61] A. S. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction*. De Gruyter, 2012.

[62] S. Pirandola, C. Lupo, V. Giovannetti, S. Mancini and S. L. Braunstein. Quantum reading capacity. *New Journal of Physics*, 13(11):113012, 2011.

[63] R. Nair. Discriminating quantum-optical beam-splitter channels with number-diagonal signal states: Applications to quantum reading and target detection. *Physical Review A*, 84:032312, 2011.

[64] A. Bisio, M. Dall'Arno and G. M. D'Ariano. Tradeoff between energy and error in the discrimination of quantum-optical devices. *Physical Review A*, 84:012310, 2011.

[65] O. Hirota. Error free quantum reading by quasi bell state of entangled coherent states. *Quantum Measurements and Quantum Metrology*, 4(1):70 – 73, 2017.

[66] M. Dall'Arno, A. Bisio, G. M. D'Ariano, M. Miková, M. Ježek and M. Dušek. Experimental implementation of unambiguous quantum reading. *Phys. Rev. A*, 85:012308, 2012.

[67] J. Peřina, M. Hamar, V. Michálek and O. Haderka. Photon-number distributions of twin beams generated in spontaneous parametric down-conversion and measured by an intensified ccd camera. *Physical Review A*, 85:023816, 2012.

[68] A. Allevi, M. Lamperti, M. Bondani, J. Peřina, V. Michálek, O. Haderka and R. Machulka. Characterizing the nonclassicality of mesoscopic optical twin-beam states. *Phys. Rev. A*, 88:063807, 2013.

[69] V. Giovannetti, S. Lloyd and L. Maccone. Quantum-enhanced measurements: Beating the standard quantum limit. *Science*, 306(5700):1330–1336, 2004.

[70] F. Paleari, A. Andreoni, G. Zambra and M. Bondani. Thermal photon statistics in spontaneous parametric downconversion. *Opt. Express*, 12(13):2816–2824, 2004.

[71] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, 1995.

[72] A. Allevi and M. Bondani. Can nonclassical correlations survive in the presence of asymmetric lossy channels? *The European Physical Journal D*, 72(10):178, 2018.

[73] E. Knyazev, F. Y. Khalili and M. V. Chekhova. Overcoming inefficient detection in sub-shot-noise absorption measurement and imaging. *Opt. Express*, 27(6):7868–7885, 2019.

[74] G. Frascella, S. Agne, F. Y. Khalili and M. V. Chekhova. Overcoming detection loss and noise in squeezing-based optical sensing. *npj Quantum Information*, 7(1):72, 2021.

[75] A. Meda, E. Losero, N. Samantaray, F. Scafirimuto, S. Pradyumna, A. Avella, I Ruo-Berchera and M Genovese. Photon-number correlation for quantum enhanced imaging and sensing. *Journal of Optics*, 19(9):094002, 2017.

[76] G. Brida, M. Genovese, I. Ruo-Berchera, M. Chekhova and A. Penin. Possibility of absolute calibration of analog detectors by using parametric down-conversion: a systematic study. *Journal of the Optical Society of America B*, 23(10):2185–2193, 2006.

[77] O. A. Ivanova, T. S. Iskhakov, A. N. Penin and M. V Chekhova. Multiphoton correlations in parametric down-conversion and their measurement in the pulsed regime. *Quantum Electronics*, 36(10):951, 2006.

[78] G. Brida, M. Chekhova, M. Genovese, M.-L. Rastello and I. Ruo-Berchera. Absolute calibration of analog detectors using stimulated parametric down conversion. *Journal of Modern Optics*, 56(2-3):401–404, 2009.

[79] A. Meda, I. Ruo-Berchera, I. P. Degiovanni, G. Brida, M. L. Rastello and M. Genovese. Absolute calibration of a charge-coupled device camera with twin beams. *Applied Physics Letters*, 105(10):101113, 2014.

[80] G. Brida, I.P. Degiovanni, M. Genovese, M. L. Rastello and Ivano Ruo-Berchera. Detection of multimode spatial correlation in pdc and application to the absolute calibration of a ccd camera. *Optics Express*, 18(20):20572, 2010.

[81] C. N. Gagatsos, A. I. Karanikas, G. Kordas and N. J. Cerf. Entropy generation in gaussian quantum transformations: applying the replica method to continuous-variable quantum information theory. *npj Quantum Information*, 2(1):15008, 2016.

[82] M. Bondani, A. Allevi, G. Zambra, M. G. A. Paris, and A. Andreoni. Sub-shot-noise photon-number correlation in a mesoscopic twin beam of light. *Phys. Rev. A*, 76:013833, 2007.

[83] A. Allevi, A. Andreoni, F. A. Beduini, M. Bondani, M. G. Genoni, S. Olivares and M. G. A. Paris. Conditional measurements on multimode pairwise entangled states from spontaneous parametric downconversion. *Europhysics Letters*, 92(2):20007, 2010.

[84] D. A. Kalashnikov, S.-H. Tan, T. S. Iskhakov, M. V. Chekhova and Leonid A. Krivitsky. Measurement of two-mode squeezing with photon number resolving multipixel detectors. *Opt. Lett.*, 37(14):2829–2831, 2012.

[85] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[86] G. Ortolano, P. Boucher, I. Ruo Berchera, S. F. Pereira, and M. Genovese. Phase retrieval enhanced by quantum correlation, 2021.

Appendix A

Photon number distribution after a loss

To convince oneself of Eq.(3.16) one can consider the more intuitive unipartite case. A generic bosonic input state ρ_I can be written, in the Fock basis, as:

$$\rho_I = \sum_{m,n=0}^{\infty} c_{n,m} |n\rangle\langle m| = \sum_{m,n=0}^{\infty} c_{n,m} \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}} |0\rangle\langle 0| \frac{(\hat{a})^m}{\sqrt{m!}} \qquad (A.1)$$

The effect of a pure loss channel, as already pointed out, is the same as the one of a beam splitter (BS), where only one port is considered at the output (see Fig.(A.1)). Let us consider the state ρ entering one of two input ports of a BS, that we denote I. The state of the other port, denoted as E, is the vacuum state, $\rho_E = |0\rangle_E\langle 0|$, since we are considering *pure losses*. If necessary one could consider, for example, a thermal states in this port, and the result would describe an evolution consisting in both losses and addition of thermal noise. Let us denote T the output of interest of the beam splitter, i.e. the one corresponding to the transmission of the input I, and R the other one. Both T and R are initially in a vacuum state, $\rho_T = |0\rangle_T\langle 0|$ and $\rho_R = |0\rangle_R\langle 0|$. The initial total state of the system is the tensor product of all the subsystems, $\rho_{TOT} = \rho_I \otimes \rho_E \otimes \rho_R \otimes \rho_T$. For bosonic input the evolution of a beam splitter is described by the already mentioned input-output relations. Inverting the input-output relations yields for the input field, in terms of the transmitted and reflected ones, $\hat{a} = \sqrt{\tau}\hat{a}_t - i\sqrt{(1-\tau)}\hat{a}_r$. The total evolution of ρ_{TOT} after the BS can be then computed by substituting $\hat{a}$ in terms of the transmitted and reflected
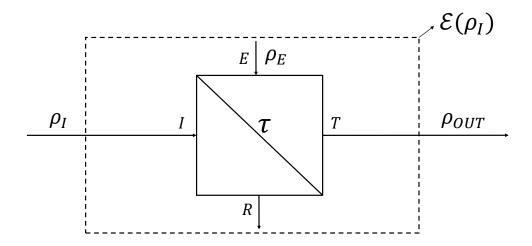
Fig. A.1 *Pure loss channel as a Beam splitter.* A pure loss channels can be seen as a beam splitter in which one of the output ports is traced out.

fields:

$$\rho_{OUT} = \sum_{m,n=0}^{\infty} c_{n,m} \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}} (|0\rangle_I \langle 0| \otimes |0\rangle_E \langle 0| \otimes |0\rangle_R \langle 0| \otimes |0\rangle_T \langle 0|) \frac{(\hat{a})^m}{\sqrt{m!}} =$$

$$= |0,0\rangle_{I,E} \langle 0,0| \otimes \sum_{m,n=0}^{\infty} c_{n,m} \frac{(\sqrt{\tau}\hat{a}_t^\dagger + i\sqrt{(1-\tau)}\hat{a}_r^\dagger)^n}{\sqrt{n!}} (|0,0\rangle_{R,T} \langle 0,0|) \frac{(\sqrt{\tau}\hat{a}_t - i\sqrt{(1-\tau)}\hat{a}_r)^m}{\sqrt{m!}}$$

The state of the subsystem RT, $\rho_{R,T}$, is then obtained by tracing out the systems I and E:

$$\rho_{R,T} = tr_{E,I}(\rho_{OUT}) =$$

$$= \sum_{m,n=0}^{\infty} c_{n,m} \frac{(\sqrt{\tau}\hat{a}_t^\dagger + i\sqrt{(1-\tau)}\hat{a}_r^\dagger)^n}{\sqrt{n!}} (|0,0\rangle_{R,T} \langle 0,0|) \frac{(\sqrt{\tau}\hat{a}_t - i\sqrt{(1-\tau)}\hat{a}_r)^m}{\sqrt{m!}}$$

$$(A.2)$$

For bosonic fields $\hat{a}_t$ and $\hat{a}_r$ are commuting operators so that the powers can be written in terms of their binomial expansions:

$$\rho_{R,T} = \sum_{m,n=0}^{\infty} c_{n,m} \sum_{k=0}^{n} \sum_{j=0}^{m} w_{m,n,k,j} |n-k,k\rangle_{R,T} \langle m-j,j| \qquad (A.3)$$

$$w_{m,n,k,j} := \binom{n}{k} \binom{m}{j} \frac{\sqrt{\tau^{n+m-k-j}(1-\tau)^{k+j}}\sqrt{n-k!}\sqrt{m-j!}\sqrt{k!}\sqrt{j!}}{\sqrt{n!}\sqrt{m!}}$$

Tracing out the system R yields the output state of the channel $\mathcal{E}_\tau(\rho_I)$:

$$\mathcal{E}_t(\rho_I) = tr_R(\rho_{R,T}) = \sum_{l=0}^{\infty} \langle l|\rho_{R,T}|l\rangle =$$

$$= \sum_{m,n=0}^{\infty} c_{n,m} \sum_{j=0}^{min\{m,n\}} \binom{n}{j}\binom{m}{j} \frac{\sqrt{\tau^{n+m-2j}}(1-\tau)^j \sqrt{n-j!}\sqrt{m-j!}j!}{\sqrt{n!}\sqrt{m!}} |n-j\rangle_T \langle m-j|$$

$$(A.4)$$

The photon number distribution for the output is then:

$$\langle n|\mathcal{E}_\tau(\rho_I)|n\rangle = \sum_{m=0}^{\infty} c_{m,m} \sum_{j=0}^{m} \binom{m}{j}\binom{m}{j} \frac{\tau^{m-j}(1-\tau)^j (m-j)! j!}{m!} \delta(n,m-j) =$$

$$= \sum_{m=n}^{\infty} c_{m,m} \binom{m}{n} \tau^n (1-\tau)^{n-N} = \sum_{m=n}^{\infty} p_0(m)\mathcal{B}(n|m,\tau) \qquad (A.5)$$

the term $\binom{m}{n}\tau^m(1-\tau)^{n-m} = \mathcal{B}(n|m,\tau)$ is, in fact, a binomial distribution with m trials and probability of success τ and $c_{m,m} = \langle m|\rho_I|m\rangle = p_0(m)$ is the photon number distribution of the input state. The process can be seen as each photon undergoing a Bernulli trial with probability of success τ giving raise to a binomial distribution with m trials. Eq.(A.5) shows how the effect of the channel $\mathcal{E}_\tau$ on an arbitrary initial photon number distribution $p_0(n)$ is to *compound* it with a binomial distribution. Eq.(3.16) follows from the fact that only the identity acts on the idler modes so that the initial distribution is composed with a binomial only on the signal system.

Appendix B

Equivalence between individual and collective measurements for independent modes

In this appendix we show how there is *no advantage in measuring the photons number separately for each mode rather than performing a collective measurement*. Consider the M independent modes to be measured separately. The joint distribution of the M pairs $\{\mathbf{n^i}\}_{i=1...M}$, $\mathbf{n^i} = (n_S^i, n_I^i)$, follows from Eq.(3.29) and from the independence of the modes:

$$P(\{\mathbf{n^i}\}|\tau) = \prod_{i=1}^{M} |c_n(n_I^i)|^2 \mathscr{B}(n_S^i|n_I^i, \tau) \tag{B.1}$$

Let's now consider a discrimination, between τ_0 and τ_1, done using each of the outcomes $\mathbf{n^i}$. In particular let's consider the ensemble $\mathscr{M} = \{\{\mathbf{n^i}\} | \Sigma_i^M n_S^i = n_S, \Sigma_i^M n_I^i = n_I\}$. The probability of success in the discrimination of this ensemble , $p_s(\mathscr{M})$, is the average over it of the probability of success of the single configurations, $p_s(\{\mathbf{n^i}\})$. Suppose that for a given configuration $\{\mathbf{n^i}\}$ the value chosen is τ_u, then the *a posteriori* probability is given by:

$$p(\tau_u|\{\mathbf{n^i}\}) = \frac{p(\{\mathbf{n^i}\}|\tau_u)}{p(\{\mathbf{n^i}\}|\tau_0) + p(\{\mathbf{n^i}\}|\tau_1)} \tag{B.2}$$

that give also the probability of success for the configuration $\{\mathbf{n^i}\}$. If τ_u is the choice for one of the configurations, it must be the choice for all the other elements in $\mathscr{M}$

as well, in fact substituting Eq.(B.1) in the condition for the choice $P(\tau_u|\{\mathbf{n^i}\}) \geq P(\tau_{1-u}|\{\mathbf{n^i}\})$ yields:

$$\tau_u^{\sum_i^M n_S^i}(1-\tau_u)^{\sum_i^M n_I^i - \sum_i^M n_S^i} \geq$$
$$\tau_{1-u}^{\sum_i^M n_S^i}(1-\tau_{1-u})^{\sum_i^M n_I^i - \sum_i^M n_S^i} \qquad (B.3)$$

That depends only on the sums of the occupation numbers and not on the specific configuration, since the sums are fixed in $\mathcal{M}$. Indeed solving the inequality above for $\sum_i^M n_1^i = n_S$ yields the same threshold in Eq. (3.34). The probability of success for each configuration in $\mathcal{M}$ is given by Eq.(B.2) so that the ensemble averaged probability of success is:

$$p_s(\mathcal{M}) = \mathbb{E}[p_s(\{\mathbf{n^i}\})]_{\mathcal{M}} =$$
$$= \sum_{\{\mathbf{n^i}\}\in\mathcal{M}} \frac{p(\{\mathbf{n^i}\}|\tau_u)}{p(\{\mathbf{n^i}\}|\tau_0) + P(\{\mathbf{n^i}\}|\tau_1)} p_{\mathcal{M}}(\{\mathbf{n^i}\}) \qquad (B.4)$$

Where $P_{\mathcal{M}}(\{\mathbf{n^i}\})$ is the distribution of the configurations in $\mathcal{M}$ given by:

$$p_{\mathcal{M}}(\{\mathbf{n^i}\}) = \frac{p(\{\mathbf{n^i}\})}{\sum_{\mathcal{N}} p(\{\mathbf{n^i}\})} =$$
$$= \frac{p(\{\mathbf{n^i}\}|\tau_0) + p(\{\mathbf{n^i}\}|\tau_1)}{\sum_{\mathcal{N}} (p(\{\mathbf{n^i}\}|\tau_0) + p(\{\mathbf{n^i}\}|\tau_1))} =$$
$$= \frac{p(\{\mathbf{n^i}\}|\tau_0) + p(\{\mathbf{n^i}\}|\tau_1)}{\mathcal{Z}} \qquad (B.5)$$

substituting in Eq.(B.4) we get:

$$p_s(\mathcal{M}) = \frac{1}{\mathcal{Z}} \sum^{\mathcal{N}} P(\{\mathbf{n^i}\}|\tau_u) \qquad (B.6)$$

The summation on the right hand side of Eq.(B.6) is the sum of the independent probabilities of all the configurations in $\mathcal{M}$,i.e. of all the configurations having a certain sum $\{n_S, n_I\}$, so it must be the probability distribution of the pair $\{n_S, n_I\}$. This can also be seen by an explicit calculation of this term:

$$\sum^{\mathcal{N}} P(\{\mathbf{n^i}\}|\tau_u) = (|c_n(n_I^i)|^2)^M \tau^{n_S}(1-\tau)^{n_I-n_S} \sum_{\mathcal{M}} \prod_i^M \binom{n_I^i}{n_S^i} \qquad (B.7)$$

where $(|c_N(n_2^i)|^2)^M = \prod_i^M |c_N(n_2^i)|^2$ has been taken out of the sum because it depends only on n_S and n_I and not on the configuration. For the sum on the right-end side we have:

$$\sum_{\mathcal{M}} \prod_i^M \binom{n_I^i}{n_S^i} = \overset{\sum_i n_I^i = n_I \; \sum_i n_S^i = n_S}{\sum} \; \sum \; \prod_i^M \binom{n_I^i}{n_S^i} =$$
$$= \overset{\sum_i n_I^i = n_I}{\sum} \binom{n_I}{n_S} = \binom{n_I + M - 1}{n_I} \binom{n_I}{n_S} \tag{B.8}$$

Where the Vandermonde equality was used to solve the sum on n_S^i and the sum on n_I^i was over a quantity not depending on it so it gave just a multiplication constant. Substituting this result gives $\sum_{\mathcal{M}} p(\{\mathbf{n^i}\}|\tau_u) = P(\{\mathbf{n}\}|\tau_u)$, $\{\mathbf{n^i}\} = \{n_s, n_I\}$, as expected. Eq.(B.6) can be then rewritten as:

$$p_s(\mathcal{M}) = \frac{p(\{\mathbf{n}\}|\tau_u)}{p(\{\mathbf{n}\}|\tau_0) + p(\{\mathbf{n}\}|\tau_1)} = p_s(\{\mathbf{n}\}) \tag{B.9}$$

Where the constant $\mathscr{Z}$ can be evaluated both by using the same identity used for the numerator or by normalization of $p_s(\mathcal{M})$. Eq. (B.9) shows that as anticipated the best performance offered by M modes separate measurements of independent modes coincides with the one given by a collective measurement.

Appendix C

QCT Data analysis

The total probability of error in the QCT procedure, p_{err}, can be divided in two separate contributions, as shown in Eq.(4.1). p_{10} is the probability of error conditioned to the object being produced by the process $\mathscr{P}_0$, while p_{01} refers to the conditional case in which the correct process is $\mathscr{P}_1$.

The process $\mathscr{P}_0$ is assumed to be strongly peaked around a value of transmission τ_0, as discussed in Sec.(4.1.8). In this scenario the contribution p_{10} can be evaluated experimentally using an object with transmittance τ_0 and repeating the discrimination procedure $N_{\mathscr{D}}$ times, to construct the experimental dataset of outcomes $\mathscr{D}$. The frequency of error $f_{\mathscr{D}}(\mathscr{P}_1|\mathscr{P}_0)$ on $\mathscr{D}$ will converge to the probability of error p_{10} as $N_{\mathscr{D}} \to \infty$.

On the other hand, the process $\mathscr{P}_1$ will have an arbitrary, but known, probability density, denoted $g_1(\tau)$. To estimate the probability p_{01}, a dataset of outcomes $\mathscr{D}$ needs to be constructed in a way that is representative of a test performed on objects distributed according to the density $g_1(\tau)$. This can be achieved by experimentally creating a collection of L datasets $\mathscr{D}_i$ each generated from a different value of transmittance τ_i, $1 \leq i \leq L$, in a range $[\tau_{min}, \tau_{max}]$ determined by the distribution $g_1(\tau)$ one wants to approximate. Consider, for example, the target distribution to be a uniform one having mean value $\bar{\tau}$ and half-width δ. This distribution can be approximated by taking measurements with L different object having transmittance τ equispaced in the interval $[\tau_{min} = \bar{\tau} - \delta, \tau_{max} = \bar{\tau} + \delta]$. In the case of a Gaussian distribution, having mean value $\bar{\tau}$ and variance σ^2, measurements can be taken in the interval $[\bar{\tau} - k\sigma, \bar{\tau} + k\sigma]$, where k can be chosen depending on the accuracy

required for the approximation. In general, for an arbitrary distribution $g_1(\tau)$ a value $0 \leq r \leq 1$ can be set and τ_{min} and τ_{max} determined such that:

$$\int_{\tau_{min}}^{\tau_{max}} g_1(\tau)d\tau = r \tag{C.1}$$

The final dataset used to estimate p_{01} can be constructed in different ways. Let us denote as N_i the number of points in each dataset $\mathscr{D}_i$. The most straightforward one is to take a number of measurements N_i, such that the union dataset $\mathscr{D}_U = \bigcup_{i=1}^{L} \mathscr{D}_i$ directly approximates the outcome of $g_1(\tau)$, i.e. taking proportionally more measurement for more probable values τ_i.

An approach that gives more flexibility is to fix the number of measurements taken for each τ to a given $\bar{N}$, i.e. $N_i = \bar{N} \; \forall i$. In this case, the union dataset $\mathscr{D}_U$ will be composed of $N_U = L\bar{N}$ measurements extracted ideally from a uniform distribution in $[\tau_{min}, \tau_{max}]$. A procedure of statistical weighting, consisting in discarding a certain amount of measurement generated from selected τ_i, can then be used to rearrange the union dataset, to create the final one $\mathscr{D}_g$, approximating the outcomes of $g_1(\tau)$. This procedure is described in detail in the following subsection. The downside of using this approach is that the number of measurements performed will be higher than the number of measurements that will compose $\mathscr{D}_g$. However, there are different practical advantages. In practical scenarios in fact there is a limited control in selecting an exact value for the transmittance τ, so it may be difficult to have exactly equispaced values. This situation is shown in Fig.(C.1). Panels **A** and **B** report the plots of an equispaced sampling of the interval $[\tau_{min} = 0.5, \tau_{max} = 1]$ in the ideal and noisy cases respectively. When the values τ_i are affected by random noise their distribution is affected as well as shown in panels **C** and **D**. Panel **C** refers to the ideal case where the empirical distribution of τ is uniform as expected. In this ideal case, selecting before the measurements appropriate coefficients N_i could be an effective way to achieve a good approximation of the target transmittance distribution. In the noisy case however, whose distribution is reported in panel **D**, the distribution is not only non uniform but, in most cases, unknown before the experiment is performed. For this reason, supposing *a priori* a uniform distribution of the sampling and selecting the values N_i accordingly could result in a bad approximation of $g_1(\tau)$ by the empirical transmittance distribution.

Since the amount of noise varies depending on experimental conditions in cases in which the deviation from a uniform distribution of the empirical one is small, or in
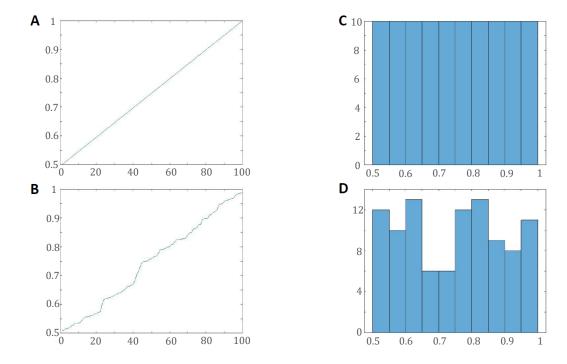
Fig. C.1 *Sampling of a transmittance interval.* Panel **A** reports an ideal equispaced sampling for the transmittance τ in the interval $[0.5, 1]$. Panel **B** reports the same plot in a noisy case, i.e. where the τ_i are taken equispaced but a value can be assigned up to a certain precision. Panels **C** and **D** are the histograms referring to panels **A** and **B** respectively. Image taken from the supplementary materials of Ref.[2].

cases in which it can be well characterized beforehand, acting on the coefficient N_i remains a viable solution. Another advantage of the a posteriori statistical weighting procedure is the fact that the procedure can be repeated with the same initial data with different target distributions, meaning that the probability of error over different distributions can be evaluated without having to perform the experiment again.

A formal description of the statistical weighting procedure used to create a dataset $\mathscr{D}_g$ approximating one generated by a probability density $g_1(\tau)$ will be given in the following.

Approximation using statistical weighting

Using the notation of the previous section, the initial experimental union dataset $\mathscr{D}_U$ will be composed by $N_U = L\bar{N}$ measurement results where each group of $\bar{N}$ measurements refers to a different transmittance τ_i, $0 \leq i \leq L$, where the values τ_i are a sampling of the interval $[\tau_{min}, \tau_{max}]$ selected according to Eq.(C.1).

The aim is to generate a dataset $\mathscr{D}_g$ to be a good approximation of an ideal dataset generated with measurement with objects having transmittance distributed as $g_1(\tau)$.

In order to do this let us define the map $\mathscr{H}_\rho^\Pi$, from a distribution on τ onto a distribution $G(\mathbf{X})$ of the experimental outcomes $\mathbf{X}$ (e.g. the distribution of photon pairs):

$$G(\mathbf{X}) = \mathscr{H}_\rho^\Pi[g_1(\tau)] \tag{C.2}$$

The subscript ρ and superscript Π denote the dependence of the map on the initial state used to probe τ and the measurement performed (for brevity they will be omitted from here on after).

We want to operate a transformation W on our initial union dataset, $\mathscr{D}_U \xrightarrow{W} \mathscr{D}_g$, such that the points in $\mathscr{D}_g$ are distributed as $\mathbf{X} \in \mathscr{D}_g \sim G(\mathbf{X})$. Let us introduce the weight vector $\mathbf{w} = [w_1, ..., w_L]$. $\mathscr{D}_U$ is composed of L subsets $\mathscr{D}_i$, each composed of $\bar{N}$ experimental points, $\mathscr{D}_i = [\mathbf{X}_i^1, ..., \mathbf{X}_i^{\bar{N}}]$. From $\mathbf{w}$ we define the transformation W_i acting on each $\mathscr{D}_i$, $\mathscr{D}_i \xrightarrow{W_i} \mathscr{D}_i'$, to extract a random fraction $N_i' = \lfloor w_i \bar{N} \rfloor$ of experimental points from $\mathscr{D}_i$. We can then define the transformation W as the one acting on

$\mathscr{D}_U$ and yielding the union set of all the $\mathscr{D}'_i$:

$$\mathscr{D}_U \xrightarrow{W} \mathscr{D}_g = \bigcup_{i=1}^{L} \mathscr{D}'_i \tag{C.3}$$

A successful approximation is therefore dependent on the optimization of the vector **w**.

To perform this optimization a figure of merit has to be introduced. To do so let us define a probability density function, $f(\tau)$, approximating the empirical distribution in $\mathscr{D}_U$. This can be done using a normalized histogram. The interval $[\tau_{max}, \tau_{min}]$ is divided into K sub-intervals with equal width, where K is selected via an adequate algorithm. A categories vector, $\tau^C = [\tau_1^C, ..., \tau_K^C]$, is defined taking τ_i^C as the middle point between the edges of the i-th sub-interval. The empirical distribution $f(\tau)$ is then defined as:

$$f(\tau) = \sum_{i=1}^{K} \frac{c_i}{2lL} \Theta_i(\tau) \tag{C.4}$$

where c_i are the number of occurrences on the i-th sub-interval, $\Theta_i(\tau)$ is a step, L is the number of different values of τ and l is the half width of each sub-interval:

$$\Theta_i(\tau) = \begin{cases} 1 & \tau_i - l \leq \tau \leq \tau_i + l \\ 0 & \text{otherwise} \end{cases} \tag{C.5}$$

$$l = \frac{\tau_{max} - \tau_{min}}{2K} \tag{C.6}$$

Using Eq.(C.2) the probability distribution in Eq.(C.4) defines a distribution for the experimental outcomes, $F(\mathbf{X})$:

$$F(\mathbf{X}) = \mathscr{H}[f(\tau)] \tag{C.7}$$

By construction W transforms $F(X)$ in $F'(X) \sim G(X)$ and since the map $\mathscr{H}$ does not depend on the probability distribution, the effect of W is equivalent to transforming $f(\tau)$ in $f'(\tau) \sim g_1(\tau)$:

$$\mathscr{H}[f(\tau)] = F(\mathbf{X}) \xrightarrow{W} \mathscr{H}[f'(\tau)] = F'(\mathbf{X}) \sim G(\mathbf{X}) = \mathscr{H}[g_1(t)]$$
$$f(\tau) \xrightarrow{W} f'(\tau) \sim g_1(\tau) \tag{C.8}$$

The weight vector $\mathbf{w}$ can be optimized by introducing the parameterized function $f_{\mathbf{w}}(\tau)$:

$$f_{\mathbf{w}}(\tau) = \sum_{i=1}^{K} \frac{c_i w_i}{2lL} \Theta_i(\tau) \tag{C.9}$$

depending on $\mathbf{w} = [w_1, ..., w_K]$ under the normalization constraint:

$$\int f_{\mathbf{w}}(\tau)d\tau = 1 \longrightarrow \sum_{i=1}^{K} \frac{c_i w_i}{L} = 1 \tag{C.10}$$

Note how in this configuration, the transformation acts separately on each of the K sub interval rather than on each of the L values of τ measured, so that the vector $\mathbf{w}$ has dimensionality $K \leq L$.

Clearly the initial empirical distribution is $f(\tau) = f_{\mathbf{w}_0}$, where $\mathbf{w}_0 = [1, ..., 1]$ and the target function is $f'(\tau) = f_{\mathbf{w}'}$, where $\mathbf{w}'$ is the optimal weight defined to obtain the best approximation $f'(\tau) \sim g_1(\tau)$. Formally, let us define the objective function to be maximized:

$$T(\mathbf{w}) = \int \sqrt{g_1(\tau) f_{\mathbf{w}}(\tau)} dt \tag{C.11}$$

and the optimal vector $\mathbf{w}'$:

$$\mathbf{w}' = \text{argmax}_{\mathbf{w}} T(\mathbf{w}) \tag{C.12}$$

$T(\mathbf{w})$ is the Bhattacharyya coefficient (BC) between the distribution $f_{\mathbf{w}}(\tau)$ and $g_1(\tau)$, a measure of their similarity, and ranges between 0 and 1. $T(\mathbf{w}')$ gives a quantitative measure of how close the experimental dataset can be arranged to a dataset produced by the distribution $g_1(\tau)$. A threshold value $0 \leq T_{th} \leq 1$ can be defined such that the approximation is accepted if $T(\mathbf{w}') \geq T_{th}$ and rejected otherwise. Some other constraint must be imposed before performing the optimization, namely:

$$0 \leq w_i \leq 1 \qquad \forall i \tag{C.13}$$

The first inequality follows from the fact that the statistical weight has to be non negative and the second one from the fact that the statistical weight represents the fraction of data that are selected from each dataset and since the data are considered fixed, i.e. new data cannot be added after the initial dataset has been created, all w_i must be less than 1.

The fixed data condition in Eq.(C.13) is not compatible with the normalization constraint in Eq.(C.10), but this issue is easily solved taking into account the number of experimental points in each dataset in the formulation of our problem.

The number of experimental data in the original union dataset is $N_{\mathscr{U}} = L\bar{N}$. Since the transformation W consists exclusively in discarding data, the number of points in the final dataset will be $N_T \leq N_{\mathscr{U}}$. It is clear that N_T will influence the variance of the estimated gain, so it makes sense to fix it before the optimization process rather then leaving it as a free parameter. We can then redefine $f_{\mathbf{w}}(\tau)$ as:

$$f_{\mathbf{w}}(\tau) = \sum_{i=1}^{K} \frac{N_{\mathscr{U}}}{N_T} \frac{c_i w_i}{2lL} \Theta_i(\tau) \qquad (\text{C.14})$$

This yields the new normalization condition:

$$\sum_{i=1}^{K} \frac{N_{\mathscr{U}}}{N_T} \frac{c_i w_i}{L} = 1 \qquad (\text{C.15})$$

The new definition for $f_{\mathbf{w}}(\tau)$ shifts the normalization condition in Eq.(C.10), imposed on the number L of different τ, to the new one in Eq.(C.15) where it is imposed on the number of experimental data in the final dataset. With this new definition, the initial distribution is $f(\tau) = f_{\mathbf{w}_0}(\tau)$, where in this case $\mathbf{w}_0 = [w_0, ..., w_0]$ and $w_0 = N_T/N_{\mathscr{U}}$. In this formulation, $\mathbf{w}$ is exactly a statistical weight vector, i.e. defines the fraction of experimental data taken from each sub interval dataset, and the optimal vector $\mathbf{w}'$, can be found maximizing $T(\mathbf{w})$ under the fixed data constraint of Eq.(C.13) and the normalization one in Eq.(C.15), that are now compatible conditions. The optimization of eq.(C.11) under the $K+1$ constraint described can be solved, in general, by numerical methods. This procedure, will yield the optimal value $\mathbf{w}'$. The choice of the sub-intervals number K could influence the outcome of the procedure, but we decided not to include it in the optimization process to avoid unnecessary numerical complications, as good approximation can be reached in a wide variety of cases performing the binning before the optimization, using well known algorithms. The binning in our case was performed using Sturges rule.

Fig.(C.2) shows some examples of the approximation of a distribution from simulated experimental datasets. In the examples reported the target distribution is a Gaussian one with mean $\bar{\tau} = 0.65$ and standard deviation $\sigma = 0.1$ and is reported in each panel as a red line. In the first row, Fig.(C.2).**A**, the experimental values of
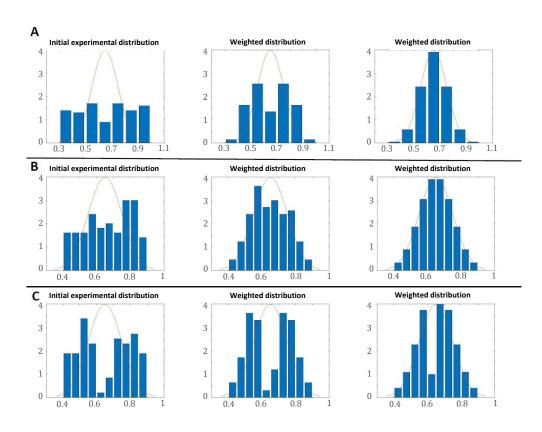
Fig. C.2 *Approximating a distribution with different initial dataset.* Row **A** shows the initial dataset and the best approximation for different numbers of experimental data taken, $\bar{N} = 150$ and $\bar{N} = 500$. Row **B** reports the same situation when the initial dataset is taken in a smaller interval. Row **C** considers a "faulty" initial dataset, having few measurements around the peak of the target distribution. The image is taken from the supplementary materials of Ref.[2].

τ are taken from a uniform distribution in $[\bar{\tau} - 3.5\sigma, \bar{\tau} + 3.5\sigma]$, meaning that τ_{min} and τ_{max} are selected imposing $r = 0.9995$ in Eq.(C.1). Starting from the left, the first panel shows the initial experimental distribution. The second and third panels show the best approximation in the situation in which are considered measurement with $L = 100$ different values of τ and, for each τ, the experimental points taken are $\bar{N} = 150$ and $\bar{N} = 500$ for the second and third panel respectively. The number of points of the final dataset is fixed to $N_T = 10000$ in both cases. If not enough points can be discarded, as it is the case in the second panel, a good approximation is difficult to reach. In the third panel a lot of experimental points are taken and, as can be expected, a good approximation can be reached for most of the distributions . In Fig.(C.2).**B** the same scenario considered in row **A** is reported but considering an initial dataset taken in a smaller interval, such that $r = 0.997$ (half width of $[\tau_{min}, \tau_{max}]$ equal 3σ). Since the datasets are randomly generated a direct comparison may be unfair. Nonetheless, it can seen that reducing the interval may be useful since more data will be taken at the peak of the target distribution. In the second panel, referring again to the situation of $\bar{N} = 150$ we can see a better approximation with respect to the same panel in row **A**. Finally, row **C** reports the results in the same situation but in case of a faulty initial dataset, i.e. one having very few values around the peak of the target distribution. In this scenario even with $\bar{N} = 500$ experimental data for each τ (third panel) a good approximation cannot be reached, suggesting that in those cases it may be worth performing the measurements again.