Mémoire

Auteur : Lejeune, Théo
Promoteur(s) : Damanet, François; Martin, John
Faculté : Faculté des Sciences
Diplôme : Master en sciences physiques, à finalité approfondie
Année académique : 2023-2024
URI/URL : http://hdl.handle.net/2268.2/20376

Faculty of Science
Department of Physics

Attacks and decoherence in quantum cryptography

Author: Théo LEJEUNE

Supervisor: François DAMANET
Co-supervisor: John MARTIN

Master's thesis submitted in the partial fulfillment of the requirements for the Master's degree in Physical Science.

Academic year 2023-2024

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisor, Mr. F. Damanet, for his invaluable guidance, continuous support, permanent disponibility, and encouragement throughout my master's study and research. His profound knowledge and expertise were instrumental in the completion of this thesis. I also wish to thank Mr. J. Martin and Mr. J. Denis for their guidance and presence through this last year.

I would also like to extend my sincere thanks to Mr. E. Bousquet, Mr. T. Bastin, and Mr. E. Opsomer, for accepting to be members of my reading committee. I wish them an enjoyable reading.

I thank Ms. R. Wolf and Mr. O. Gühne from the Theoretical Quantum Optics department of Siegen university, for useful discussions about this work, and more generally, about quantum cryptography.

Finally, I would like to thank my family and friends for their continuous support and help throughout the past five academic years.

Contents

Introduction

In the last decades, the demand for fast, secure, and reliable data connections has significantly increased. To meet this demand, it is essential to enhance the computational power of network systems through high-performance technologies. Quantum computing is one such technology, showing a clear potential to outperform current classical computing systems. Quantum computing combines principles from computer science and quantum mechanics to create more advanced systems capable of handling complex computational tasks. The basic operational unit in QC is the quantum bit (qubit), which utilizes fundamental concepts like superposition and entanglement, contributing to its high performance. Quantum computing-assisted communications have therefore been extensively studied and developed in recent years, and hold great promise for improving communications and security in today's networks [1, 2].

However, challenges emerge from the rapid development of quantum computing, notably related to some asymmetric cryptographic algorithms such as RSA (Rivest-Shamir-Adleman) [3], a public key cryptosystem still used in many secure data transmissions to this day. Standard encryption techniques such as RSA, while very powerful, could be broken through Shor's algorithm, a quantum algorithm factoring large integers exponentially faster than the best-known classical algorithms [4]. While current quantum computing technology is still far from being enough advanced to break RSA, this motivated the elaboration of new encryption techniques, notably based on quantum mechanical properties. This new branch of cryptography, called Quantum Cryptography, offers unconditional security independent of the current computing power and gives rise to the use of quantum key distribution (QKD). Its aim is to implement the exchange of private key bits over a public insecure channel between two parties. These key bits can then be used to implement a classical private key cryptosystem, thus enabling the two parties to securely communicate. Quantum Key Distribution is the most famous category of quantum cryptography protocols as it offers a theoretically secure solution to the challenge of private key exchange between communicating users [5]. The BB84 is the first QKD protocol, proposed by Charles Bennett and Gilles Brassard in 1984 [6], and uses linearly polarized photons travelling in an optical fiber.

The advantage of quantum over classical cryptography is its exploitation of quantum mechanical properties. Among these properties are the Heisenberg uncertainty principle, the no-cloning theorem which stipulates it is impossible to create an independent and identical copy of an arbitrary unknown quantum state, and most importantly the perturbative nature of measurement. Indeed, the latter implies that any spy acting on a communication channel will influence the state of the qubits traveling inside, because of the collapse of the wavefunction. Therefore a measurement would modify the information contained in the qubits, making it more easily detectable than in classical communication protocols.
The effects on QKD protocols of measurements have been studied in the past, although not extensively [7, 8, 9, 10].

The aim of this Master's thesis is to study the security of the BB84 protocol when implemented in realistic conditions and against advanced type of attacks. We analyze the impact on the photons, used as the qubits in the quantum communication channel, of intrinsic dissipation in the channel and of different types of measurement, such as projective or weak measurements, that can be viewed as attacks from a third party. In addition, we investigate the implementation of quantum feedback based on the spy measurement outcomes that they could use to cover their tracks, i.e., decrease the influence of their measurements on the qubits. A representative sketch is displayed in Figure 1.

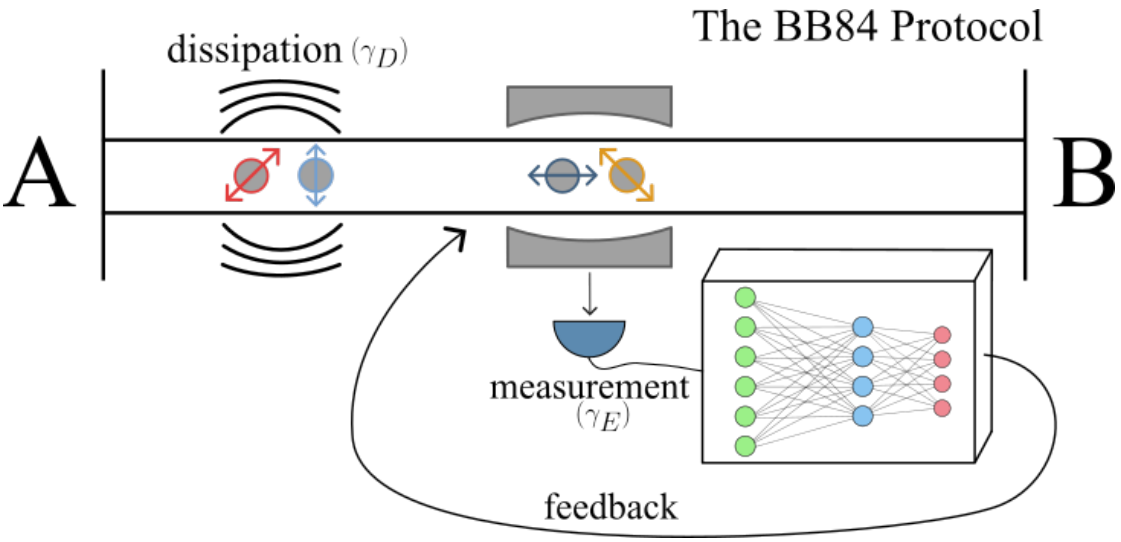


Figure 1: Sketch representation of this Master's thesis aim. Alice (A) sends linearly polarized photons to Bob (B) via an optical fiber, while travelling they are subject to dissipation with a rate γ_D and to weak measurements via monitoring of a cavity output field with a rate γ_E. The output of this measurement is treated using a neural network to determine the initial state, and used to apply feedback on the photons.

In Chaper 1 we start by recalling the fundamental concepts of quantum mechanics needed to describe open quantum systems. Then we derive the different equations needed, such as the Lindblad master equation to describe the interaction of a quantum system (e.g., photons) with its environment (e.g., optical fiber), the master equations conditioned on measurement, such as the homodyne master equation, and finally the homodyne mediated feedback master equation.

In chapter 2 we explain the BB84 protocol, which is the most common QKD protocol, step by step, considering first an idealized case without noise nor eavesdropping, then a more realistic case with their presence.

In Chapter 3, we first evaluate the impact of noise and third party attacks, in the form of projective measurements, independently. Then we evaluate their joint impact on the protocol probability of success. In the fourth section, we perform an in-depth examination of how weak measurements could be used by the spy to decrease its impact on the photons while still extracting information. We finish this Chapter by evaluating the impact on single photons of such weak measurements through the trace distance metric, which measures how different two quantum states are from each other.

In Chapter 4 we start by exploring standard tomography to reconstruct the state of photons from the result of a single homodyne measurement on a photon traveling in the optical fiber.

We also motivate the use of neural networks for this task, given their high capacities in pattern recognition, data analysis, and complex problem-solving. We then introduce the theoretical background needed to understand and construct a neural network, which we use to perform initial state tomography (i.e., recovering the initial state of the qubits used in the protocol) from the results of homodyne measurements.

In Chapter 5, we introduce the concept of information gain to quantify the information a given measurement extracts on a given system. The second Section of this Chapter is dedicated to finding the trade-off between the amount of information extracted from the qubits measurement and the impact that this measurement has on their state.

In Chapter 6 we investigate the introduction, in the system, of quantum feedback by the spy to cover its tracks.

Chapter 1

Open quantum systems

Since we will be interested in looking at the influence of dissipative mechanisms and measurements on quantum systems, it is necessary to introduce all the necessary tools to the description of open quantum systems. In this Chapter, we first recall the concepts of pure and mixed states, the density operator formalism, and quantum mechanics postulates in this formalism. We then derive three master equations, respectively the GKSL (or Lindblad) one to describe quantum systems interacting with their environment, the homodyne conditioned stochastic master equation for systems subjected to weak measurements, and finally the homodyne mediated feedback master equation to account for the additional introduction of quantum feedback.

1.1 Pure and mixed states

Let us consider a quantum system S, e.g., a two-level atom prepared in an excited state $|e\rangle$. In this example, at $t = 0$, we know that this system has a probability of one to be in the state $|e\rangle$, and is thus said to be in a pure state. However, at a later time $t \neq 0$, there is a non-zero probability the atom has emitted a photon to reach its ground state $|g\rangle$, and we thus not know exactly in which state the atom is. In other words, the state of the whole system, made of the atom and the surrounding electromagnetic field can be written as:

$$|\psi_{SE}\rangle = C_1(t)\,|e,0\rangle + C_2(t)\,|g,1\rangle\,, \tag{1.1}$$

where $|e,0\rangle = |e\rangle \otimes |0\rangle$ denotes the state of the whole system when the atom is in its excited state and no photon has been emitted, $|g,1\rangle$ is the state when the atom has emitted a photon and is in its ground state, $C_1(t)$ and $C_2(t)$ are time-dependant complex coefficients that determine the probabilities of the system to be in each state.

From Eq. (1.1), it is clear that one cannot identify exactly the state of the atom without knowing the state of the electromagnetic field (they are entangled). In this case, the atom is said to be in a mixed state or in a statistical mixture of states, which is more generally defined as

$$\sum_i p_i\,|\psi_i\rangle\,\langle\psi_i|\,, \tag{1.2}$$

where $0 \leq p_i \leq 1$ is the probability to be in the pure state $|\psi_i\rangle$. Therefore we have $\sum_i p_i = 1$.

Usually, the coupling of a system to an environment is likely to turn an initially pure state into a mixed state. In Sec. 1.4, we will show how to derive explicitly Eq. (1.2) for the example given above, and what are the explicit expressions for the probabilities p_i involved.

The fact that the state of an open system is generally not pure motivates the introduction of a more general tool to describe open systems dynamics: the density operator, as elaborated on below.

1.2 Density operator

A noteworthy mathematical instrument employed universally is the density operator, also called density matrix or statistical operator. Usually noted $\hat{\rho}$, the density operator is mainly used to describe the state of open quantum systems. For a system in a pure state, the density operator is defined as the projector on the state of the system considered:

$$\hat{\rho}(t) = |\psi(t)\rangle \langle\psi(t)| . \tag{1.3}$$

Given Eqs. (1.2) and (1.3), for a system in a mixed state

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i, \tag{1.4}$$

where p_i is the probability that the system is in the pure state $\hat{\rho}_i = |\psi_i\rangle \langle\psi_i|$. The following fundamental properties of $\hat{\rho}$ are immediately inferred from its definition (1.2) [11]:

- $\hat{\rho}$ is hermitian: $\hat{\rho}^\dagger = \hat{\rho}$
- $\hat{\rho}$ has unit trace: $\mathrm{Tr}(\hat{\rho}) = 1$
- $\hat{\rho}$ is a positive operator: $\langle\phi| \hat{\rho} |\phi\rangle \geq 0, \quad \forall |\phi\rangle \in \mathcal{H}$, the hilbert space of the system.
- For a pure state, $\hat{\rho}$ is idempotent: $\hat{\rho}^2 = \hat{\rho}$

We can define the purity of a state $\hat{\rho}$ as $\mathrm{Tr}(\hat{\rho}^2)$, which is equal to one for a pure state and smaller than one for a mixed state. Another important property of the density operator is its use to compute the expectation value of any other operator:

$$\left\langle \hat{A} \right\rangle_\psi = \mathrm{Tr}(\hat{\rho}\hat{A}). \tag{1.5}$$

1.3 Quantum mechanics postulates

We can reformulate the postulates of quantum mechanics in terms of the density operator. The two first postulates characterize the state of a system and its temporal evolution.

Postulate 1

The state of an isolated physical system is represented, at a fixed time t, by a density operator $\hat{\rho}(t)$, belonging to a Hilbert space $\mathcal{H}$ defined over the field of complexes, called the state space. This operator must be positive and have a unit trace.

Postulate 2

The time evolution of the density operator $\hat{\rho}(t)$ of an isolated physical system is described by the Liouville-von Neumann equation

$$\dot{\hat{\rho}}(t) = \frac{-i}{\hbar}[\hat{H}, \hat{\rho}(t)], \tag{1.6}$$

where $\hbar$ is the reduced Planck constant and $\hat{H}$ is the Hamiltonian of the considered system.

The third postulate characterizes measurements and their impact on the system.

Postulate 3

Measurements on a quantum system are described by a set of operators $\{\hat{M}_k\}$ called measurement operators acting in the state space $\mathcal{H}$, where k denotes the possible measurement outcomes. They satisfy the completeness relation

$$\sum_k \hat{M}_k^\dagger \hat{M}_k = \hat{1}. \tag{1.7}$$

If the state of the system is described by the density operator $\hat{\rho}$ just before the measurement, then the probability that the outcome of a physical quantity (represented by $\hat{A}$) measurement is one of its discrete eigenvalues k is

$$\mathbb{P}(k) = \mathrm{Tr}\left(\hat{\rho}\hat{M}_k^\dagger \hat{M}_k\right). \tag{1.8}$$

If the outcome is k, the state of the system after the measurement is

$$\hat{\rho}' = \frac{\hat{M}_k \hat{\rho} \hat{M}_k^\dagger}{Tr\left(\hat{\rho}\hat{M}_k^\dagger \hat{M}_k\right)}. \tag{1.9}$$

The last postulate describes the state of a composite system, i.e. a system composed of several sub-systems.

Postulate 4

The state space of a composite system of N physical systems is the tensor product of the state spaces of each sub-systems

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_N. \tag{1.10}$$

If the ith sub-system is in the state $\hat{\rho}_i$ $(i = 1, ..., N)$, the state of the composite system is given by the density operator

$$\hat{\rho} = \hat{\rho}_1 \otimes \hat{\rho}_2 \otimes ... \otimes \hat{\rho}_N. \tag{1.11}$$

The elements of the density operator matrix can be interpreted as physical quantities. Indeed, the diagonal elements are the amplitudes and the non-diagonal ones are the coherences. The latter will go to zero if the system considered is interacting with its environment, this process is called phase damping.

1.4 The Lindblad master equation

Since the physical systems considered in this work are single photons travelling through an optical fiber and since optical fibers, such as any other real system, are not perfectly insulating, the photons will interact with their environment by being diffused or polarized. To handle such open systems we need a new formalism, which is to replace the state vector $|\psi(t)\rangle$ used previously by the density operator $\hat{\rho}(t)$ defined in Section 1.2.

In this open quantum system formalism we consider an enlarged system, composed of the open system of interest S and its environment E, to be a closed quantum system whose state is described by a density operator $\hat{\rho}_{SE}(t)$ as explained in Figure 1.1. Following postulate 4, this state lives in a Hilbert space $\mathcal{H}_{SE}$ which is the tensor product of the Hilbert space of the environment and of the system : $\mathcal{H}_{SE} = \mathcal{H}_S \otimes \mathcal{H}_E$.

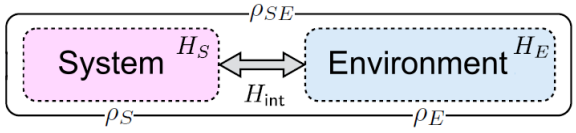


Figure 1.1: Schematic representation of the open quantum system formalism [12]. H_S, H_E and H_{int} are respectively the Hilbert spaces of the system, its environment, and the interaction between them. ρ_S, ρ_E and ρ_{SE} are respectively the density operators representing the states of the system, the environment and the total system formed by both of them.

Going back to the example of Section 1.1, the system considered is the two-level atom and the environment is a bosonic mode of the emitted (or not) photon. The initial state describing the system and environment is thus $|\psi_{SE}\rangle = |e,0\rangle$. At time t, this state is mixed and we can describe it as (1.1) but also using the density operator

$$\hat{\rho}_{SE} = |\psi_{SE}\rangle\langle\psi_{SE}| = |C_1(t)|^2 |e,0\rangle\langle e,0| + |C_2(t)|^2 |g,1\rangle\langle g,1|, \tag{1.12}$$

where $|C_1(t)|^2$ and $|C_2(t)|^2$ are the probabilities of the excited state and the ground state respectively. To obtain the state of the system starting from the state of the System + Environment, we must perform the partial trace operation :

$$\hat{\rho}_S = \mathrm{Tr}_E(\hat{\rho}_{SE}) = \sum_j \left\langle \phi_E^{(j)} \right| \hat{\rho}_{SE} \left| \phi_E^{(j)} \right\rangle, \tag{1.13}$$

where $\{\left|\phi_E^{(j)}\right\rangle\}$ is an orthonormal basis of $\mathcal{H}_E$, the Hilbert space of the environment. In the example of the atom and surrounding electromagnetic field, this basis could be a Fock basis. In

this example, the state of the atom is

$$\begin{aligned}\hat{\rho}_S &= \mathrm{Tr}_E(\hat{\rho}_{SE}) \\ &= |C_1(t)|^2 |e\rangle \langle e| + |C_2(t)|^2 |g\rangle \langle g| \\ &= \sum_i p_i |i\rangle \langle i| ,\end{aligned}$$

which corresponds to the general expression of Eq. (1.4).

As stated in postulate 2, the evolution of the total system is given by the von Neumann equation [Eq. (1.6)] but we need a way to describe the evolution of the system alone.
It was shown in 1976 by Lindblad [13] and Gorini *et al.* [14] that the generator of the quantum dynamics for a Markovian equation $\dot{\hat{\rho}} = \mathcal{L}\hat{\rho}(t)$ must be of the form:

$$\mathcal{L}\hat{\rho} = -i[\hat{H}, \hat{\rho}] + \sum_{k=1}^{K} \gamma_k \mathcal{D}[\hat{L}_k]\hat{\rho}, \tag{1.14}$$

where the superoperator $\mathcal{L}$ is called the Lindbladian or the Liouvillian, $\hat{H}$ is hermitian and contains a rate ω (e.g., $\hat{H} = \omega\hat{\sigma}$), $\{\hat{L}_j\}$ are arbitrary operators and $\mathcal{D}[\hat{L}]$ is a superoperator defined as:

$$\boxed{\mathcal{D}[\hat{L}]\hat{\rho} = \hat{L}\hat{\rho}\hat{L}^\dagger - \frac{1}{2}\left(\hat{L}^\dagger\hat{L}\hat{\rho} + \hat{\rho}\hat{L}^\dagger\hat{L}\right).} \tag{1.15}$$

The form of (1.14) is called the Lindblad or Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) form, and we must use an equation of this form to perform our analysis. Deriving this equation is beyond the scope of this master thesis, but a full demonstration can be found in [15]. This equation is derived from the von Neumann equation (1.6), where the Hamiltonian is separated into three parts: one for the system, one for the environment, and one for the interactions between the two.

The total system is characterized by three typical timescales: τ_B, τ_S and τ_R referring respectively to the typical timescale of the bath, the system and the relaxation. τ_R refers to the characteristic timescale over which the system approaches its steady state or equilibrium due to interaction with the environment. The bath correlation time τ_B refers to the characteristic timescale over which the bath loses information coming from the system. For a finite size system, τ_S is the typical timescale of the system dynamics, for example if the Hamiltonian describing the system dynamics is $\hat{H} = \omega_0\hat{\sigma}_z$, then $\tau_S = 1/\omega_0$.

We first assume a weak interaction between the system and the environment, which is also called weak system-bath coupling as the environment is also called bath. This assumption is called the Born approximation. We also assume that the largeness of the environment (the closeness of its energy levels) ensures that from one moment to another the system interacts with different parts of the environment, this is called the Markov approximation (see Figure 1.2) [16]. These two approximations together form the Born-Markov approximation, which can be interpreted as the bath having no memory effect of its interaction with the system, i.e., it returns to its original state almost instantly after the interaction while the system does not. This approximation is valid if the timescale of the bath is small compared to the relaxation timescale, i.e., $\tau_B \ll \tau_R$.

Then, we assume the system and environment are initially in a separable state:

$$\hat{\rho}_{SE}(0) = \hat{\rho}_S(0) \otimes \hat{\rho}_E(0),$$

which means there are no correlations between them. Over time, due to the interaction Hamiltonian, some correlations are expected to happen between the system and the environment. However, we may assume the typical timescales of correlation and relaxation of the environment are much smaller than the system timescale (because coupling is weak) and are thus negligible. Under this approximation, the environment state can be considered as always thermal and decoupled from the system state:

$$\hat{\rho}_{SE}(t) = \hat{\rho}_S(t) \otimes \hat{\rho}_E(0).$$

We now apply the rotating wave approximation (RWA, see Figure 1.2) which consists in considering some terms as oscillating much faster than the typical timescale of the system evolution, thus averaging to 0 over that period. This approximation is valid if the typical timescale of the system is much smaller than the relaxation one, i.e., $\tau_S \ll \tau_R$.

After these assumptions and approximations, we finally obtain the Lindblad-Gorini-Kossakowski-Sudarshan Master equation (or Lindblad equation) :

$$\boxed{\dot{\hat{\rho}} = -i[\hat{H} + \hat{H}_{LS}, \hat{\rho}(t)] + \sum_i \gamma_i \left(\hat{L}_i \hat{\rho}(t) \hat{L}_i^\dagger - \frac{1}{2}\{\hat{L}_i^\dagger \hat{L}_i, \hat{\rho}(t)\} \right) \equiv \mathcal{L}\hat{\rho}(t),} \tag{1.16}$$

where the operators $\hat{L}_i$ are usually referred to as jump operators. $\hat{H}_{LS}$ is the Lamb Shift Hamiltonian which renormalizes the system energy levels due to the interaction with the environment. The Lindblad master equation (1.16) will be our model to describe dissipative processes acting on the system we consider, i.e., the photons travelling through the optical fiber.

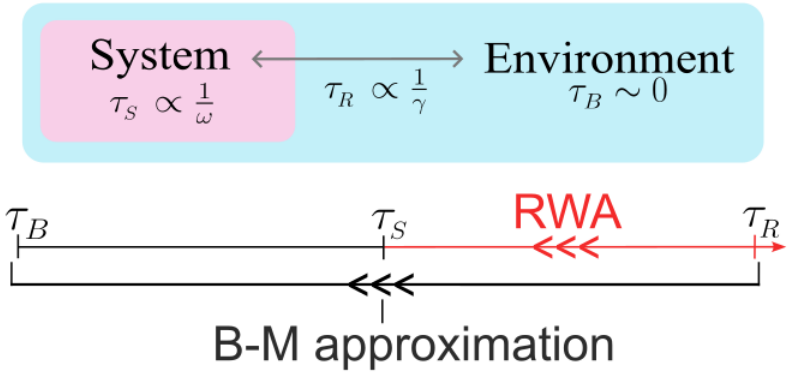


Figure 1.2: Hierarchy between the typical timescales τ_B, τ_S and τ_R of the bath, the system and the relaxation respectively. Here, ω is the natural frequency of the system and γ is the typical damping rate of the system due to its interaction with its environment. The Born-Markov and Rotating Wave approximations are valid respectively for $\tau_B \ll \tau_R$ and $\tau_S \ll \tau_R$.

1.5 Quantum trajectories

In this section, we generalize the unitary evolution of an isolated quantum system by incorporating measurements. Most results of this Section come from [16].

1.5.1 Quantum Jumps

The evolution of an isolated quantum system in the absence of measurement is Markovian and continuous:

$$\dot{\hat{\rho}}(t) = -i\left[\hat{H}(t), \hat{\rho}(t)\right] = \text{finite}. \tag{1.17}$$

In this equation, and more generally in this master thesis, we consider the Hamiltonian $\hat{H}$ to be containing the rate ω in a similar way to in Eq. (1.14). We require the measurement time T to be infinitesimal in order to obtain a differential equation for $\hat{\rho}(t)$, in this case the system is said to be monitored. Using Eq. (1.9), assuming that the measurements are performed but the results r are ignored and averaging over all possible results, the unconditioned evolution of the state is

$$\hat{\rho}(t+dt) = \sum_r \mathbb{P}_r \hat{\rho}_r(t+dt) = \sum_r \mathcal{J}[\hat{M}_r]\hat{\rho}(t), \tag{1.18}$$

where $\mathcal{J}[\hat{M}_r]\hat{\rho}(t) = \hat{M}_r\hat{\rho}(t)\hat{M}_r^\dagger$, $\mathbb{P}_r$ is given by Eq. (1.8) and the index r corresponds to the possible results of the measurement. Here the denomination 'unconditioned' refers to the state obtained by averaging over the random measurement results which condition the system [16]. As $\hat{\rho}(t+dt)$ is infinitesimally different from $\hat{\rho}(t)$, a plausible assumption would be to consider just one r (e.g., $r=0$) and set the measurement operator as

$$\hat{M}_0(dt) = \hat{1} - \left(\frac{\hat{R}}{2} + i\hat{H}\right)dt, \tag{1.19}$$

where $\hat{R}$ and $\hat{H}$ are Hermitian operators. However, this measurement operator alone does not satisfy the completeness condition that is Eq. (1.7). We have, to order dt,

$$\hat{M}_0^\dagger(dt)\hat{M}_0(dt) = \hat{1} - \hat{R}dt \neq \hat{1}. \tag{1.20}$$

Indeed, a measurement with only one possible result does not really constitute a measurement at all. For the completeness condition to be satisfiable, the measurement requires a second possible result and thus a second measurement operator. We thus consider the two possible outcomes 0 and 1, and define the measurement operators

$$\hat{M}_0(dt) = \hat{1} - \left(\frac{\hat{R}}{2} + i\hat{H}\right)dt \quad \text{and} \quad \hat{M}_1(dt) = \sqrt{\gamma_E dt}\hat{c}, \tag{1.21}$$

where $\hat{c}$ is an arbitrary operator such that

$$\gamma_E \hat{c}^\dagger\hat{c} = \hat{R}. \tag{1.22}$$

Thus with $\hat{R}$ being positive, the measurement operators satisfy the completeness condition

$$\hat{M}_0^\dagger(dt)\hat{M}_0(dt) + \hat{M}_1^\dagger(dt)\hat{M}_1(dt) = \hat{1}. \tag{1.23}$$

The respective probabilities for the results $r=1$ and $r=0$ are

$$\mathbb{P}_1(dt) = \text{Tr}\left[\mathcal{J}[\hat{M}_1]\hat{\rho}\right] = \gamma_E \text{Tr}\left[\hat{c}^\dagger\hat{c}\hat{\rho}\right]dt, \tag{1.24}$$

$$\mathbb{P}_0(dt) = \text{Tr}\left[\mathcal{J}[\hat{M}_0]\hat{\rho}\right] = 1 - \gamma_E \text{Tr}\left[\hat{c}^\dagger\hat{c}\hat{\rho}\right]dt. \tag{1.25}$$

$\mathbb{P}_1(dt)$ is infinitesimal since $\hat{c}^\dagger\hat{c}$ is bounded, and thus $\mathbb{P}_0(dt)$ is almost equal to one.

Thus for almost every infinitesimal time intervals dt the measurement result is $r = 0$, and the system evolves infinitesimally. However at random times, having a rate $\frac{\mathbb{P}_1(dt)}{dt}$, the measurement result is $r = 1$ and thus the system will undergo a finite evolution (detection). This event is called a quantum jump but does not represent a physical event, but rather a sudden modification of the observer's knowledge about the system.

1.5.2 Homodyne detection

Homodyne detection is a technique used in quantum optics to analyze and extract information from quantum states of light. It consists in mixing a quantum signal with a reference beam, typically called a local oscillator, at a beamsplitter. This creates an interference pattern that contains information about the phase and amplitude of the quantum signal. By measuring this interference pattern using photodetectors (Figure 1.3), we can extract valuable information about the quantum state being studied. Homodyne detection is said to measure one quadrature of the system. The term *homodyne* signifies that the local oscillator is derived from the same source as the signal before the modulating process. For example, in a laser scattering measurement, the laser beam is split into two part, one is the local oscillator and the other is sent to the system to be probed. The scattered light is then mixed with the local oscillator on the detector [17]. This type of measurement is called *weak measurement* because it introduces less perturbations in the system than a direct measurement, as a result the measurement does not contain all information about the state of the system, but only about one of its quadratures. Since this type of measurement perturbs less the system it monitors, it could be used in attacks against the BB84 protocol. In this section, we derive a master equation describing a system continuously monitored via homodyne detection.

The master equation in the Lindblad form, where γ_E is the dissipation rate for the operator $\hat{c}$,

$$d\hat{\rho} = -idt[\hat{H}, \hat{\rho}] + \gamma_E dt\mathcal{D}[\hat{c}]\hat{\rho}, \tag{1.26}$$

is known to be invariant under shift of the jump operator $\hat{c}$, if a new term is added to the Hamiltonian [16]:

$$\hat{c} \to \hat{c} + \gamma; \qquad \hat{H} \to \hat{H} - i\frac{1}{2}(\gamma^*\hat{c} - \gamma\hat{c}^\dagger), \tag{1.27}$$

where γ is an arbitrary complex number. Under such transformation, the measurement operators from Eq. (1.21) become

$$\hat{M}_1(dt) = \sqrt{\gamma_E dt}(\hat{c} + \gamma), \tag{1.28}$$

$$\hat{M}_0(dt) = \hat{1} - dt\left[i\hat{H} + \frac{\sqrt{\gamma_E}}{2}(\hat{c}\gamma^* - \hat{c}^\dagger\gamma) + \frac{\gamma_E}{2}(\hat{c}^\dagger + \gamma^*)(\hat{c} + \gamma)\right]. \tag{1.29}$$

This type of transformation can be used to model homodyne detection, which is a type of weak measurement where the output field of the system considered (e.g. a cavity) is sent through one port of a beam-splitter of transmittance η while a strong coherent field (having the same frequency as the system dipole) is sent through its other port. Then both output ports are measured with photo detectors which outputs are subtracted to remove the contribution of the local oscillator, as illustrated in Figure 1.3. Let's suppose $\hat{b}$ is the operator for the field entering one port and $\hat{o}$ is the operator for the other port incident field. After going through the beam-splitter they become

$$\hat{b} \to \sqrt{\eta}\hat{b} + \sqrt{1-\eta}\hat{o}, \tag{1.30}$$

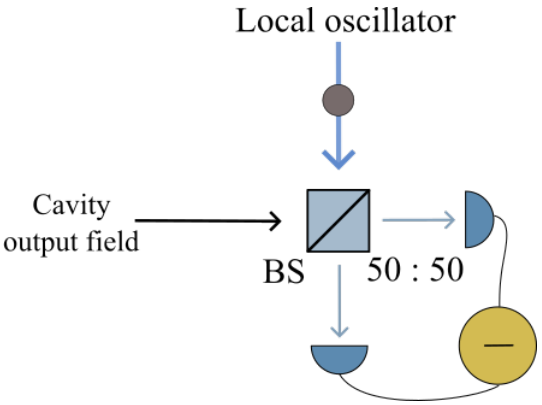


Figure 1.3: Balanced homodyne detection in the limit of an infinitely strong local oscillator. The field being emitted by the system (e.g. a cavity) is being fed to one input port of the beam-splitter while a strong coherent field (local oscillator) is injected in the other input port. The two output ports are then measured and substracted to obtain the homodyne photo current of Eq. (1.49).

$$\hat{o} \to \sqrt{\eta}\hat{o} - \sqrt{1-\eta}\hat{b}, \tag{1.31}$$

where η is an adimensional number representing the efficiency of the measurement, i.e. $\eta = 1$ for a perfect measurement. We can model a strong coherent field using

$$\hat{o} = \frac{\gamma}{\sqrt{1-\eta}} + \hat{\nu}, \tag{1.32}$$

where the first term represents the coherent amplitude of the local oscillator and the second part represents the quantum fluctuations around the expectation value ($\hat{\nu}$ is a continuum field).

Considering a measurement efficiency η close to one, the transformation (1.30) thus reduces to

$$\hat{b} \to \hat{b} + \gamma. \tag{1.33}$$

This transformation represents a displacement of the field. Supposing the coherent field γ is real, the homodyne detection measures the x quadrature of the system, which two quadratures are defined as

$$\hat{x} = \hat{c} + \hat{c}^\dagger \quad \text{and} \quad \hat{y} = -i(\hat{c} - \hat{c}^\dagger). \tag{1.34}$$

We define $N(t)$ as the number of photo detections up to time t, thus the stochastic increment $dN(t)$ follows

$$\mathrm{E}[dN(t)] = \left\langle \hat{M}_1(dt)\hat{M}_1^\dagger(dt) \right\rangle = \gamma_E dt \left\langle \psi(t) \right| \hat{c}^\dagger \hat{c} \left| \psi(t) \right\rangle . \tag{1.35}$$

We also define the rate of photo detection at the detector as

$$\mathrm{E}\left[\frac{dN(t)}{dt}\right] = \mathrm{Tr}[(\gamma^2 + \gamma\sqrt{\gamma_E}\hat{x} + \gamma_E \hat{c}^\dagger \hat{c})\hat{\rho}_I(t)], \tag{1.36}$$

where $\hat{\rho}_I(t)$ is the state of the system conditioned on the photo current $I(t) = \frac{dN(t)}{dt}$. If γ is much larger than the expectation value $\langle \hat{c}^\dagger \hat{c} \rangle$, this rate is composed of a large term, a term proportional to $\hat{x}$, and a small term.

The measurement operators $\hat{M}_0(dt)$ and $\hat{M}_1(dt)$ yield the stochastic master equation for the conditioned state

$$\mathrm{d}\hat{\rho}_I(t) = \left\{ dN(t)\mathcal{G}[\hat{c} + \gamma] + dt\mathcal{H}\left[-i\hat{H} - \gamma\sqrt{\gamma_E}\hat{c} - \frac{\gamma_E}{2}\hat{c}^\dagger \hat{c} \right] \right\} \hat{\rho}_I(t), \tag{1.37}$$

where

$$\boxed{\mathcal{H}[\hat{c}]\hat{\rho} = \hat{c}\hat{\rho} + \hat{\rho}\hat{c}^\dagger - \mathrm{Tr}\left[\hat{c}\hat{\rho} + \hat{\rho}\hat{c}^\dagger\right]\hat{\rho}}, \tag{1.38}$$

and

$$\mathcal{G}[\hat{c}]\hat{\rho} = \frac{\hat{c}\hat{\rho}\hat{c}^\dagger}{\mathrm{Tr}\left[\hat{c}\hat{\rho}\hat{c}^\dagger\right]} - \hat{\rho}. \tag{1.39}$$

When the local oscillator amplitudes tends to infinity, the rate of photo detections also tends to infinity but the effect of each detection on the system tends to zero. Indeed, the local oscillator covers almost entirely the other field and we can thus derive a continuous approximation of the photo current, yielding a continuous evolution equation for the system [18, 16].

Let us consider a time interval $[t, t+\delta t]$ with $\delta t = O(\gamma^{-3/2})$ in order to have a very large number of detections $\delta N \sim \gamma^2 \delta t$ and a very small change in the system $O(\delta t) = O(\gamma^{-3/2})$. Therefore the mean number of detections in this time is

$$\begin{aligned}\mu &= \mathrm{Tr}\left[\left(\gamma^2 + \gamma\sqrt{\gamma_E}\hat{x} + \gamma_E\hat{c}^\dagger\hat{c}\right)\left\{\hat{\rho}_I(t) + O(\gamma^{-3/2})\right\}\right]\delta t \\ &= [\gamma^2 + \gamma\sqrt{\gamma_E}\left\langle\hat{x}\right\rangle_I(t) + O(\gamma^{1/2})]\delta t.\end{aligned} \tag{1.40}$$

Since the number of counts δN is very large, its statistics are consistent with those of a Gaussian random variable of mean μ [Eq. (1.40)] and variance

$$\sigma^2 = [\gamma^2 + O(\gamma^{3/2})]\delta t. \tag{1.41}$$

Thus we can write δN as

$$\delta N = \gamma^2\delta t\left[1 + \frac{\sqrt{\gamma_E}\left\langle\hat{x}\right\rangle_I(t)}{\gamma}\right] + \gamma dW, \tag{1.42}$$

where δW is a Wiener increment satisfying

$$\mathrm{E}[\delta W] = 0, \tag{1.43}$$

$$\mathrm{E}[\delta W^2] = \delta t. \tag{1.44}$$

From the beginning, δt has been considered to be very small, and we can expand Eq. (1.37) in powers of γ^{-1}, yielding

$$\begin{aligned}\delta\hat{\rho}_I(t) =& \delta N(t)\left(\frac{\sqrt{\gamma_E}\mathcal{H}[\hat{c}]}{\gamma} + \gamma_E\frac{\left\langle\hat{c}^\dagger\hat{c}\right\rangle_I(t)\mathcal{G}[\hat{c}] - \left\langle\hat{x}\right\rangle_I(t)\mathcal{H}[\hat{c}]}{\gamma^2} + O\left(\gamma^{-3}\right)\right)\hat{\rho}_I(t) \\ &+ \delta t\mathcal{H}\left[-\mathrm{i}\hat{H} - \gamma\sqrt{\gamma_E}\hat{c} - \frac{\gamma_E}{2}\hat{c}^\dagger\hat{c}\right]\hat{\rho}_I(t).\end{aligned} \tag{1.45}$$

We can substitute the expression of δN from Eq. (1.42) into Eq. (1.45), keeping only the lowest-order terms in $\gamma^{-1/2}$ and taking $\delta t \to dt$ yields the stochastic master equation

$$\mathrm{d}\hat{\rho}_J(t) = -\mathrm{i}\left[\hat{H}, \hat{\rho}_J(t)\right]\mathrm{d}t + \gamma_E\mathrm{d}t\mathcal{D}[\hat{c}]\hat{\rho}_J(t) + \sqrt{\gamma_E}\mathrm{d}W(t)\mathcal{H}[\hat{c}]\hat{\rho}_J(t), \tag{1.46}$$

where $dW(t)$ is an infinitesimal Wiener increment satisfying the conditions

$$\mathrm{E}[dW(t)] = 0, \tag{1.47}$$

$$dW(t)^2 = dt. \tag{1.48}$$

Eq. (1.46) represents diffusive evolution and no longer a jump evolution as Eq. (1.37).
As $\gamma \to \infty$ transforms the evolution to continuous, it also changes the point-process photo-count into a continuous photo current with white noise. Removing the constant local oscillator contribution gives

$$J_{hom} = \lim_{\gamma\to\infty} \frac{\delta N(t) - \gamma^2 \delta t}{\gamma \delta t} = \sqrt{\gamma_E} \langle \hat{x} \rangle_J (t) + \frac{\xi(t)}{\sqrt{\gamma_E}}, \tag{1.49}$$

where $\xi(t) = \frac{dW(t)}{dt}$.

To obtain Eq. (1.33), we made the assumption that the measurement efficiency was close to one. In order to obtain a more realistic and general SME, we now consider an inefficient detection (i.e. $0 \leq \eta \leq 1$) which we model by a perfect detector detecting only a proportion η of the output beam [16]. The homodyne photo current can be obtained by replacing the operator $\hat{c}$ with $\sqrt{\eta}\hat{c}$ which yields

$$J_{hom} = \sqrt{\gamma_E \eta} \langle \hat{x} \rangle_J (t) + \frac{\xi(t)}{\sqrt{\gamma_E}}, \tag{1.50}$$

that we can normalize so that the deterministic part is not dependant on the efficiency. We obtain

$$J_{hom} = \sqrt{\gamma_E} \langle \hat{x} \rangle_J (t) + \frac{\xi(t)}{\sqrt{\gamma_E \eta}}. \tag{1.51}$$

Eq. (1.46) is thus modified to

$$\mathrm{d}\hat{\rho}_J(t) = -\mathrm{i}\left[\hat{H}, \hat{\rho}_J(t)\right] \mathrm{d}t + \gamma_E \mathcal{D}[\hat{c}]\hat{\rho}_J(t)\mathrm{d}t + \sqrt{\gamma_E \eta}\mathrm{d}W(t)\mathcal{H}[\hat{c}]\hat{\rho}_J(t). \tag{1.52}$$

To better model the system we consider (i.e., a photon travelling in an optical fiber), an intrinsic dissipation term is added to equation (1.52), which gives

$$\boxed{d\hat{\rho}_J(t) = -\mathrm{i}\left[\hat{H}, \hat{\rho}_J(t)\right] \mathrm{d}t + \gamma_E \mathcal{D}[\hat{c}]\hat{\rho}_J(t)\mathrm{d}t + \sqrt{\gamma_E \eta}\mathrm{d}W(t)\mathcal{H}[\hat{c}]\hat{\rho}_J(t) + \gamma_D \mathcal{D}[\hat{v}]\hat{\rho}_J(t)dt,} \tag{1.53}$$

where $\hat{v}$ is the dissipation operator and γ_D the corresponding dissipation rate.

This equation is one of the central points of this work; to describe the effects of dissipation and weak measurement on the BB84 protocol. More specifically, this equation can serve as a model of the temporal evolution of a photon travelling inside an optical fiber, where dissipation occurs, and being subject to weak measurements such as homodyne detection.

1.6 Quantum feedback control

The goal of this Section is to derive a stochastic master equation describing the evolution of a quantum system monitored with homodyne detection, to which one apply quantum feedback. Indeed, as stated in the introduction, one of the goal of this master thesis is to investigate the use of quantum feedback by a spy, in order to decrease its measurement impact on the photons, and thus become less detectable.

As shown in the previous section, the SME for homodyne detection is Eq. (1.52). We derived it from Eq. (1.37) which, since it is the starting point to describe feedback, we recall to be

$$\mathrm{d}\hat{\rho}_J(t) = \left\{ dN(t)\mathcal{G}[\hat{c} + \gamma] + dt\mathcal{H}\left[-i\hat{H} - \gamma\sqrt{\gamma_E}\hat{c} - \frac{\gamma_E}{2}\hat{c}^\dagger\hat{c}\right]\right\} \hat{\rho}_J(t). \tag{1.54}$$

Then, the expression for the effect of the feedback is

$$\left[\dot{\hat{\rho}}_J(t)\right]_{\text{fb}} = \mathcal{F}\left[t, \mathbf{J}_{hom[0,t)}\right]\hat{\rho}_J(t), \tag{1.55}$$

where $\mathbf{J}_{hom[0,t)}$ is the photo current record from time $t_0 = 0$ (beginning of the experiment) to time t (present time). $\mathcal{F}\left[t, \mathbf{J}_{hom[0,t)}\right]$ is a superoperator, i.e., a functional of the current for all past times. To completely describe the effect of feedback, we can add Eq. (1.55) to Eq. (1.54). Simplifying assumptions can be done, firstly we can consider that the feedback functional is linear, which generates an evolution of the form

$$\left[\dot{\hat{\rho}}_J(t)\right]_{\text{fb}} = \int_0^\infty h(s) J_{hom}(t-s)\mathcal{K}\hat{\rho}_J(t)\mathrm{d}s, \tag{1.56}$$

where $h(s)$ is the response function and $\mathcal{K}$ is a Liouville superoperator.

We first consider that $h(s) = \delta(s-\tau)$, modelling a fixed delay τ of the feedback. According to [19], the time evolution is given by

$$\left[\dot{\hat{\rho}}_J(t)\right]_{\text{fb}} = J_{hom}(t-\tau)\mathcal{K}\hat{\rho}_J(t). \tag{1.57}$$

$\mathcal{K}$ must be such as to give valid evolution whether the time is positive or negative since the homodyne photo current may be negative because of the substraction of the constant local oscillator. In other words, it must generate reversible evolution with

$$\mathcal{K}\hat{\rho} = -i\sqrt{\gamma_E}\left[\hat{F}, \hat{\rho}\right], \tag{1.58}$$

for $\hat{F}$ some hermitian operator. To combine Eq. (1.57) with Eq. (1.54) it is necessary to convert it from an implicit to an explicit equation, which yields

$$\hat{\rho}_J(t) + [d\hat{\rho}_J(t)] = \exp\left[\mathcal{K}J_{hom}(t-\tau)dt\right]\hat{\rho}_J(t). \tag{1.59}$$

The total conditioned evolution of the system is, according to [20],

$$\begin{aligned}\hat{\rho}_J(t+\mathrm{d}t) = &\left\{1 + \mathcal{K}\left[\sqrt{\gamma_E}\left\langle\hat{c}+\hat{c}^\dagger\right\rangle_J(t-\tau)\mathrm{d}t + \mathrm{d}W(t-\tau)/\sqrt{\eta}\right] + [1/(2\eta)]\mathcal{K}^2\ \mathrm{d}t\right\}\\ &\times\left\{1 + \mathcal{H}[-\mathrm{i}\hat{H}]\mathrm{d}t + \gamma_E\mathcal{D}[\hat{c}]\mathrm{d}t + \sqrt{\gamma_E\eta}\mathrm{d}W(t)\mathcal{H}[\hat{c}]\right\}\hat{\rho}_J(t).\end{aligned} \tag{1.60}$$

Considering a finite τ, this equation becomes

$$\begin{aligned}\mathrm{d}\hat{\rho}_J(t) =&\mathrm{d}t\left\{\mathcal{H}[-\mathrm{i}\hat{H}] + \gamma_E\mathcal{D}[\hat{c}] + \sqrt{\gamma_E}\left\langle\hat{c}+\hat{c}^\dagger\right\rangle_J(t-\tau)\mathcal{K} + \frac{1}{2\eta}\mathcal{K}^2\right\}\hat{\rho}_J(t)\\ &+\mathrm{d}W(t-\tau)\mathcal{K}\hat{\rho}_J(t)/\sqrt{\eta} + \sqrt{\gamma_E\eta}\mathrm{d}W(t)\mathcal{H}[\hat{c}]\hat{\rho}_J(t).\end{aligned} \tag{1.61}$$

Putting $\tau = 0$ in Eq. (1.60) is considering a feedback with no delay, in other words an instantaneous feedback. This yields

$$\begin{aligned}\mathrm{d}\hat{\rho}_J(t) =&\mathrm{d}t\left\{-\mathrm{i}\left[\hat{H}, \hat{\rho}_J(t)\right] + \gamma_E\mathcal{D}[\hat{c}]\hat{\rho}_J(t) - \mathrm{i}\gamma_E\left[\hat{F}, \hat{c}\hat{\rho}_J(t) + \hat{\rho}_J(t)\hat{c}^\dagger\right]\right\}\\ &+\gamma_E\mathcal{D}[\hat{F}]\hat{\rho}_J(t)\mathrm{d}t/\eta + \gamma_E\mathrm{d}W(t)\mathcal{H}[\sqrt{\eta}\hat{c} - \mathrm{i}\hat{F}/\sqrt{\eta}]\hat{\rho}_J(t).\end{aligned} \tag{1.62}$$

The non-selective evolution equation of the system can be obtained by taking the ensemble average of this equation, which removes the last term because of the stochasticity of $dW(t)$. By doing so we obtain the homodyne mediated feedback master equation [20]:

$$\dot{\hat{\rho}} = -\mathrm{i}[\hat{H}, \hat{\rho}] + \gamma_E\mathcal{D}[\hat{c}]\hat{\rho} - \mathrm{i}\gamma_E\left[\hat{F}, \hat{c}\hat{\rho} + \hat{\rho}\hat{c}^\dagger\right] + \frac{\gamma_E}{\eta}\mathcal{D}[\hat{F}]\hat{\rho}. \tag{1.63}$$

The last two terms represent the feedback applied on the system. The first one is linear in $\hat{F}$ and is the desired effect of the feedback which would dominate in classical regime. The second one causes diffusion in the variable conjugate to the operator $\hat{F}$ which can be attributed to the inevitable introduction of noise in the system by the measurement. This last equation can be rewritten in the Lindblad form:

$$\dot{\hat{\rho}} = -\mathrm{i}\left[\hat{H} + \frac{\gamma_E}{2}\left(\hat{c}^\dagger\hat{F} + \hat{F}\hat{c}\right), \hat{\rho}\right] + \gamma_E\mathcal{D}[\hat{c} - \mathrm{i}\hat{F}]\hat{\rho} + \frac{1-\eta}{\eta}\gamma_E\mathcal{D}[\hat{F}]\hat{\rho} \equiv \mathcal{L}\hat{\rho}. \tag{1.64}$$

By comparing this equation to the original Lindblad equation [Eq. (1.14)], we notice that the effect of feedback is to replace the $\hat{c}$ operator by $\hat{c} - i\hat{F}$, to add an extra term to the hamiltonian and a term vanishing for the efficiency $\eta = 1$.

Since the dissipation in the optical fiber is totally independant from the measurement and the feedback applied, we can simply add to Eq. (1.64) the same dissipation term as in Eq. (1.52) which yields the equation we use to describe the system evolution

$$\boxed{\dot{\hat{\rho}} = -\mathrm{i}\left[\hat{H} + \frac{\gamma_E}{2}\left(\hat{c}^\dagger\hat{F} + \hat{F}\hat{c}\right), \hat{\rho}\right] + \gamma_E\mathcal{D}[\hat{c} - \mathrm{i}\hat{F}]\hat{\rho} + \frac{1-\eta}{\eta}\gamma_E\mathcal{D}[\hat{F}]\hat{\rho} + \gamma_D\mathcal{D}[\hat{v}]\hat{\rho}_J(t),} \tag{1.65}$$

using the same notations as previously. This equation will serve as model for the description of the BB84 protocol subjected to dissipation, weak measurement and unconditional feedback.

1.7 Summary

In this Chapter, we derived three main equations that we will use throughout this work. They are the Linblad master equation (1.16), the stochastic master equation conditioned on homodyne measurement (1.53) and finally the homodyne mediated feedback master equation (1.65).

Now that we have obtained the mathematical tools to describe the evolution of photons in various cases, we will detail the implementation of the BB84 protocol in the next Chapter.

Chapter 2

The BB84 protocol

The BB84 protocol is the first quantum cryptography protocol, formalized by Charles Bennett and Gilles Brassard in 1984 [6]. Its primary aim is to establish a secure communication channel between two parties by exploiting the principles of quantum mechanics. It is, however, not used to directly communicate information, but rather for these two parties to agree on a shared secret key, which will then be used to implement a classical private key protocol.

In the first Section of this Chapter we explore the definition of qubits from linearly polarized photons. We detail a generic implementation of the BB84 protocol in the second section, first using the approximations that the quantum channel used is perfect (i.e., there is no dissipation on the photons) and that there is no eavesdropping. Finally, we consider the solutions to an implementation of the protocol without these assumptions.

2.1 Photon polarization and qubit definition

The BB84 protocol makes use of polarized photons to define qubits of information. As a matter of fact, we choose to make use of photons because they travel at the speed of light, therefore information reaches its destination much faster than with other means of transportation. Another reason is that photons can be sent through optical fibers, which are already known and used in many places. We thus want to create qubits using photons in order to send them via an optical fiber.

A photon polarization lives in a two-dimensional Hilbert space, which means we need two parameters to completely define a polarization state. A photon can be polarized in every direction contained in a plane perpendicular to its direction of propagation. We choose as a basis the states $|0\rangle$ and $|1\rangle$, corresponding respectively to vertical and horizontal polarizations. Every polarization state can be written as a superposition of these two basis states, such as $|+\rangle$ and $|-\rangle$ defined as

$$\begin{aligned} |+\rangle &= \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \\ |-\rangle &= \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \end{aligned} \tag{2.1}$$

These 2 states respectively correspond to a polarization of the photons at angles of $\frac{\pi}{4}$ and $\frac{-\pi}{4}$ with respect to the vertical. We call the basis $\{|0\rangle,|1\rangle\}$ the computational basis (or Pauli-Z eigenbasis) and the basis $\{|+\rangle,|-\rangle\}$ the diagonal basis (or Pauli-X eigenbasis) [21]. These two bases

are mutually unbiased, because if any state of one basis is measured in the other, the results are equally likely [22].

To prepare photons in such states, the simplest technique is to send a weak beam of light through an oriented polaroid filter. To measure a photon state, we can either send it through an oriented birefringent crystal and observe which beam it emerges in, or observe whether or not it passes through another oriented polaroid filter. Indeed, if the photon is polarized vertically, it will not pass through a horizontally oriented filter, and vice versa. For diagonally polarized photons, if a vertical (resp. horizontal) filter is used, the polarization will change to vertical (resp. horizontal) after the measure [21].

2.2 Simplified case

To understand the protocol, let us first assume a perfect implementation, meaning we consider a perfectly noiseless channel and that no measure is done on the travelling photons. As explained earlier, no physical system can be perfectly isolated from its environment and thus these assumptions, especially the first one, are not scientifically possible.

Step 1: Alice chooses $(4 + \delta)n$ random data bits, which are her qubits. Use of the variable δ is explained later, at the end of this section.

Step 2: Alice chooses a random $(4 + \delta)n$-bit string b. She encodes each data bit as the quantum states $|0\rangle$ or $|+\rangle$ if the corresponding bit of b is 1 and $|1\rangle$ or $|-\rangle$ if the corresponding bit of b is 0.

Step 3: Alice sends the resulting qubits to Bob via an optical fiber.

Step 4: Bob receives the qubits and announces this fact. He then measures each of the qubits in the Pauli-X eigenbasis or the Pauli-Z eigenbasis at random, also following a random bit string he chose.

Step 5: Alice announces b via the public channel.

Step 6: Via the public channel Alice and Bob compare, for each qubit, the basis chosen by Alice to encode it and the basis chosen by Bob to measure this same qubit. They discard all the qubits where the two bases do not correspond. With high probability, there are at least $2n$ bits left (if not, abort the protocol). They thus have $2n$ bits, which is the key they need to implement a classical private key protocol [23].

The protocol is illustrated in Figure 2.1.

In the step 6 it is mentioned there are at least $2n$ bits left. This result is due to the use of the δ in the steps 1 and 2. Indeed, it is employed to ensure that for δ large enough, the probability Alice and Bob have at least $2n$ identical choices of polarization bases is large enough. Indeed, if $\delta = 0$, there is a 50% chance that they did not share $2n$ identical polarization bases as they both have $4n$ bits to start with and there are 2 distinct bases.

2.3 Realistic case

Without the assumptions made in the previous section, we consider a case closer to reality, where the qubits sent through the optical fiber will be impacted by noise and measurements. In the following Chapter, we will analyze in detail the impact of these two factors on the protocol

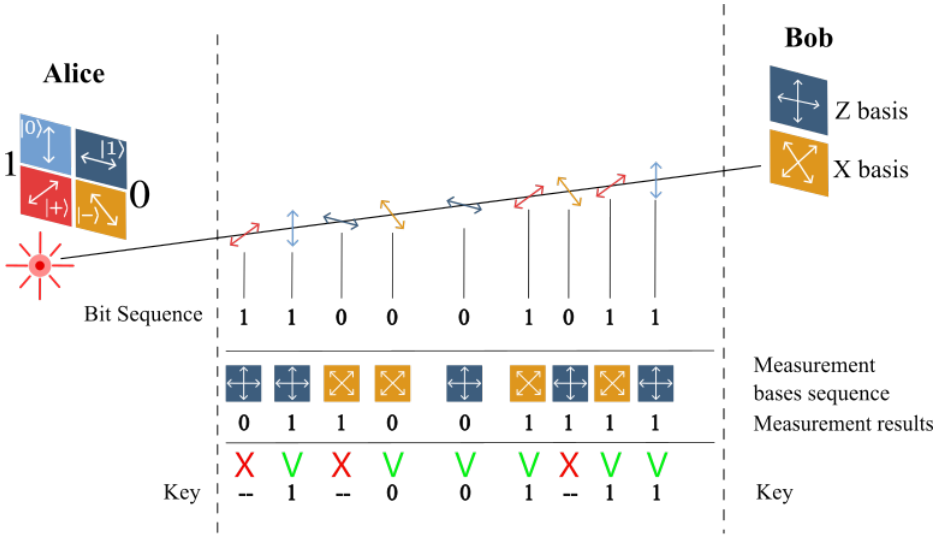


Figure 2.1: Illustration of the BB84 protocol in the perfect case. In this example, Alice sends 12 photons to Bob, among which half are polarized in the Pauli-X eigenbasis and the other half in the Pauli-Z eigenbasis. Bob measures randomly in one of these two bases each of the photons received, and the basis chosen is the right one for seven of these. The shared secret key is thus '100110'.

success probability. Three steps can be added to the protocol to counter the effects of dissipation and eavesdropping.

Step 7: Alice selects a subset of n bits to check on the interference caused by Eve, and tells Bob which bits she chose. They both announce and compare the values of the n check bits via the public channel. If more than a threshold number of qubits disagree, they abort the protocol.

Step 8: Alice and Bob perform information reconciliation, which only requires publicly comparing a random portion of their respective strings to estimate the bit error rate (BER). They assume the rest of their bits have the same proportion of error than the part they compared. The next step is to correct the remaining errors (being at unknown positions), this can be done by divulging only a negligible amount of bits, if the error rate is reasonably small. These divulged bits are then discarded, there are $p < n$ bits left.

At this point, Alice and Bob have the same key but Eve could hold a similar key. The next step is performed to decrease as much as possible Eve's mutual information with them.

Step 9: They implement privacy amplification on their shared private key. They thus both agree (via the public channel) on a universal hash function g that takes a p-bit string as input and gives an m-bit string as output. By using this hash function on their p-bit key string, they obtain an m-bit key that is more secure. Indeed, universal hash functions are uniform, meaning a good hash function should map the expected inputs as evenly as possible over its output range. That is, every hash value in the output range should be generated with roughly the same probability. This implies that even if Eve had a key similar to Alice and Bob, her hash value would not be close at all to theirs. The privacy of their key is thus enhanced and Eve's mutual information is reduced.

Chapter 3

Impact of noise and attacks on the protocol security

In this Chapter, we first assess the impact, on the protocol success probability, of projective measurements and noise separately, then jointly. We then analyze the impact of weak homodyne measurements on the photons states via the trace distance metric.

3.1 Impact of projective measurement

If we assume a noiseless channel but eavesdropping, we can accurately determine the probability that Alice and Bob detect the presence of Eve. Indeed, when Eve is not present, we know that Alice and Bob will measure the same state if choosing the same basis.

We consider Alice sends a photon in the state

$$|\psi_a\rangle = \cos a\,|0\rangle + \sin a\,|1\rangle\,, \tag{3.1}$$

where a is the angle between the photon polarization direction and the vertical axis and can then have 4 values: 0, $\frac{\pi}{2}$ (Pauli-Z basis) or $\pm\frac{\pi}{4}$ (Pauli-X basis).
The photon reaches Eve and she measures it in the state

$$|\psi_s\rangle = \cos s\,|0\rangle + \sin s\,|1\rangle\,, \tag{3.2}$$

where s is the equivalent to a for Eve's measurement direction. The probability of this result is

$$\mathbb{P}_s = |\langle\psi_s|\psi_a\rangle|^2 = |\cos s\cos a + \sin s\sin a|^2 = \cos^2(s-a). \tag{3.3}$$

The state after the measurement is

$$\begin{aligned} |\psi'\rangle &= \frac{|\psi_s\rangle\,\langle\psi_s|\psi_a\rangle}{\sqrt{\mathbb{P}_s}} \\ &= \frac{|\psi_s\rangle\cos(s-a)}{\cos(s-a)} \\ &= |\psi_s\rangle\,. \end{aligned}$$

This photon then reaches Bob who measures it in the state

$$|\psi_b\rangle = \cos b\,|0\rangle + \sin b\,|1\rangle\,, \tag{3.4}$$

where b is still the same but for Bob's measure direction. The probability of this result is

$$\mathbb{P}_b = |\langle\psi_b|\psi_s\rangle|^2 = |\cos b \cos s + \sin b \sin s|^2 = \cos^2(b-s). \tag{3.5}$$

Let us first consider the case where Alice and Bob choose the same basis but Eve does not, i.e., $a = b$. The probabilities are summarized in the Table 1. In this case, the probability $\mathbb{P}_b$ represents the probability that Bob measures the same state as Alice sent him, in other words, it represents the probability that Eve is not detected. If they both choose the same basis and Eve chooses randomly the basis in which she performs her measurement, there are four distinct cases each equally probable as described in the Table 1. The probability $\mathbb{P}$ that Eve is not detected is then

$$\mathbb{P} = \frac{1}{4}1 + \frac{1}{4}\frac{1}{2} + \frac{1}{4}1 + \frac{1}{4}\frac{1}{2} = \frac{3}{4}, \tag{3.6}$$

i.e., there is a 75% chance that Alice and Bob have the same result for a considered qubit. If Alice and Bob also choose randomly the basis of use, the probability that Eve is discovered is divided by 2 and is thus $\frac{1}{8}$. When generalizing to N qubits, the probability that Eve is discovered is $1 - (\frac{7}{8})^N$. For example, if $N = 200$ qubits, $(\frac{7}{8})^{200} = 1,5.10^{-12}$ and we can conclude Eve is discovered almost every time she attempts to eavesdrop. In other words the BB84 protocol is secure for an arbitrarily high number of qubits.

In Section 2.3, step 7, we mentioned the use of a threshold value of disagreeing qubits. This threshold is chosen arbitrarily in function of the number of qubits employed (N) and the level of security wanted. For example if $N = 10$, the probability that Alice and Bob have different results (i.e., Eve is discovered) is $1 - (\frac{7}{8})^{10} = 0,7369 = 73,69\%$ thus the threshold value could be defined as 70%. We can conclude that the more qubits are used in the protocol, the more secure it is to eavesdropping.

a	$\|\psi_a\rangle$	s	$\mathbb{P}_s$	$\|\psi_s\rangle$	$\mathbb{P}_b$
0 $\frac{\pi}{2}$	$\|0\rangle$ $\|1\rangle$	0 $\frac{\pi}{2}$	1	$\|0\rangle$	1
		$\pm\frac{\pi}{4}$	$\frac{1}{2}$	$\frac{\|0\rangle\pm\|1\rangle}{\sqrt{2}}$	$\frac{1}{2}$
$\pm\frac{\pi}{4}$	$\frac{\|0\rangle\pm\|1\rangle}{\sqrt{2}}$	0 $\frac{\pi}{2}$	$\frac{1}{2}$	$\|0\rangle$ $\|1\rangle$	$\frac{1}{2}$
		$\pm\frac{\pi}{4}$	1	$\frac{\|0\rangle\pm\|1\rangle}{\sqrt{2}}$	1

Table 3.1: Probabilities that Eve ($\mathbb{P}_s$) and Bob ($\mathbb{P}_b$) obtain the same result than Alice depending on the initial angle a and Alice's state $|\psi_a\rangle$. The variable s is the angle with respect to the vertical at which Eve performs her measurement on the qubit, with probability $\mathbb{P}_s$ to obtain the state Alice sent. $|\psi_s\rangle$ is the state of the qubit after the measure of Eve. We see that if Eve measured in the same basis as Alice encoded her qubit, Bob has a probability of 1 to measure the same state, which means Eve is not detected. However, if Eve does not select the same basis, this probability drops to $\frac{1}{2}$.

3.2 Impact of noise

The noise in the quantum channel (here optical fiber) can be computed by various models, we choose here the Pauli Noise model where the dissipation is represented, in the basis $\{|0\rangle, |1\rangle\}$, by the Pauli operators [24]:

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.7}$$

The motivation behind this choice is that all kind of dissipation can be modeled with these 3 types of errors. Indeed, $\hat{\sigma}_x$ represents a bit flip error, i.e., a bit being changed from the state $|0\rangle$ to the state $|1\rangle$ or inversely, while $\hat{\sigma}_z$ represents a phase flip error and $\hat{\sigma}_y$ a combination of these two.

Among these 3 types of errors, we select only the $\hat{\sigma}_x$ one. As said in Section 1.4, we employ the Lindblad master equation (1.16) with $\hat{L}$ being the operator of dissipation ($\hat{\sigma}_x$ here) and

$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \tag{3.8}$$

is the matrix representation of the density operator of the system written in the basis $\{|0\rangle, |1\rangle\}$, so that

$$\hat{\rho}_{ij} = Tr\left(|j\rangle \langle i| \hat{\rho}\right) = \langle i| \hat{\rho} |j\rangle \qquad \text{with} \qquad i, j = 0, 1. \tag{3.9}$$

We employ one unique Lindblad operator thus the summing index i in Eq. (1.16) has only one value. In addition, the first term of Eq. (1.16) is not of interest here (i.e., the Hamiltonian is set to 0) and we can thus study only the second term. Therefore we have

$$\dot{\hat{\rho}} = \gamma_D \left(\hat{L}\hat{\rho}\hat{L}^\dagger - \frac{1}{2}\hat{L}^\dagger\hat{L}\hat{\rho} - \frac{1}{2}\hat{\rho}\hat{L}^\dagger\hat{L} \right). \tag{3.10}$$

In our case,

$$\hat{L} = \hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{3.11}$$

and thus,

$$\hat{L}\hat{\rho}\hat{L}^\dagger = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{10} \\ \rho_{01} & \rho_{00} \end{pmatrix}. \tag{3.12}$$

Inserting that in Eq. (3.10) yields a system of four differential equations (one for each element of the density matrix):

$$\begin{cases} \dot{\rho}_{00} = \gamma_D(\rho_{11}(t) - \rho_{00}(t)) \\ \dot{\rho}_{01} = \gamma_D(\rho_{10}(t) - \rho_{01}(t)) \\ \dot{\rho}_{10} = \gamma_D(\rho_{01}(t) - \rho_{10}(t)) \\ \dot{\rho}_{11} = \gamma_D(\rho_{00}(t) - \rho_{11}(t)) \end{cases} \tag{3.13}$$

These equations are coupled two by two. Resolving this system gives :

$$\hat{\rho}(t) = \begin{pmatrix} \frac{e^{-2\gamma_D t}}{2}(1 + e^{2\gamma_D t})\rho_{00}(0) + \frac{e^{-2\gamma_D t}}{2}(-1 + e^{2\gamma_D t})\rho_{11}(0) & \frac{e^{-2\gamma_D t}}{2}(1 + e^{2\gamma_D t})\rho_{01}(0) + \frac{e^{-2\gamma_D t}}{2}(-1 + e^{2\gamma_D t})\rho_{10}(0) \\ \frac{e^{-2\gamma_D t}}{2}(-1 + e^{2\gamma_D t})\rho_{01}(0) + \frac{e^{-2\gamma_D t}}{2}(1 + e^{2\gamma_D t})\rho_{10}(0) & \frac{e^{-2\gamma_D t}}{2}(-1 + e^{2\gamma_D t})\rho_{00}(0) + \frac{e^{-2\gamma_D t}}{2}(1 + e^{2\gamma_D t})\rho_{11}(0) \end{pmatrix} \tag{3.14}$$

Which describes the state of the qubit in the channel at time t.

As discussed in Chapter 2, there are four possible states for Alice to send :$\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$.

1. In the case $\hat{\rho}(0) = |0\rangle\langle 0|$, we have $\rho_{00}(0) = 1$ and Eq. (3.14) gives

$$\hat{\rho}(t) = \begin{pmatrix} \frac{e^{-2\gamma_D t}}{2} + \frac{1}{2} & 0 \\ 0 & \frac{-e^{-2\gamma_D t}}{2} + \frac{1}{2} \end{pmatrix}. \tag{3.15}$$

   The probability that a system in a mixed state is in a state $|\psi\rangle$ is $\mathbb{P}(|\psi\rangle) = Tr(|\psi\rangle\langle\psi|\hat{\rho})$. In this case, the purity is equal to $\frac{e^{-4\gamma_D t}}{2} + \frac{1}{2}$, which means the state will evolve from initially pure to more and more mixed with time.

   The probability that the qubit is in the state $|0\rangle$ after going through the optical fiber is thus given by $\rho_{00}(t) = \frac{e^{-2\gamma_D t}}{2} + \frac{1}{2}$ which for a time $t \to \infty$ is equal to the probability to have the state $|1\rangle$, equal to $\frac{1}{2}$.

2. In the case $\hat{\rho}(0) = |1\rangle\langle 1|$, $\rho_{11}(0) = 1$, we find

$$\hat{\rho}(t) = \begin{pmatrix} \frac{-e^{-2\gamma_D t}}{2} + \frac{1}{2} & 0 \\ 0 & \frac{e^{-2\gamma_D t}}{2} + \frac{1}{2} \end{pmatrix}. \tag{3.16}$$

   The probability to be in state $|1\rangle$ is thus given by $\rho_{11}(t) = \frac{e^{-2\gamma_D t}}{2} + \frac{1}{2}$.

3. In the case $\hat{\rho}(0) = |+\rangle\langle +|$, we find

$$\hat{\rho}(t) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{3.17}$$

4. In the case $\hat{\rho}(0) = |-\rangle\langle -|$, we find

$$\hat{\rho}(t) = \begin{pmatrix} \frac{1}{2} & \frac{-1}{2} \\ \frac{-1}{2} & \frac{1}{2} \end{pmatrix}. \tag{3.18}$$

As the states $|+\rangle$ and $|-\rangle$ are eigenstates of the $\hat{\sigma}_x$ jump operator, they will not change when traveling through the optical fiber: they are insensitive to the dissipation process. These states are sometimes called decoherence-free states (DFS) or dark states.

With these results, we can estimate the total probability that the polarization state of the photon reaching Bob is the same that the one Alice sent. Indeed, the probability that Alice sends one of the four states is $\frac{1}{4}$ because she chooses randomly the polarization basis and the state in this polarization. Thus we have

$$\mathbb{P}(\text{same results}) = \frac{1}{4}\frac{1}{2} + \frac{1}{4}\frac{1}{2} + \frac{1}{4}1 + \frac{1}{4}1 = \frac{3}{4}, \tag{3.19}$$

for $t \to \infty$. Interestingly, this probability coincides with the result in Eq. (3.6). This means that a spy in a noiseless channel or no spy in an infinitely-long noisy channel have the same effect on the probability for Alice and Bob to get the same result, meaning that eavesdropping and dissipation are indistinguishable in this case.

If we consider the photon spends a limited time in the optical fiber , which is a more realistic consideration, we find:

$$\mathbb{P}(\text{same results}) = \frac{1}{4}\left(\frac{e^{-2\gamma_D t}}{2} + \frac{1}{2}\right) + \frac{1}{4}\left(\frac{e^{-2\gamma_D t}}{2} + \frac{1}{2}\right) + \frac{1}{4}1 + \frac{1}{4}1 = \frac{e^{-2\gamma_D t}}{4} + \frac{3}{4}. \tag{3.20}$$

For a very small t, we can approximate this probability to 1 and for $t \to \infty$ it is equal to $\frac{3}{4}$ which is coherent with our previous calculations.

3.3 Impact of both noise and projective measurement

In this section, we analyze the impact of both the processes considered in the previous sections simultaneously acting on the system. Suppose Alice wants to send her photon in an optical fiber of length L such as its travel time is τ. At the halfway point of the photon's path is Eve that measures the photon before it goes through the other half of the optical fiber to reach Bob. This is shown in Figure 3.1.

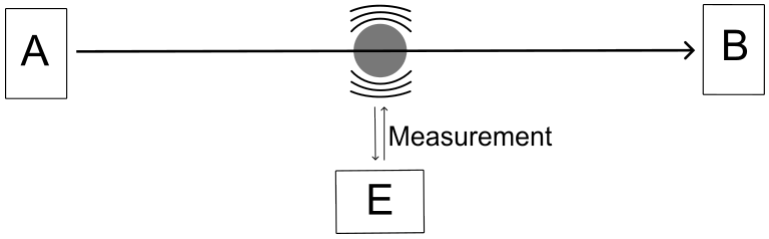


Figure 3.1: Illustration of the system considered to study the impact of noise and eavesdropping simultaneously. A is Alice, B is Bob and E is Eve.

When traveling from Alice to Eve and from Eve to Bob, the state of the photon will evolve due to the noise following Eq. (3.14). Therefore the state reaching Eve is

$$\hat{\rho}\left(\frac{\tau}{2}\right) = \begin{pmatrix} \frac{e^{-\gamma_D\tau}}{2}(1+e^{\gamma_D\tau})\rho_{00}(0) + \frac{e^{-\gamma_D\tau}}{2}(-1+e^{\gamma_D\tau})\rho_{11}(0) & \frac{e^{-\gamma_D\tau}}{2}(1+e^{\gamma_D\tau})\rho_{01}(0) + \frac{e^{-\gamma_D\tau}}{2}(-1+e^{\gamma_D\tau})\rho_{10}(0) \\ \frac{e^{-\gamma_D\tau}}{2}(-1+e^{\gamma_D\tau})\rho_{01}(0) + \frac{e^{-\gamma_D\tau}}{2}(1+e^{\gamma_D\tau})\rho_{10}(0) & \frac{e^{-\gamma_D\tau}}{2}(-1+e^{\gamma_D\tau})\rho_{00}(0) + \frac{e^{-\gamma_D\tau}}{2}(1+e^{\gamma_D\tau})\rho_{11}(0) \end{pmatrix} \tag{3.21}$$

As discussed earlier, Eve has two choices: measuring in the basis $\{|0\rangle, |1\rangle\}$ or in the basis $\{|+\rangle, |-\rangle\}$. If she chooses the first one, the probabilities of her measuring the states $|0\rangle$ or $|1\rangle$ are :

$$\mathbb{P}(|0\rangle) = \mathrm{Tr}\left(|0\rangle\langle 0|\,\hat{\rho}\left(\frac{\tau}{2}\right)\right) = \frac{e^{-\gamma_D\tau}}{2}(1+e^{\gamma_D\tau})\rho_{00}(0) + \frac{e^{-\gamma_D\tau}}{2}(-1+e^{\gamma_D\tau})\rho_{11}(0) = \rho_{00}\left(\frac{\tau}{2}\right), \tag{3.22}$$

$$\mathbb{P}(|1\rangle) = \mathrm{Tr}\left(|1\rangle\langle 1|\,\hat{\rho}\left(\frac{\tau}{2}\right)\right) = \frac{e^{-\gamma_D\tau}}{2}(-1+e^{\gamma_D\tau})\rho_{00}(0) + \frac{e^{-\gamma_D\tau}}{2}(1+e^{\gamma_D\tau})\rho_{11}(0) = \rho_{11}\left(\frac{\tau}{2}\right). \tag{3.23}$$

If she chooses the other basis, the probabilities of her measuring the states $|-\rangle$ or $|+\rangle$ are :

$$\mathbb{P}(|-\rangle) = \mathrm{Tr}\left(|-\rangle\langle -|\,\hat{\rho}\left(\frac{\tau}{2}\right)\right) = \frac{1}{2}(\rho_{00}(0) + \rho_{11}(0) - \rho_{01}(0) - \rho_{10}(0)), \tag{3.24}$$

$$\mathbb{P}(|+\rangle) = \mathrm{Tr}\left(|+\rangle\langle+|\,\hat{\rho}\left(\frac{\tau}{2}\right)\right) = \frac{1}{2}(\rho_{00}(0) + \rho_{11}(0) + \rho_{01}(0) + \rho_{10}(0)). \tag{3.25}$$

As we recall from Section 3.1, after Eve's measurement on a photon its polarization becomes the same as the measurement filter one (e.g., if Eve measures in the basis $\{|0\rangle, |1\rangle\}$, the photon will be in one of these 2 states, depending on the result). The probabilities calculated above are thus the probabilities of the photon being in these states after the measurement. We also have that the general state of Eq. (3.14) reaching Bob is the same as the one reaching Eve and the probabilities are the same too. Indeed, Eve's measurement is in fact a projection of the mixed state $\hat{\rho}(\frac{\tau}{2})$ onto a pure state belonging to the two bases considered.

We can now consider individually each state that Alice can send initially i.e., $|0\rangle, |1\rangle, |+\rangle$ and $|-\rangle$. Using Eqs. (3.21) to (3.25), we can calculate Eve's probabilities of measure. If Alice sends the state $|0\rangle$:

Eve's probabilities of measure are

$$\mathbb{P}(|0\rangle) = \frac{e^{-\gamma_D\tau}}{2}(1 + e^{\gamma_D\tau}) = \frac{1}{2} + \frac{e^{-\gamma_D\tau}}{2}, \tag{3.26}$$

$$\mathbb{P}(|1\rangle) = \frac{e^{-\gamma_D\tau}}{2}(1 + e^{\gamma_D\tau}) = \frac{1}{2} - \frac{e^{-\gamma_D\tau}}{2}, \tag{3.27}$$

if she chose the $\{|0\rangle, |1\rangle\}$ basis, and

$$\mathbb{P}(|+\rangle) = \mathbb{P}(|-\rangle) = \frac{1}{2}, \tag{3.28}$$

if she chose the $\{|+\rangle, |-\rangle\}$ basis.

Following the same logic we can do the same for the three other initial states possible, the results are summarized in Table 3.2

	Alice's qubit state			
	$\|0\rangle$	$\|1\rangle$	$\|+\rangle$	$\|-\rangle$
$\mathbb{P}(\|0\rangle)$	$\frac{1}{2} + \frac{e^{-\gamma_D\tau}}{2}$	$\frac{1}{2} - \frac{e^{-\gamma_D\tau}}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\mathbb{P}(\|1\rangle)$	$\frac{1}{2} - \frac{e^{-\gamma_D\tau}}{2}$	$\frac{1}{2} + \frac{e^{-\gamma_D\tau}}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\mathbb{P}(\|+\rangle)$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
$\mathbb{P}(\|-\rangle)$	$\frac{1}{2}$	$\frac{1}{2}$	0	1

Table 3.2: Probabilities that Eve measures each possible state (rows) depending on the initial state Alice sent (columns).

Since Bob has the same 2 possibilities of measurement basis, there are 16 different cases for him (e.g., one of them is Alice sends the state $|0\rangle$, Eve measures in the $\{|0\rangle, |1\rangle\}$ basis and Bob measures in the $\{|+\rangle, |-\rangle\}$ basis.). As discussed above, Eve's measurement is a projection of a mixed state on one of the four basis states, therefore the probabilities calculated above can be

used for Bob's measurement too. We can calculate the total probability that Bob has the same result than Alice, but first it is useful to do it for a specific case. Suppose Alice sent the state $|0\rangle$ ($\frac{1}{4}$ probability), then Eve measured in one of the two bases ($\frac{1}{2}$ probability), and Bob did it in the same basis as Alice. The probabilities of Bob measuring the state $|0\rangle$, depending on the basis chosen by Eve are thus:

$$\begin{cases} \mathbb{P}_b\left(|0\rangle \,|\{|0\rangle, |1\rangle\}\right) = \frac{1}{4}\frac{1}{2}\left(\left(\frac{1}{2}+\frac{e^{-\gamma_D \tau}}{2}\right)\left(\frac{1}{2}+\frac{e^{-\gamma_D \tau}}{2}\right)+\left(\frac{1}{2}-\frac{e^{-\gamma_D \tau}}{2}\right)\left(\frac{1}{2}-\frac{e^{-\gamma_D \tau}}{2}\right)\right) \\ \qquad\qquad\qquad = \frac{1}{16}\left(1+e^{-2\gamma_D \tau}\right) \\ \mathbb{P}_b\big(|0\rangle \,|\{|+\rangle, |-\rangle\}\big) = \frac{1}{8}\left(\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2\right) = \frac{1}{16} \end{cases} \tag{3.29}$$

By repeating the same logic for all the other cases, we obtain:

$$\begin{cases} \mathbb{P}_b\big(|1\rangle \,|\{|0\rangle, |1\rangle\}\big) = \frac{1}{16}(1+e^{-2\gamma_D \tau}) \\ \mathbb{P}_b\big(|1\rangle \,|\{|+\rangle, |-\rangle\}\big) = \frac{1}{16} \end{cases}, \tag{3.30}$$

if Alice sent the state $|1\rangle$;

$$\begin{cases} \mathbb{P}_b\big(|+\rangle \,|\{|0\rangle, |1\rangle\}\big) = \frac{1}{16} \\ \mathbb{P}_b\big(|+\rangle \,|\{|+\rangle, |-\rangle\}\big) = \frac{1}{8} \end{cases}, \tag{3.31}$$

if Alice sent the state $|+\rangle$; and

$$\begin{cases} \mathbb{P}_b\big(|-\rangle \,|\{|0\rangle, |1\rangle\}\big) = \frac{1}{16} \\ \mathbb{P}_b\big(|-\rangle \,|\{|+\rangle, |-\rangle\}\big) = \frac{1}{8} \end{cases}, \tag{3.32}$$

if she sent the state $|-\rangle$.

Now that we have calculated all the probabilities separately, the total probability of Bob having the same result as Alice when choosing the same basis of measurement is obtained by summing the above probabilities:

$$\begin{aligned} \mathbb{P}_b(\text{same result}|\text{same basis}) &= \frac{1}{16}(1+e^{-2\gamma_D \tau})+\frac{1}{16}+\frac{1}{16}(1+e^{-2\gamma_D \tau})+\frac{1}{16}+\frac{1}{16}+\frac{1}{8}+\frac{1}{16}+\frac{1}{8} \\ &= \frac{5}{8}+\frac{e^{-2\gamma_D \tau}}{8}. \end{aligned} \tag{3.33}$$

Finally, we can consider the most general case in which Bob arbitrarily chooses his measurement basis. If he does not choose the same basis as Alice, his probability of measuring the same state drops to 1/2. Indeed, if he measures in the $\{|+\rangle, |-\rangle\}$ basis, he has a probability of 1/2 to measure the states $|0\rangle$ or $|1\rangle$, and vice versa. Therefore the total probability of Bob measuring the same state as Alice sent (impacted by eavesdropping and dissipation) is :

$$\mathbb{P}_b(\text{same result}) = \frac{1}{2}\mathbb{P}_b(\text{same result}|\text{same basis}) = \frac{5}{16}+\frac{e^{-2\gamma_D \tau}}{16}. \tag{3.34}$$

Since Alice and Bob compare the basis they chose to measure each qubit and discard the ones where the basis is not the same (see Section 2.2) this result is less important than (3.33). However, (3.34) is useful to determine in advance the number of qubits they have to use to ensure that they have an amount of corresponding qubits large enough. The probabilities of Eqs. (3.19), (3.20) and (3.33) are summarized in Figure 3.2.

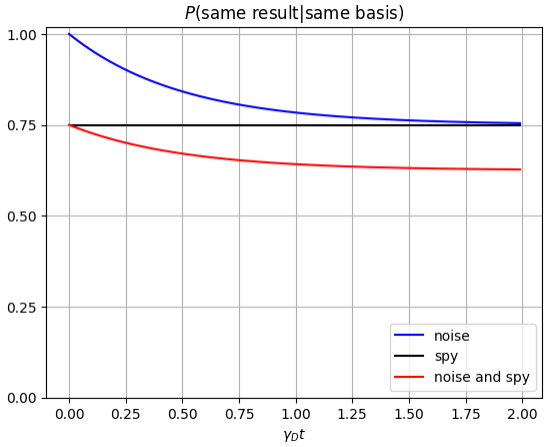


Figure 3.2: Probabilities that Bob measures the state Alice sent him for three cases, where the protocol is subject to spying only (black curve), to noise only (blue curve), to both (red curve).

3.4 Weak measurement

One of the techniques Eve could employ to try to decrease its impact on the qubits are weak measurements such as homodyne detection, in which case the photons are said to be continuously monitored. To do so, a single bosonic mode cavity could be placed around the optical fiber in which the photons travel. Then, the output field of this cavity is measured using the balanced homodyne scheme described in Section 1.5.2.

We can model the evolution of such photons by using Eq. (1.53) where we first decide to use $\hat{\sigma}_x$ and $\hat{\sigma}_z$ as the $\hat{v}$ and $\hat{c}$ operators respectively. It is worth recalling that these operators represents the dissipation in the channel and the spy's weak measurement respectively. The dissipation is thus still modeled as bit flip errors, such as in the previous sections. We also choose to set the Hamiltonian to $\omega\hat{\sigma}_z$ for the two initial states $|0\rangle$ and $|1\rangle$ and to $\omega\hat{\sigma}_x$ for the states $|+\rangle$ and $|-\rangle$, following [25].

With this specific choice of operators, simplifications can be made to Eqs. (1.15), (1.38) and (1.53), because the Pauli operators are hermitian and involutory. We thus have

$$\mathcal{D}[\hat{\sigma}_z]\hat{\rho} = \hat{\sigma}_z\hat{\rho}\hat{\sigma}_z - \frac{1}{2}\left(\hat{\sigma}_z\hat{\sigma}_z\hat{\rho} + \hat{\rho}\hat{\sigma}_z\hat{\sigma}_z\right) = \hat{\sigma}_z\hat{\rho}\hat{\sigma}_z - \hat{\rho}, \tag{3.35}$$

and

$$\mathcal{H}[\hat{\sigma}_z]\hat{\rho} = \hat{\sigma}_z\hat{\rho} + \hat{\rho}\hat{\sigma}_z - \mathrm{Tr}\left[\hat{\sigma}_z\hat{\rho} + \hat{\rho}\hat{\sigma}_z\right]\hat{\rho}. \tag{3.36}$$

It is obvious that using $\hat{\sigma}_x$ instead of $\hat{\sigma}_z$ yields the same simplifications and thus the same equations with the operator index being the only difference. The equation modelling our system

is thus

$$\begin{aligned} d\hat{\rho}_J(t) = &-i\left[\hat{H}, \hat{\rho}_J(t)\right] dt \\ &+ \gamma_E \left(\hat{\sigma}_z \hat{\rho}_J(t) \hat{\sigma}_z - \hat{\rho}_J(t)\right) dt \\ &+ \sqrt{\gamma_E \eta} dW(t) \left(\hat{\sigma}_z \hat{\rho}_J(t) + \hat{\rho}_J(t) \hat{\sigma}_z - 2 \left\langle \hat{\sigma}_z \hat{\rho}_J(t) \right\rangle \hat{\rho}_J(t)\right) \\ &+ \gamma_D \left(\hat{\sigma}_x \hat{\rho}_J(t) \hat{\sigma}_x - \hat{\rho}_J(t)\right) dt. \end{aligned} \tag{3.37}$$

As stated in Section 1.5.2, the information a spy would obtain with this type of weak measurement concerns a quadrature of the system, called homodyne photo current:

$$J_{hom}(t) = \gamma_E \left\langle \hat{c} + \hat{c}^\dagger \right\rangle + \frac{1}{\sqrt{\gamma_E \eta}} \frac{dW}{dt}, \tag{3.38}$$

which, in our case, can be simplified to

$$J_{hom}(t) = \gamma_E \mathrm{Tr}\left(2\hat{\sigma}_z \hat{\rho}_J(t)\right) + \frac{1}{\sqrt{\gamma_E \eta}} \frac{dW}{dt}. \tag{3.39}$$

In practice, when it is obtained through numerical simulations (using Eq. (3.37)), a typical homodyne photo current will be as depicted in Figure 3.3. Unless stated otherwise, the efficiency $\gamma_E \eta$ is set to 0.5 for all the numerical simulations of this master thesis.

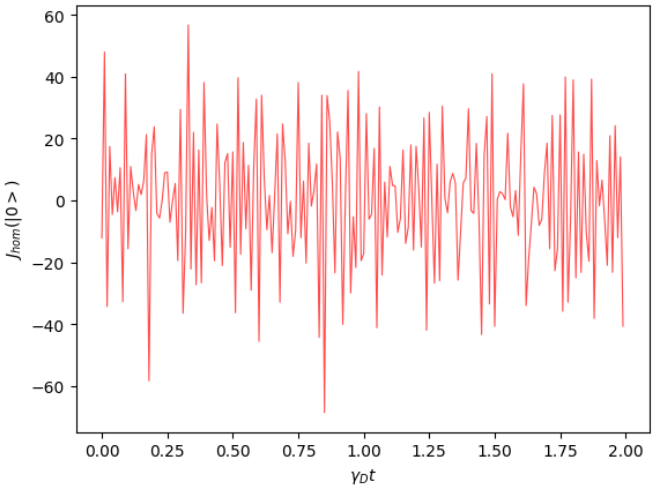


Figure 3.3: Typical homodyne photo current for the initial state $|0\rangle\langle 0|$ as a function of the time, obtained by numerical simulations with the time step parameter dt set to 0.01. Other parameters are set to $\omega = \gamma_E = \gamma_D = 1$.

We notice that the current is noisy, however we can compute statistics on it, which will show that information can still be extracted (as we show using a neural network in Section 4.3). Indeed, its mean and standard deviation will change depending on the initial state we consider, these values are summarized in Table 3.3. We notice that the means all have different values and could discriminate one state from the other, however the standard deviations are very high compared to the means so that it makes it impossible to distinguish different initial states, based on their individual means, from a single homodyne current.

Another statistics we can look at is the autocorrelation of the currents. The autocorrelation measures the correlation between a signal and a delayed version of itself, varying with the amount of delay. Essentially, it indicates how similar observations of a random variable are, based

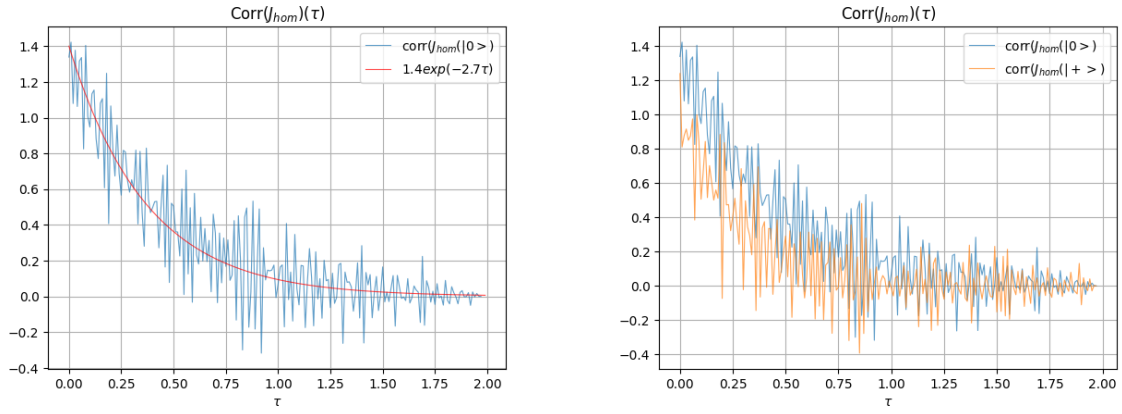
Initial state	Mean	Standard deviation
$\|0\rangle\langle 0\|$	0.48751	20.02372
$\|1\rangle\langle 1\|$	-0.48563	20.02785
$\|+\rangle\langle +\|$	-0.00146	20.04641
$\|-\rangle\langle -\|$	-0.00187	20.02813

Table 3.3: Means and standard deviations of the homodyne photo current measured for each of the four initial states.

on the time difference between them. This analysis serves as a mathematical technique for detecting recurring patterns within data, like identifying periodic signals that might be hidden by background noise. For a discrete function, it is computed as

$$\mathrm{corr}(J_{hom})(\tau) = \frac{1}{N}\sum_{t=1}^{N} \langle J_{hom}(t) J_{hom}(t+\tau)\rangle \,, \tag{3.40}$$

where N is the total number of time steps. We can thus plot it as a function of the delay τ, which yields the Figure 3.4. Figure 3.4 (a) shows that the autocorrelation behaves similarly to a decaying exponential (red curve) for the state $|0\rangle$, indicating that the current values over time are strongly correlated in the short term. However, this correlation rapidly decreases as we consider medium and long terms. Figure 3.4 (b) shows a similar behavior for the initial state $|+\rangle$.



(a) (b)

Figure 3.4: (a) Autocorrelation of the homodyne photo current measured for the initial state $|0\rangle$. The red curve is an exponential fitting the blue curve. (b) Autocorrelation of homodyne photo currents for the two initial states $|0\rangle$ (blue) and $|+\rangle$ (orange).

There are two parameters in Eq. (3.37), the efficiency of the measurement $\gamma_E\eta$ and the dissipation rate γ_D. Their respective values impact the dynamic of the system with respect to its environment: the larger γ_D is, the more the dissipation term will be large and thus the dynamic of the photon will be dominated by noise. If $\gamma_E\eta$ is set to 1, the measurement performed on the system is considered to be perfect and the photons states will be more impacted, while the spy will extract more information about their state through time. On the other hand, if $\gamma_E\eta$ is set

to 0.001 the measurement almost does not impact the photons but the spy is unable to extract any information from the states. Since the second term of Eq (3.39) is stochastic and includes $\gamma_E\eta$ in its denominator, such value of the efficiency leads to a photo current dominated by noise as displayed in Figure 3.5.

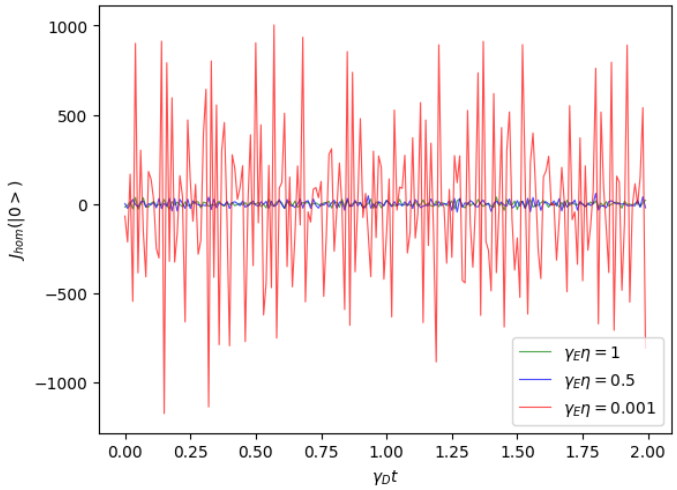


Figure 3.5: Homodyne photo current for the initial state $|0\rangle$, for efficiency ($\gamma_E\eta$) values of 1 (blue), 0.5 (green) and 0.001 (red). The latter yields a photo current with values between -1000 and 1000, almost entirely due to the stochastic term $\frac{1}{\sqrt{\gamma_E\eta}}\frac{dW}{dt}$. Other parameters are set to $\omega = \gamma_D = 1$.

A change in the value of the efficiency parameter actually impacts the standard deviations of the photo currents but not their means, which are summarized for different values in Table 3.4.

Initial state	$\gamma_E\eta = 0.001$	$\gamma_E\eta = 0.5$	$\gamma_E\eta = 1.0$
$\|0\rangle\langle 0\|$	0.47514 (447.52)	0.48751 (20.02)	0.48406 (14.20)
$\|1\rangle\langle 1\|$	-0.47786 (447.06)	-0.48563 (20.02)	-0.48226 (14.19)
$\|+\rangle\langle +\|$	-0.01763 (447.07)	-0.00146 (20.04)	-0.00276 (14.18)
$\|-\rangle\langle -\|$	0.01185 (447.33)	-0.00187 (20.02)	0.00196 (14.20)

Table 3.4: Means and standard deviations of the homodyne photo current measured for each of the four initial states, for different values of $\gamma_E\eta$.

3.5 Effect of weak measurement and dissipation

An important metric to define is the trace distance between two states. The trace distance between $\hat{\rho}$ and $\hat{\sigma}$ is [26]

$$D(\hat{\rho},\hat{\sigma}) = \frac{1}{2}\mathrm{Tr}\sqrt{(\hat{\rho}-\hat{\sigma})^\dagger(\hat{\rho}-\hat{\sigma})}, \tag{3.41}$$

where $\mathrm{Tr}|\hat{\rho}| = \mathrm{Tr}\sqrt{\hat{\rho}^\dagger\hat{\rho}}$ is the trace norm. In other words, the trace distance is defined as half the trace norm of the difference between the two considered states.

This quantity can be interpreted as a measure of state distinguishability. Suppose Alice prepares a quantum system in the state $\hat{\rho}$ with probability $\frac{1}{2}$ and in the state $\hat{\sigma}$ with probability

$\frac{1}{2}$. She gives the system to Bob, who performs a measurement to distinguish the two states. It can be shown that Bob's probability of correctly identifying which state Alice prepared is $\frac{1}{2} + \frac{D(\hat{\rho},\hat{\sigma})}{2}$. That is, $D(\hat{\rho},\hat{\sigma})$ can be interpreted, up to the factor $\frac{1}{2}$, as the optimal bias in favour of Bob correctly determining which of the two states was prepared.

For simplicity, in the following discussions we will denote a state subject to measurement ($\equiv$ spying) and noise evolving through time by $\hat{\rho}_{ns}(t)$ and a state subject to only noise by $\hat{\rho}_n(t)$. For example the time evolution of the initial state $|0\rangle\langle 0|$ under measurement and noise would be $|0\rangle\langle 0|_{ns}(t)$. We also denote the trace distances as follows:

$$D\left(|0\rangle\langle 0|_{ns}(t), |0\rangle\langle 0|\right) \equiv D_{ns}(|0\rangle\langle 0|)(t), \tag{3.42}$$

$$D\left(|0\rangle\langle 0|_{ns}(t), |0\rangle\langle 0|_{n}(t)\right) \equiv D_{nsn}(|0\rangle\langle 0|)(t), \tag{3.43}$$

where the last trace distance compares the state evolving due to the noise only, and the state evolving due to the noise and measurement, thus extracting the effect of the measurement only. The same logic is applied to the other states.

In this Section we first show how to derive an analytical solution of the stochastic master equation we use and compute different trace distances to analyze. We recall the SME to be

$$d\hat{\rho}_J(t) = -i\left[\hat{H}, \hat{\rho}_J(t)\right]dt + \gamma_E\mathcal{D}[\hat{c}]\hat{\rho}_J(t)dt + \sqrt{\gamma_E\eta}dW(t)\mathcal{H}[\hat{c}]\hat{\rho}_J(t) + \gamma_D\mathcal{D}[\hat{v}]\hat{\rho}_J(t)dt, \tag{3.44}$$

where $\hat{v}$ is still $\hat{\sigma}_z$ and the Hamiltonian $\hat{H} = \omega\hat{\sigma}_z$ for $|0\rangle$ and $|1\rangle$, and $\hat{H} = \omega\hat{\sigma}_x$ for $|+\rangle$ and $|-\rangle$. We also define a new measurement operator as depending on an angle θ, which we call the measurement angle:

$$\hat{c} = \cos(\theta)\hat{\sigma}_x + \sin(\theta)\hat{\sigma}_z. \tag{3.45}$$

For values of $0 + k\pi$ ($k \in \mathbb{Z}$), the cosinus is equal to plus or minus one while the sinus is null and $\hat{c} = \pm\hat{\sigma}_x$. On the other hand, when $\theta = \frac{\pi}{2} + k\pi$ ($k \in \mathbb{Z}$), $\hat{c} = \pm\hat{\sigma}_z$. It is important to recall that the problem we consider is symmetrical, the two bases of initial states $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ are respectively eigenstates of the Pauli-Z operator $\hat{\sigma}_z$ and the Pauli-X operator $\hat{\sigma}_x$. Therefore two states from the same basis will display the same behavior under the noise and the measurement we model using Pauli operators, which means we can consider one state of each basis (e.g. $|0\rangle$ and $|+\rangle$) throughout our analysis.

We start by averaging the SME (3.44) over all possible measurement results to cancel the stochastic term:

$$d\hat{\rho}_J(t) = -i\left[\hat{H}, \hat{\rho}_J(t)\right]dt + \gamma_E\mathcal{D}[\hat{c}]\hat{\rho}_J(t)dt + \gamma_D\mathcal{D}[\hat{v}]\hat{\rho}_J(t)dt. \tag{3.46}$$

Indeed, the expectation value of the Wiener increment $dW(t)$ is null [Eq. (1.47)]. Since the density matrix $\hat{\rho}_J(t)$ is a 2×2 matrix, this equation is in fact a system of four coupled differential equations, one for each matrix elements. Solving this system yields four evolution equations, which can be used to compute the trace distance of Eq. (3.41) between the initial state considered (e.g. $|0\rangle\langle 0|$) and its time evolution (described by Eq. (3.46)). This function is plotted as a function of θ and t in the panels (a) and (b) of Figure 3.6. We can plot this trace distance for the initial quantum state $|+\rangle\langle +|$ using the same method, as is shown in the panels (c) and (d) of Figure 3.6. While for the initial state $|0\rangle\langle 0|$ the trace distance increases with time until stagnation to a value of 0.50, it exhibits a completely different behavior for the initial state $|+\rangle\langle +|$ where it increases with time except for the specific values $\theta = 0, \pi$ and 2π where it stays null. A null trace distance means that the two states it compares are the same. In our case, $D_{ns}(|+\rangle\langle +|)(t,\theta) = 0$ means that for some values of theta (i.e., $0, \pi, 2\pi$) the impact of both measurement and dissipation on the $|+\rangle\langle +|$ state evolution is null.

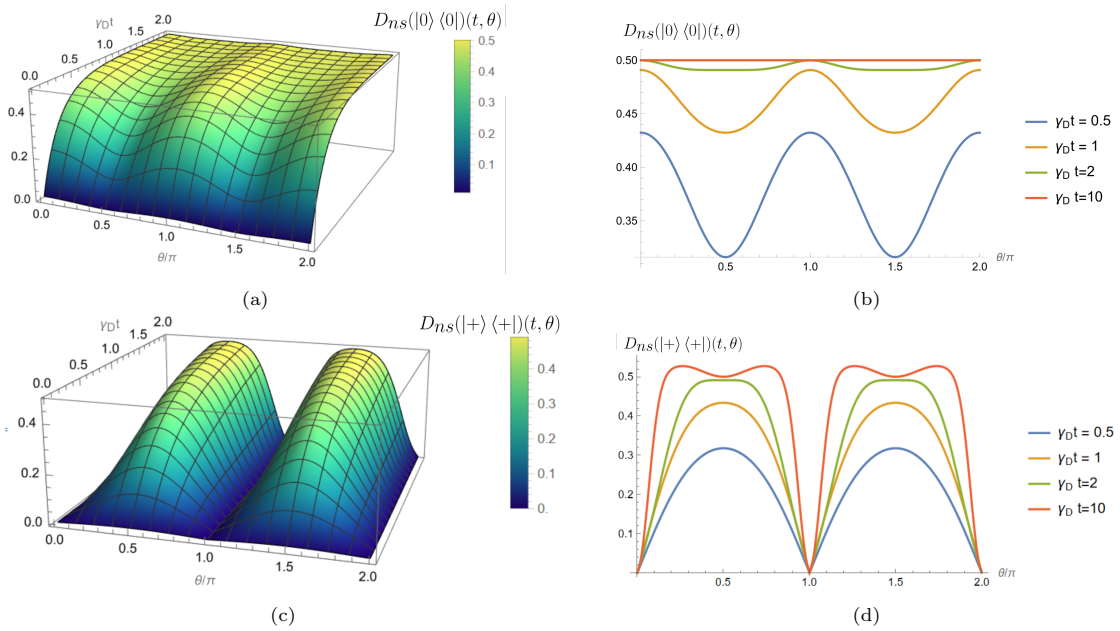


Figure 3.6: (a) $D_{ns}(|0\rangle\langle 0|)(t,\theta)$ as a function of the measurement basis (θ) and the time of evolution t. (b) $D_{ns}(|0\rangle\langle 0|)(t,\theta)$ as a function of the measurement basis (θ), for different values of t. (c) and (d) are the same graphs as (a) and (b) but for the initial state $|+\rangle\langle +|$. The larger the time t is, the less impact θ has for $|0\rangle\langle 0|$. Indeed, for $t = 10$ (red curve in (b)) the trace distance is totally independent of the angle. However for $|+\rangle\langle +|$ we observe the opposite behavior, the trace distance is always null for θ values of $0, \pi$ and 2π while it grows larger with time for other values. The parameters ω and γ_E are set to $\omega = \gamma_E = \gamma_D$.

3.5.1 Effect of measurement only

To analyze in more detail the effect on the photons states of the spy measurement alone, we computed the trace distance between the time evolution of an initial state subject to only dissipation and the time evolution of this same state subject to dissipation and measurement, i.e. $D_{nsn}(.)(t,\theta)$. Therefore we obtain the trace distance corresponding to the effect of the measurement only, on the considered state. The results for the states $|0\rangle\langle 0|$ and $|+\rangle\langle +|$ are respectively plotted in the panels (a), (b), (c) and (d) of Figure 3.7.

The trace distances $D_{ns}(|+\rangle\langle +|)(t,\theta)$ and $D_{nsn}(|+\rangle\langle +|)(t,\theta)$ of the panels 3.7c and 3.6c are identical, meaning that whether we consider the trace distance between the time evolved state (with measurement and noise) and the initial state $|+\rangle\langle +|$ or its time evolution under dissipation only, the results are the same. This behavior was expected, since the state $|+\rangle$ is an eigenstate of the dissipation operator $\hat{\sigma}_x$, the noise does not affect it and the state thus stays equal to its initial value, causing the two trace distances to be equal. The trace distance is maximum for measurement angles $\theta + \frac{\pi}{2}$ and $\theta = \frac{3\pi}{2}$ and minimum for 0, π and 2π which makes sense given our definition of the measurement operator:

$$\hat{c} = \cos(\theta)\hat{\sigma}_x + \sin(\theta)\hat{\sigma}_z. \tag{3.47}$$

Indeed, for values of $0 + k\pi$ ($k \in \mathbb{Z}$), $\hat{c} = \pm\hat{\sigma}_x$, which means the measurement does not affect the state. On the other hand, when $\theta = \frac{\pi}{2} + k\pi$ ($k \in \mathbb{Z}$), $\hat{c} = \pm\hat{\sigma}_z$ and the state is maximally affected.

For the initial state $|0\rangle\langle 0|$, the trace distance $D_{ns}(|0\rangle\langle 0|)(t,\theta)$ (panel 3.6a) increases with time until 0.5 where it stabilizes. On the other hand $D_{nsn}(|0\rangle\langle 0|)(t,\theta)$ (panel 3.7a) is maximum,

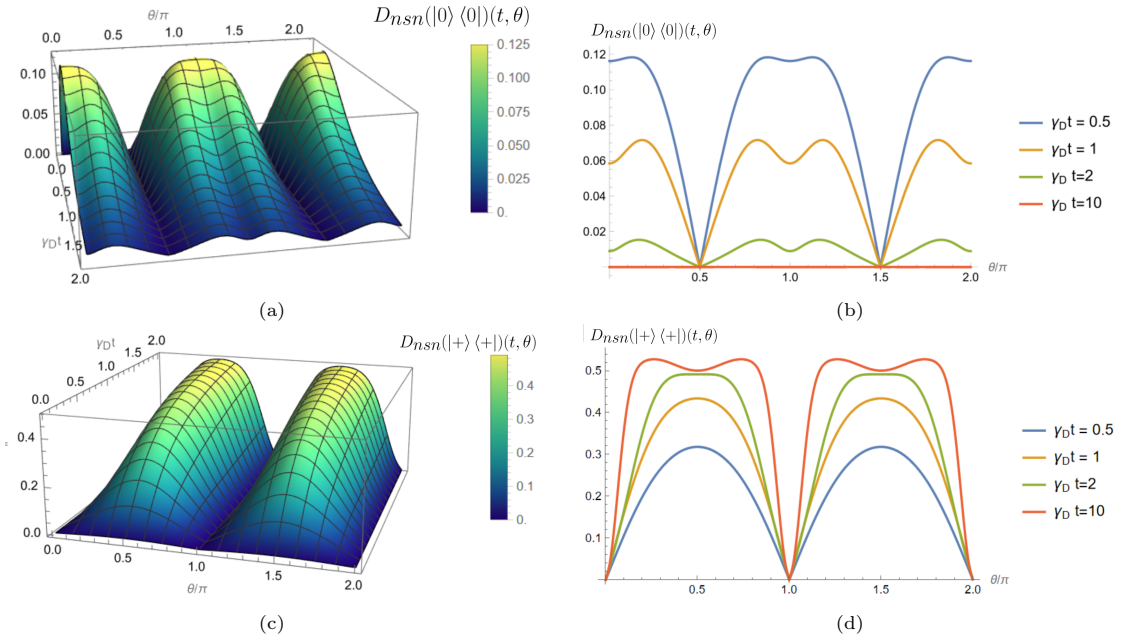


Figure 3.7: (a) $D_{nsn}(|0\rangle\langle 0|)(t,\theta)$ as a function of the measurement basis (θ) and the time of evolution t. (b) Same trace distance as in (a), as a function of the measurement basis (θ), for different values of t. (c) and (d) are the same graphs as (a) and (b) but for the $|+\rangle\langle +|$ initial state. The larger the time t is, the less impact θ has for $|0\rangle\langle 0|$. Indeed, for $t = 10$ (red curve in (b)) the trace distance is totally independent of the angle. However for $|+\rangle\langle +|$ we observe the opposite behavior, the trace distance is always null for θ values of $0, \pi$ and 2π while it grows larger with time for other values. The parameters ω and γ_E are set to $\omega = \gamma_E = \gamma_D = 1$.

for a small time t, at $\theta = 0, \pi$ or 2π where it reaches 0.10 while it tends to 0 for $\theta = \frac{\pi}{2}$ or $\frac{3\pi}{2}$. These behaviors are coherent with our definition of $\hat{c}$: $|0\rangle$ is an eigenstate of the operator $\hat{c} = \pm\hat{\sigma}_z$ ($\theta = \frac{\pi}{2} + k\pi$) while it is not for $\hat{c} = \pm\hat{\sigma}_x$ ($\theta = 0 + k\pi$).
Since for both of these trace distances the bigger t is, the less θ has impact, and D_{nsn} is at most equal to 20% of D_{ns}, we can conclude that the process generating the most perturbations on this state is the dissipation.

3.6 Summary

In this Chapter, we obtained that the probabilities of Bob measuring the photon state Alice initially sent were $\frac{3}{4}$ with Eve performing projective measurements at the middle of the channel, $\frac{e^{-2\gamma t}}{4} + \frac{3}{4}$ with a $\hat{\sigma}_x$-modeled noise only, and $\frac{e^{-2\gamma t}}{8} + \frac{5}{8}$ with both noise and eavesdropping. We also obtained that the angles minimizing the impact on the photons of homodyne measurements, when the corresponding operator was defined as $\hat{c} = \sin\theta\hat{\sigma}_z + \cos\theta\hat{\sigma}_x$, were $k\frac{\pi}{2}$ for the initial states $\{|0\rangle, |1\rangle\}$, and $0 + k\pi$ for $\{|+\rangle, |-\rangle\}$. These last results will be further used in Chapter 5.

After computing the impact of measurement on the photons, the question arising is "How does the spy harness these measurements results to obtain information about the shared secret key ?", which we will try to answer in the next Chapter.

Chapter 4

Quantum state tomography and neural networks

Now that the impact of weak measurements on the photons of the protocol has been assessed, we must examine how such measurement outputs could be used to recover the initial state Alice sent in the optical fiber. This task is called quantum state tomography (QST).

In this Chapter we first explore standard QST, and motivate the possible use of machine learning for such task. Then we develop the theoretical tools needed to understand and implement recurrent neural networks to perform QST and we analyze the results (i.e., the accuracy) achieved with this approach.

4.1 Standard tomography

Quantum state tomography (QST) is the process of reconstructing the quantum state of a system from repeated measurements of a set of observable. This set must be complete and it thus requires a number of copies of equally prepared quantum systems. In [27], D'Ariano and Yuen showed the impossibility of any measurement scheme for determining the wave function from a single copy of the system. They reviewed a variety of concrete measurement schemes based on vanishingly weak quantum non-demolition measurements [28], weak measurements on "protected" states [29], "logically reversible" [30], and "physically reversible" [31, 32] measurements. The conclusion of these reports is that it is practically impossible to measure the wave function from a single copy of a system. Indeed, in [28] the weakness of the measuring interaction prevents one from gaining enough information on the wave function, in [29] the method of protecting the state requires some *a priori* knowledge on the state, and in [32] quantum measurements can be physically reverted only with a probability of success equal to $\frac{1}{2}$.

In addition more recent works on tomography all use a set of measurements, and could not reconstruct the initial state using only one measurements or one copy of the system. These include Plain averaging or Maximum Likelihood methods [33], direct inversion, distance minimization, maximum likelihood estimate with radial priors and Bayesian mean estimate [34]. In other works such as [35], the photonic state tomography of a single qubit has been studied but as is said by Alteper *et al.*: "Exact single-qubit tomography requires a sequence of three linearly independent measurements.". [36] explores Bayesian Homodyne and Heterodyne tomography, in which the measurements must be repeated K times. Finally, in the context of QST, without the

measurement of a complete set of observables (a quorum), there is not enough information for the reconstruction as different states may give the exact same statistics on an incomplete set of observables [37]. For the state of a single mode of the radiation field, the quorum typically used is composed of the field quadratures, which can be measured one by one using homodyne detection.

In conclusion, there is no standard tomography technique that works for a single homodyne measurement on a photon. In the following sections, we will therefore attempt to tackle this problem using the capacities of neural networks.

4.2 Neural networks

The fundamental idea behind neural networks is the threshold logic unit [38] which models the biological neuron. Assuming boolean inputs, it is defined as

$$f(x) = \text{sign}\left(\sum_i w_i x_i + b \geq 0\right), \tag{4.1}$$

and work as follows: there is a weight w_i assigned to each input x_i defining its importance among other inputs, and a threshold b (also called bias) which will define if the sign function outputs 1 ($w_i x_i \geq -b$) or 0 ($w_i x_i \leq -b$).
The generalization to real inputs is called the perceptron [39] and is the building block of all neural networks. Indeed, this logic unit can be composed in parallel to form layers, which can be composed in series to form a multi-layer perceptron, also called artificial neural networks.

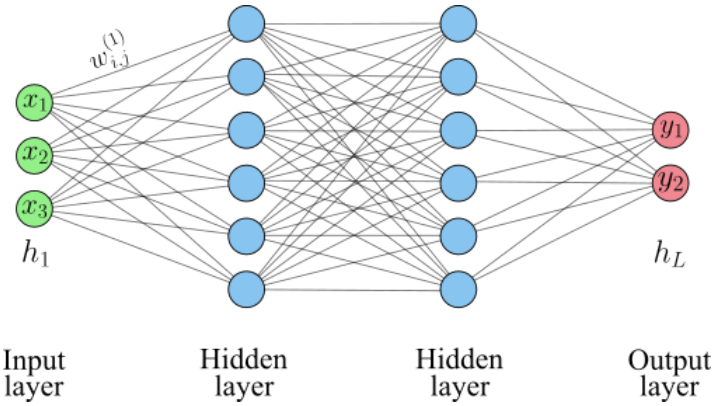


Figure 4.1: Fully connected multi-layer perceptron composed of L layers (here $L = 4$): input (h_1), 2 hidden layers (h_2 and h_3) and output (h_L). The ith input is denoted by x_i, the weight between the neuron i of layer l and the neuron j of layer $l+1$ by $w_{i,j}^{(l)}$ and the ith output by y_i.

The main purpose of neural networks is, from a data set $\mathbf{d} = \{(\mathbf{x}, \mathbf{y})\}$ (pairs of input-ground truth), to learn dependencies between the input variables $\mathbf{x}$ so as to be able to make a prediction $\hat{\mathbf{y}}$ as close as possible to the ground truth $\mathbf{y}$. The latter can either be a class (classification problem) or a function to approximate (regression problem). The bold notation is used here to designate a vector, as it is a standard in machine learning.

As illustrated in Figure 4.1 artificial neural networks are built as follows: there is an input layer of neurons, hidden layers and an output layer which are all partially or fully connected. Each connection between two neurons has a weight defining how strongly the output of the first

one will be taken into account in the second. Indeed, we notice in Eq. (4.1) that the output of each neuron is multiplied by the weight, then fed to the neurons of the next layer it is connected to. Therefore, a small weight is equivalent to a small importance of the neuron output.

Each layer has a non linear activation function (sigmoid, ReLU, hyperbolic tangent) which, for each neuron of the layer, takes as input the weighted output of the neurons it is connected to. The network has a loss function noted $\mathcal{L}$ which defines how far away its prediction is from the true output. The choice of a loss function is highly problem-dependant, whether it is a regression or classification problem, multi-class or binary, etc. Among the most popular loss functions are:

- Mean square error: $\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

  This error is used for regression problems. Where y_i is the ith ground truth and $\hat{y}_i$ the ith prediction.

- Mean absolute error: $\mathcal{L}_{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$.

  which is also used for regression problems.

- Cross-entropy : $\mathcal{L}_{CE} = -\sum_{i}^{n} t_i log(p_i)$.

  This error is used for multi-class classification problems. Where t_i is the ground truth class and p_i is the Softmax probability for the i^{th} class. The Softmax activation function is used to transform a vector of k real numbers into a probability distribution over k choices (i.e. classes here).

When training a neural network we aim at minimizing this loss function so that the network predictions on are as accurate as possible during training. Nevertheless, this loss does not represent the error on data the network did not see during training. It is therefore good practice to monitor the error on an independent data set (test set) during training, to prevent it from being too specialized with respect to the true data generating process and thus have a bad generalization error. In other words the objective is to prevent the network from over fitting. In general, the loss function do not admit a minimizer that can be expressed analytically in closed form [40]. However it can be found numerically using a general optimization technique such as gradient descent, which uses local linear information to iteratively move towards a local minimum. Since the true loss function can be problematic to minimize, we locally approximate it as a parabola around the considered point θ_c, as illustrated in Figure 4.2:

$$\hat{\mathcal{L}}(\epsilon; \theta_c) = \mathcal{L}(\theta_c) + \epsilon^T \boldsymbol{\nabla}_\theta \mathcal{L}(\theta_c) + \frac{1}{2\gamma}\|\epsilon\|^2, \tag{4.2}$$

where γ is the learning rate, ϵ is the distance between the point of the parabola we consider and the point where we want to approximate the loss (i.e. θ_c), and θ_t denotes the parameters of the network (e.g. the weights w and bias b) at time t of the training. Indeed, when training the network we update its parameters at each step, thus their value will change over time but remain fixed once the training is finished. The parabola approximating the loss can be minimized with respect to ϵ as:

$$\boldsymbol{\nabla}_\epsilon \hat{\mathcal{L}}(\epsilon; \theta_0) = \boldsymbol{\nabla}_\theta \mathcal{L}(\theta_0) + \frac{1}{\gamma}\epsilon = 0, \tag{4.3}$$

giving the best update for the step

$$\epsilon = -\gamma \boldsymbol{\nabla}_\theta \mathcal{L}(\theta_0). \tag{4.4}$$

We thus have an update rule for the parameters:

$$\theta_{t+1} = \theta_t - \gamma \boldsymbol{\nabla}_\theta \mathcal{L}(\theta_t), \tag{4.5}$$

where θ_0 are the initial parameters of the network.

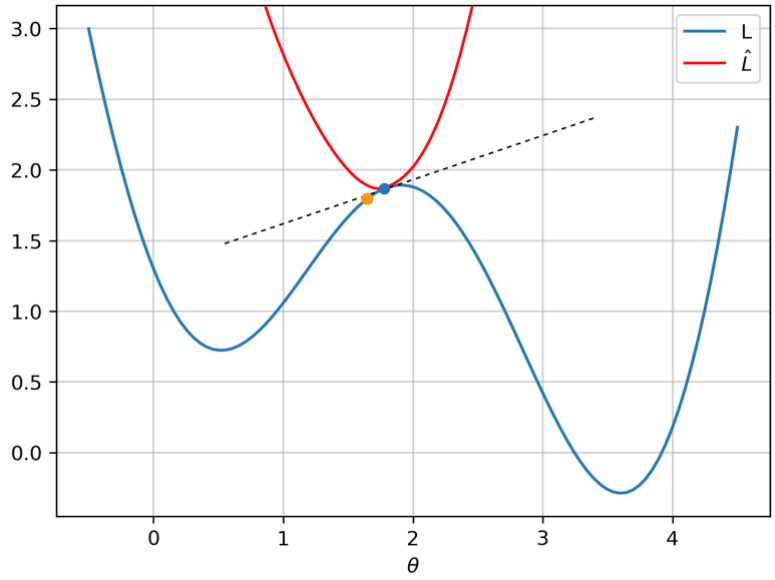


Figure 4.2: Illustration of a gradient descent step. If it is the first of the training, the point we consider is $\theta_c = \theta_0$ (i.e. the initial parameters), represented by the blue point. At this point we approximate the true loss function (blue curve) by a parabola (red curve) following Eq. (4.2). Then we compute the gradient with respect to θ (dotted line), which we multiply by the learning rate before updating the parameters. This process yields the new parameter value θ_1 (orange point), and will be repeated until convergence to a minimum.

In general the parameter we update is the weight matrix, as depicted in Eq. (4.5) this update is proportional to the partial derivative of the loss function with respect to the current weights ($\equiv$ gradient). In order to minimize $\mathcal{L}(\theta)$ we must compute all of its partial derivatives with respect to θ, which can be computed efficiently using the backpropagation algorithm. Indeed, since a neural network is a composition of the differentiable activation functions of each layer, the total derivatives of the loss can be evaluated backward, by applying the chain rule recursively. The algorithm can be summarized in 3 steps [41]:

1. Compute inputs ($z_i^{(l)}$) and outputs ($a_i^{(l)}$) for all neurons using

$$a_i^{(l+1)} = f^{(l+1)}\left(b^{(l)} + \sum_{j=1}^{s_l} w_{i,j}^{(l)} a_j^{(l)}\right), \text{and} \qquad z_i^{(l)} = b^{(l-1)} + \sum_{j=1}^{s_{l-1}} w_{i,j}^{(l-1)} a_j^{(l-1)}. \tag{4.6}$$

2. Compute $\delta_i^{(l)}$ for all neurons using

$$\delta_i^{(l)} = \left(\sum_{j=1}^{s_{l+1}} \delta_i^{(l+1)} w_{i,j}^{(l)}\right) f'(z_i^{(l)}). \tag{4.7}$$

   This equation is recurrent since we need to know the δ of the higher layer $(l+1)$ to compute the one of the layer l.

3. Compute

$$\frac{\partial \mathcal{L}}{\partial w_{i,j}^{(l)}} = \delta_i^{(l+1)} a_j^{(l)}, \tag{4.8}$$

which are the partial derivatives of the loss with respect to all the network parameters (i.e., weights), allowing to perform a single update. This process is repeated until convergence of the loss. The first step is called *forward propagation* while the second is called *backward propagation*. The training of a neural network can be in batch mode or in online mode, the first mode refers to the training samples being fed in groups (e.g. 50 at a time) to the network and thus the weight update is a mean on the samples from the batch. On the other hand, online mode designates the samples being fed one at a time to the network. In both cases, one update of the parameters is called an *iteration*, one sweep over all training examples is called an *epoch*

4.2.1 Training a neural network

When we train a neural network, no matter how complex the model is, we always go through a training loop. In this loop we feed the data to the model and get its predictions. We then compare the predictions of the network to the ground truth with the chosen loss function, and adjust the parameters of the model by performing gradient descent. During the training stage we would like to keep track whether our model will improve over the different iterations. It is therefore good practice to monitor whether the loss we are minimizing decreases over time, and whether the overall performance of the model increases the more training iterations we perform.

Once the data is fed to the model we can obtain its predictions, which is usually called the forward pass. Once predictions are obtained, we can compare how close they are to the ground truth, by feeding them together with the true labels through the loss function we are minimizing. At its early training stages the network will perform poorly, but it will improve as its weights are updated by gradient descent. We can do this very easily by obtaining the gradients of the loss function we are minimizing with respect to the weights (backward pass) and adjusting these. Finally, we can measure the performance of our model by evaluating its accuracy on an independent test set.

Even though neural networks can tackle efficiently a wide variety of regression or classification problem, it can be computationally expensive to train them and it might require a very large dataset, which is not always easily accessible. In many cases, simpler machine learning models might offer comparable performance with less computational overhead. Also, there exists various refinements such as recurrent neural networks, long short-term memory, Boltzmann machine or convolutional networks, designed to be more problem-specific as we detail in the next subsections.

4.2.2 Recurrent Neural Network (RNN)

Recurrent neural networks are special types of neural networks adapted to handle time series data or data involving sequences, by sharing parameters between time steps. In fact, in traditional deep neural networks, we assume that inputs are independent of each other and so are the outputs, while in recurrent neural networks, the outputs depend on the prior elements within the sequence, the network acting with something similar to a memory [42]. Their capacity to process any sequence of inputs using internal state (memory) makes them specialized in solving problems that involve time series data such as connected handwriting recognition or speech recognition. Considering an input sequence $\mathbf{x}^{(t)}$, the internal state of the RNN at the time step

t is

$$\mathbf{h}^{(t)} = f\left(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}\right). \tag{4.9}$$

Indeed, The RNN consists of a unit cell that is repeated at every new input of the time-series data $\mathbf{x}^{(t)}$, producing an output $\mathbf{h}^{(t+1)}$ known as the hidden state. This hidden state is then combined with the next time-series input $\mathbf{x}^{(t+1)}$, allowing information to propagate through the sequence and have an impact on the outputs at future times. As shown in Figure 4.3, a recurrent neural network can be represented in a folded or an unfolded way (respectively left and right on the Figure). The first representation allows to better visualize the recursive definition and the sharing of the parameters (i.e., weights and activation functions f) throughout the time steps.

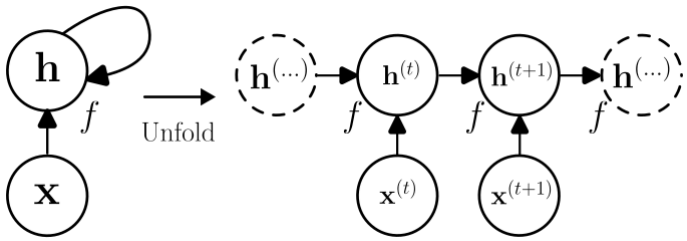


Figure 4.3: The same recurrent neural network with no output, seen as a folded (left) and an unfolded (right) computational graph.

4.2.3 Long Short-Term Memory network

The problem we aim to resolve, as explained in previous sections, is to recover the initial state sent by Alice by inferring it from the homodyne photo-current J_{hom} measured on the system. Therefore we have a classification problem with each class corresponding to one of the four initial states, and the input data (J_{hom}) is represented as time-series. That is why among all the existing neural networks architecture, the most interesting one for our purpose is the recurrent neural network (RNN) and more precisely the long short-term memory (LSTM), invented to resolve the vanishing gradients problem.

Studied in 1991 by Sepp Hochreiter [43], the vanishing gradient problem can be summarized as the update to the weights of the network (done at each iteration) becoming vanishingly small and thus preventing the weights values from changing at each time step. This constitutes a major problem in a RNN since it is completely stopping its learning process. To overcome this issue, Sepp Hochreiter and Jürgen Schmidhuber invented the Long short-term memory network [44]. To retain valuable long-term dependencies for prediction-making in both present and future time-steps, the Long Short-Term Memory network (LSTM) selectively outputs pertinent information from the current state. In fact, an input gate, an output gate, a forget gate, and a cell state make up a typical LSTM recurrent unit as shown in Figure 4.4. The three gates control the information flow into and out of the cell state $\mathbf{c}^{(t)}$ to retain values for arbitrarily long periods of time. Forget gates use a value between 0 and 1 to indicate which information from a prior state should be discarded in relation to the present input. To retain the information, a (rounded) value of 1 is indicated, and to discard it, a value of 0. Using the same mechanism as forget gates, input gates determine which new pieces of information to store in the existing state. By allocating a value, output gates regulate which bits of data in the present state are output in the hidden

state. The presence of a cell state $\mathbf{c}^{(t)}$ is exclusive to the LSTM networks and is what carries its long-term memory property, while the hidden state $\mathbf{h}^{(t)}$ is common to LSTMs and RNNs.

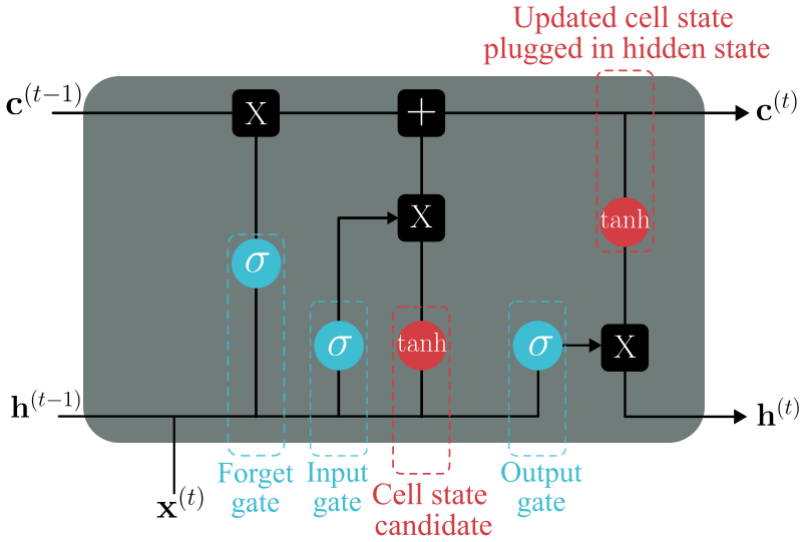


Figure 4.4: One LSTM recurrent unit, composed of three gates with sigmoid activation functions (forget, input and output). These 3 gates determine which information from the prior hidden state must be erased, taken into account and stored respectively.

4.3 Recurrent neural network tomography

The situation we simulate in this Section is the following: the spy Eve has access to the optical fiber through which Alice photons travel, and can perform a continuous homodyne measurement on these. To determine the initial state of each photon she measures, she has beforehand trained a LSTM neural network to reconstruct the state starting from the measured photo current. To do so she has access to a QKD emitter, a device that can create and send single polarized photons through an optical fiber. In this Section we first present the architecture and hyper-parameters used for the neural network, the construction of the dataset, and then the results that Eve could obtain with it.

4.3.1 Architecture

The network input layer is a LSTM with 40 units in its hidden state. We have opted for a hidden layer composed of 40 neurons, with a rectified linear unit (ReLU) activation function which is defined as

$$\mathrm{ReLU}(x) = \max(0, x). \tag{4.10}$$

This function is designed to output zero if the input is negative and the input otherwise. This hidden layer is connected to a dense hidden layer of 20 sigmoid neurons, itself connected to a dense output layer of four neurons, meaning every output neuron is connected to all the previous hidden layer ones. Since our problem is to classify each photo current in one of the four existing class (i.e. the four initial states), we used the sparse categorical cross entropy loss function which

is commonly used for classification problems with more than two classes. Cross entropy is defined as

$$\mathcal{L}_{CE} = -\sum_{i}^{N} t_i log(p_i), \tag{4.11}$$

where t_i is the true label and p_i is the Softmax probability for the i^{th} class. Sparse categorical cross entropy is built on the same equation except that the outputs of the network are encoded as integers depending on the predicted class (going from 0 to 3 in our case).

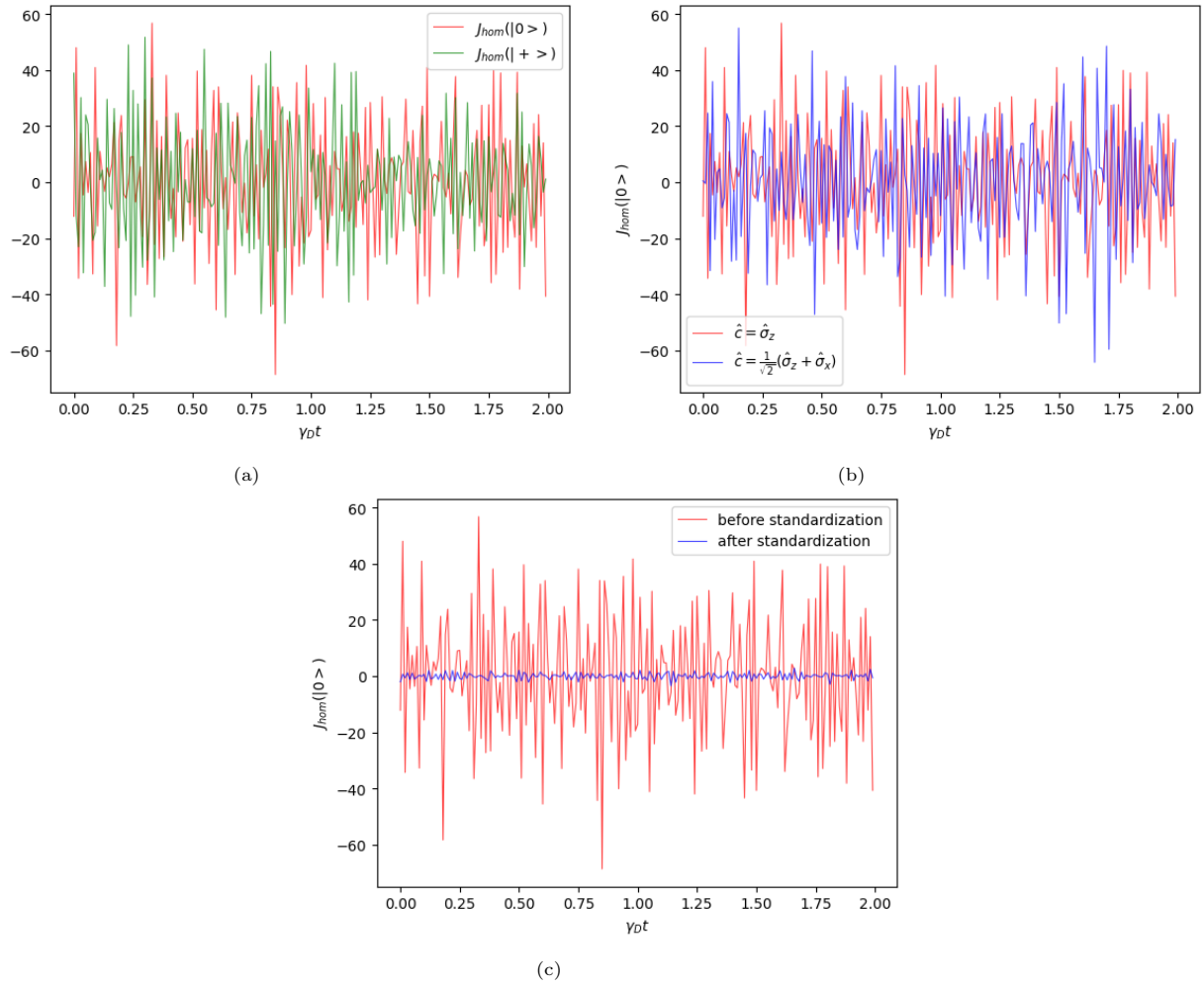


Figure 4.5: (a) Measured homodyne photo currents for the two initial states $|0\rangle$ (red) and $|+\rangle$ (green) as a function of $\gamma_D t$. The operator used to model the spy measurement is the pauli z operator $\hat{\sigma}_z$. (b)Measured homodyne photo currents for the initial state $|0\rangle$ as a function of $\gamma_D t$. The operators used to model the spy measurement are the pauli z operator $\hat{\sigma}_z$ (red) and the Breidbart one $\frac{1}{\sqrt{2}}(\hat{\sigma}_z + \hat{\sigma}_x)$ (blue). These two graphs demonstrate that the presence of noise in the measured photo current remains unaffected by changes in either the initial state selection or the measurement operator utilized by the spy. (c) Homodyne photo currents for the initial state $|0\rangle$, original (red) and after standardization (blue). The parameters ω and γ_E are set to $\omega = \gamma_E = \gamma_D$.

4.3.2 Data set

The data set used to train the neural network is composed of 10^5 photo currents each associated to the initial state of the photon, before its propagation in the optical fiber, randomly drawn from one of the four possible states. Each photo current has a length of 200 time steps, which are set to $\gamma_D dt = 0.01$, thus the total duration is $\gamma_D t = 2$.

We divided this data set into a learning set and a test set as 70% and 30% of the whole dataset respectively. Finally, the batch size was set to 50 and the number of training steps to 3000, i.e. the network was trained on 3000 batches of 50 samples randomly drawn from the learning set.

As is depicted in Figure 4.5, the input data is noisy, to reduce the impact of noise on the model (i.e. the network) we standardized the input data. The purpose of standardization is to put everything on the same scale, it works as follows:

$$\underline{x} \leftarrow \frac{(\underline{x} - \tilde{\mu})}{\tilde{\sigma}}, \tag{4.12}$$

where $\underline{x}$ is the input data (vector form), $\tilde{\mu}$ and $\tilde{\sigma}$ are, respectively, the estimated mean and standard deviation of the data. Standardized data values will be much closer while still having the same distribution and proportion to each other. Also, they are more easily interpretable, for example a standardized value $x = 2$ means that this observation lies two standard deviations above the mean. The effect of this process is displayed in Figure 4.5c.

The inherent noise of the data (i.e. the photo currents) as can be seen in Figure 4.5, is the main reason that led us to the use of neural networks.

4.3.3 Results

The training presented in Section 4.1 yields the losses depicted in Figure 4.6, where the test loss not being above the training loss shows that the model does not overfit the data, which is a good indicator of the network's ability to generalize to data it has not yet seen.

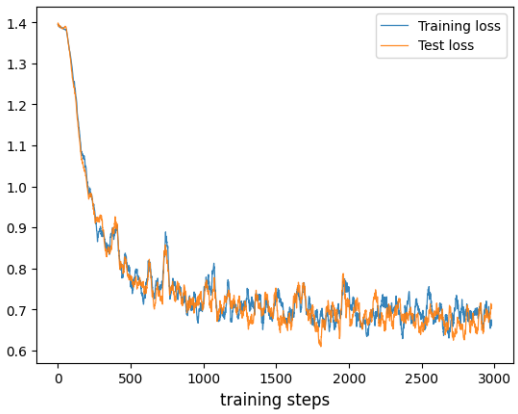


Figure 4.6: Sparse categorical cross entropy losses evolution during training, where the loss is averaged over 20 batches. The blue curve is the training loss while the orange one is the test loss, which means at each step of the training the network makes a prediction on a batch of the test set and this prediction is evaluated with respect to the true output.

For the case where the spy weak measurement is modeled with the operator $\hat{\sigma}_z$, the network predicts the correct initial state 75% of the time. This result is logical since the quantum states

$|0\rangle$ and $|1\rangle$ are eigenstates of this operator, which means Eve's measurement does not affect these two states and thus cannot extract any information differentiating them using $\hat{\sigma}_z$. However, when we model the same measurement with the generic operator

$$\hat{c} = \cos(\theta)\hat{\sigma}_x + \sin(\theta)\hat{\sigma}_z. \tag{4.13}$$

We can evaluate the accuracy of the system as a function of θ as is done in the Figure 4.7.

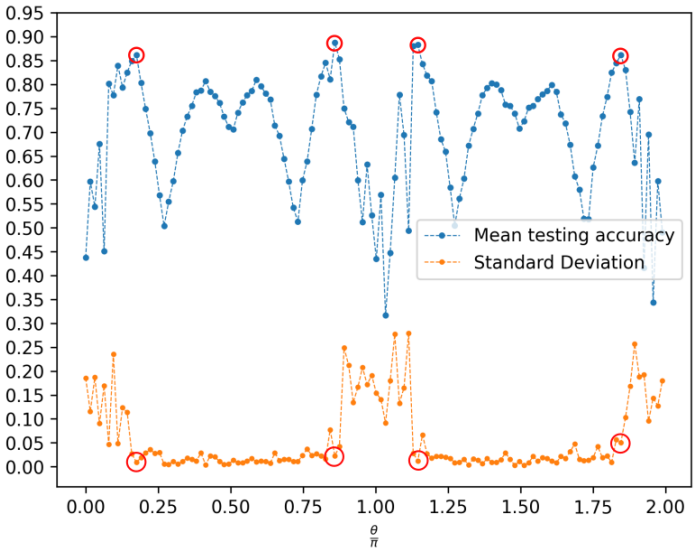


Figure 4.7: Mean estimated accuracy (blue) and standard deviation (orange) of the model, on the test set, as a function of $\frac{\theta}{\pi}$. Despite some noise in the values, the accuracy is symmetrical with respect to $\theta = \pi$, as expected. Circles in red are the four values of θ maximizing the test accuracy and the corresponding standard deviations.

This Figure shows the estimated mean accuracy (blue) and standard deviation (orange) of the neural network as a function of the measurement angle θ. There are preferential measurement bases for the spy Eve, i.e. there are angles θ maximizing the information obtained from the weak measurement. It also shows that the test accuracy is a noisy but symmetrical function of θ with respect to the value $\theta = \pi$. There are four main values of the measurement angle leading to a test accuracy between 85% and 90%, as summarized in Table 5.1. The measurement angles

θ	mean accuracy	standard deviation
0.17507π	0.86164	0.00922
0.85944π	0.88794	0.02285
1.14592π	0.88265	0.01217
1.84620π	0.86167	0.05026

Table 4.1: Mean accuracy of the neural network and corresponding standard deviations for different values of theta.

around $\theta = 0$, $\theta = \pi$ and $\theta = 2\pi$ lead to accuracies barely above the one obtained from random guess (i.e. 25%), which means that the network could only obtain few useful information from the photo currents. The neural network is also unstable for these angles, which correspond to accuracies with high standard deviation (up to 0.30). However the angles of Table 5.1 (circled

in red in Figure 4.7), yield high accuracies with relatively low standard deviations. The spy can thus predict the initial state with a high certainty while keeping the same measurement angle, where the network is stable.

4.4 Summary

In this Chapter we figured out that it is impossible for a spy to perform traditional quantum state tomography using a single homodyne photo current. However, using the powerful capacities of a recurrent neural network and assuming that the spy could train it beforehand using a QKD emitter, initial state tomography is possible. Indeed, with the measurement operator $\hat{\sigma}_z$ the neural network achieves an accuracy of 75%, and this accuracy can be maximized to 88.265% using a measurement angle $\theta = 1.14592\pi$. This angle will be called the optimal measurement angle for the rest of this master thesis.

We have obtained a way to maximize the extracted information of the measurement, however the metric we use is not formal. In addition, intuition tells us that the more information is extracted from the photons by a measurement, the more perturbation is induced, just like a projective measurement yields the most information about a state (i.e., the state itself), but completely changes the state if the basis is not appropriate. In the next sections, we will consider the trade-off between extracted information and impact of the measurement.

Chapter 5

Photon state integrity and information gain

In the previous Chapters we have determined the impact of homodyne measurements on the photon states of the BB84 protocol and the achievable accuracy of RNN tomography. Therefore the relationship between the amount of information a measurement extracts and the impact that it yields must be investigated.

In this Chapter we first consider a more formal way to describe the information gain from a measurement. Then relate the extracted information to the average impact on the initial states, to finally obtain the optimal trade-off. We also try to make the measurement scheme more realistic and then optimize it.

5.1 Extracted information

In the previous section, we characterized the amount of extracted information from the system using the neural network accuracy as metric. However there is a more formal way to quantify extracted information, using the von Neumann entropy.

In 1948, Claude Shannon introduced a way to quantify the information contained in a message, or at least that some message contained more information than another one. Indeed, a message provides information if it reveals, among a set of possibilities, which one have occurred. Formally, the Shannon entropy of a given probability distribution p_i is defined in [45] as

$$H\left[\{p_i\}\right] \equiv -\sum_i p_i \ln p_i. \tag{5.1}$$

It quantifies the information that a random variable having the probability distribution p_i contains. To better understand what the entropy means, we can look at the simple example of sending English texts encoded in bits. Each letter of the alphabet sent is sampled from a random distribution that can be determined by analyzing a huge dataset of typical English texts. We assume that each letter is independent of the next and the previous ones, even if we know it is not perfectly true since some letters (e.g., "h") appear more frequently after some (e.g., "t") than after others (e.g., "z"). In the case of sending N letters, we must send on average H bits per letter, where H is the entropy of the distribution (relative frequency) of English letters [45].

The amount of communication resources (bits) required to inform us about something is also a measure of our initial uncertainty about that thing. Indeed, the more we know about the state

of a system, the less uncertain we are, and the less information we need to fully know this state. Therefore, the entropy can be understood as measuring both the uncertainty about something, and the amount of information that it "contains".

The Shannon entropy measures the uncertainty associated with a classical probability distribution [23], but we know that quantum states can be described with density operators, replacing probability distributions. We can thus generalize Shannon entropy to quantum states by defining the Von Neumann entropy of a quantum state $\hat{\rho}$ as

$$S(\hat{\rho}) \equiv -\mathrm{Tr}\left[\hat{\rho}\ln\hat{\rho}\right]. \tag{5.2}$$

It can be understood as the minimum uncertainty that we have about the future behavior of a quantum system. If we make a measurement giving complete information about the system then the von Neumann entropy is the minimum possible entropy of the measurement outcomes.

Now that we have defined a measure of the uncertainty about the state of a quantum system, we can define the information gain (or uncertainty reduction) that a measurement provides us with. This information gain, denoted ΔI, is defined as the average reduction in the von Neumann entropy, and is thus given by the difference of the entropy of the averaged state and the mean entropy of the conditioned state

$$\Delta I(t) = S(E\left[\hat{\rho}(t)\right]) - E\left[S(\hat{\rho}(t))\right]. \tag{5.3}$$

It is also called *Groenewold's information* as he was the first to consider it [46]. Since the second term is an average of the entropy of conditioned states (i.e., it is a property of the full ensemble and not just its mean), we cannot express it as a function of the averaged state, and its evaluation thus requires many different realizations of the stochastic master equation describing the evolution of $\hat{\rho}$. By "different realizations" we mean here many simulations of the evolution of each possible initial state. On the other hand, the first term is just the entropy of the average state, and thus require only one simulation, for each initial state, of the deterministic master equation

$$d\hat{\rho}_J(t) = -\mathrm{i}\left[\hat{H}, \hat{\rho}_J(t)\right]\mathrm{d}t + \gamma_E\mathcal{D}[\hat{c}]\hat{\rho}_J(t)\mathrm{d}t + \gamma_D\mathcal{D}[\hat{v}]\hat{\rho}_J(t)dt, \tag{5.4}$$

obtained by averaging Eq. (1.53) over all possible measurement results. The information gain $\Delta I(t)$ from homodyne measurement using different measurement operators is displayed in Figure 5.1. We see that the information extracted using $\theta = 1.14592\pi$ is the highest, which confirms the fact that this angle extracts more information than $\theta = \frac{\pi}{2}$ which corresponds to the measurement operator $\hat{c} = \hat{\sigma}_z$.

Even though the information gain is more formal, using the network accuracy directly reflects how much of the private key a spy can retrieve, and the latter will thus be used as metric in the next sections.

5.2 Trade-off between information extraction and state perturbation

As discussed in Section 4.3, there are measurement angles θ that maximize the probability of finding the initial state for a spy. To study the impact of such measurement on the whole protocol, we computed an average of the trace distance $D_{ns}(.)(t,\theta)$ over the four possible initial states, denoted by $\overline{D}_{ns}(t,\theta)$. As illustrated in Figure 5.2b, the angles minimizing the measurement

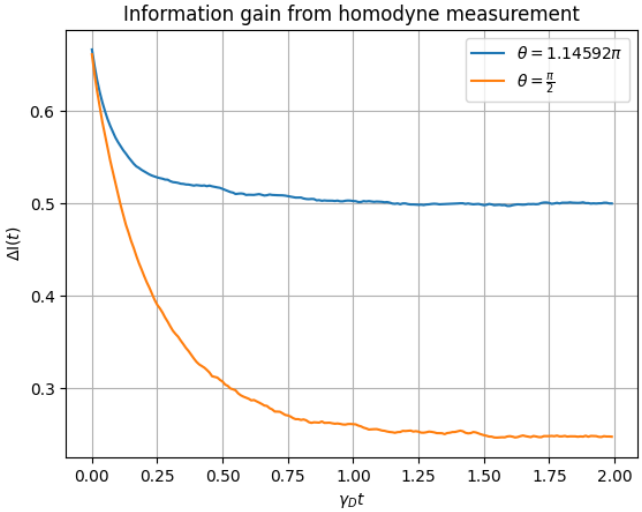
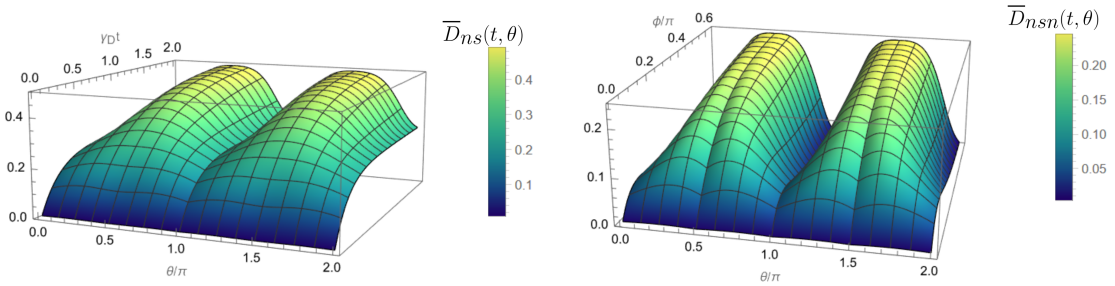


Figure 5.1: Information gain ΔI for the measurement angles $\theta = \frac{\pi}{2}(\hat{c} = \hat{\sigma}_z$, in orange) and $\theta = 1.14592\pi$, the optimal angle derived earlier (in blue). Other parameters are set to $\omega = \gamma_E = \gamma_D = 1$



(a) (b)

Figure 5.2: (a) and (b) Trace distances $\overline{D}_{ns}(t,\theta)$ and $\overline{D}_{nsn}(t,\theta)$ respectively, as a function of the time t and the measurement angle θ. Other parameters ω and γ_E are set to $\omega = \gamma_E = \gamma_D = 1$.

impact on the state are $\theta = 0 + k\pi$ $(k \in \mathbb{Z})$, which corresponds to a measurement modeled by the operator $\hat{c} = \pm\hat{\sigma}_x$.

There is a trade-off between the amount of information we can extract from the photons and the perturbations that this measurement induces. Indeed, the angles that minimize, in average, the trace distance are also minimizing the information the spy obtains, thus minimizing the accuracy of the neural network (Figure 4.7). To better assess this trade-off, a new quantity can be defined as the trace distance plus one divided by the accuracy of the network, for a given measurement basis (i.e. a given θ):

$$\lambda(t,\theta) = \frac{\overline{D}_{nsn}(t,\theta) + 1}{\text{accuracy(t},\theta)}. \tag{5.5}$$

To simplify this expression, the time dimension can be ignored by setting $\gamma_D t = 2$, which yields an expression depending only on θ.

Figure 5.3 is a plot of $\lambda(\theta)$ for values between 0 and 2π. $\lambda(\theta)$ is almost minimized for all

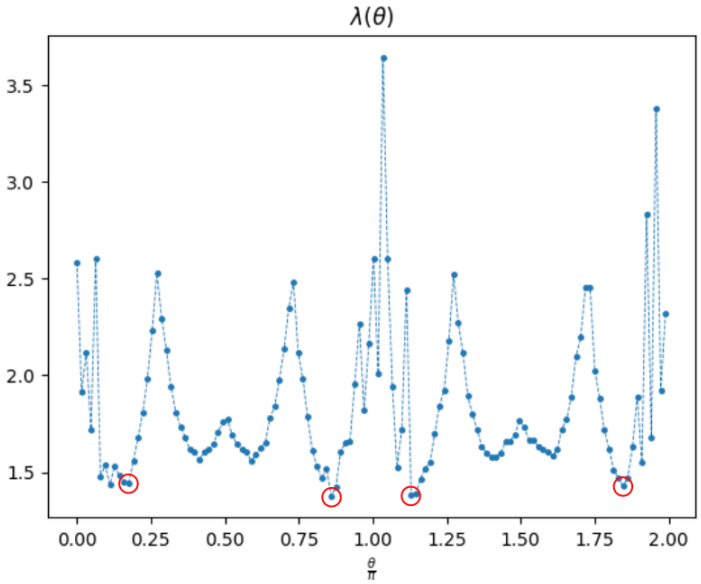


Figure 5.3: $\lambda(\theta)$ for $\theta = 0$ to 2π. Circled in red are the four values of θ that maximize the test accuracy, which also minimize $\lambda(\theta)$.

θ	mean accuracy	$\overline{D}_{nsn}(\theta)$	$\lambda(\theta)$
0.17507π	0.86164	0.24260	1.44213
0.85944π	0.88794	0.22324	1.37761
1.14592π	0.88265	0.22645	1.38951
1.84620π	0.86167	0.23105	1.42867

Table 5.1: Mean accuracy of the neural network and corresponding standard deviations for the values of theta maximizing the test accuracy.

four of the θ values maximizing the test accuracy of the network. The mean accuracies and standard deviations for these angles are summarized in Table 5.1. By choosing one of these four measurement basis the spy could thus achieve a high accuracy when predicting the initial state, while perturbating the photons as less as possible.

5.3 Optimization of the measurement

Until now, we have considered that the homodyne measurement was made during the entire photon travel time. First this is not plausible, the spy could not measure directly from the source (Alice) and until the end (Bob) or it would be detected, i.e. Alice or Bob would see Eve with their own eyes. To solve this problem, the starting time of the measure is set to 10 time steps from now on such that Eve is not too close to the source but she still start measuring at the early stage of the photon evolution, to have access to as much information as possible. Indeed, we deduced from Figure 3.4 and 5.1 that most of the information is contained at the start of the photo currents. Another modification would be to consider that the measurement is much shorter than the travel time (e.g. one quarter) to decrease the impact of the measurement.

In this section, we test the performances of the neural network for different lengths of measurement.

To determine the optimal measurement duration we train the neural network and test it on different lengths of the measured photo currents and compute its average test accuracy as a function of the length (Figure 5.4). The test accuracy reaches a maximum of 85.410% for a length of 40 time steps, with a standard deviation of 0.02183. This maximum is higher than the accuracy when training on the whole photo currents (still sarting at time step 10). The measurement can thus be performed only from time step 10 to 50 while decreasing the accuracy by only 3%.

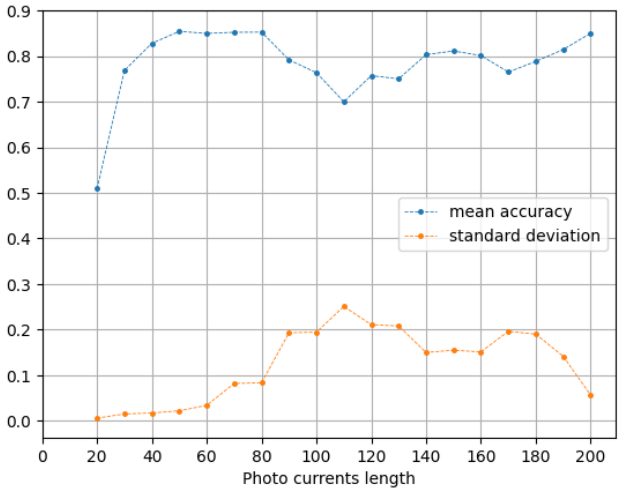


Figure 5.4: Mean accuracy and standard deviation of the network as functions of the measurement length.

5.4 Summary

The most important result of this Chapter, and even of this master thesis, is that the measurement angle $\theta = 1.14592\pi$ optimizes the $\lambda(\theta)$ coefficient, thus maximizing the extracted information while minimizing the average trace distance of the photons state to their evolution only subject to noise. Indeed, this angle achieves an accuracy of 88.265% with a trace distance $\overline{D}_{nsn}$ of 0.226. In the non-idealized measurement scheme (going from 10 to 50 time steps), this accuracy is 85.41%. A question the spy could ask is "Are there ways to decrease the impact on the photons or to increase the accuracy ?". In the next Chapter we investigate the use of quantum feedback to decrease the perturbations caused by the measurement.

Chapter 6

Quantum feedback

We determined in the previous Chapter that a spy could extract a relatively high amount of information from the qubits using homodyne measurements along with a recurrent neural network. However, its impact on the photons could still make the spy detectable by Alice and Bob. Consequently, we investigate in this Chapter the use of quantum feedback to decrease the perturbations caused on the system, thus covering the tracks of the spy.

The equation describing the deterministic, and thus averaged, evolution of an open quantum system subject to dissipation, homodyne detection and unconditional feedback is

$$\dot{\hat{\rho}} = -\mathrm{i}\left[\hat{H} + \frac{\gamma_E}{2}\left(\hat{c}^\dagger\hat{F} + \hat{F}\hat{c}\right), \hat{\rho}\right] + \gamma_E\mathcal{D}[\hat{c} - \mathrm{i}\hat{F}]\hat{\rho} + \frac{1-\eta}{\eta}\gamma_E\mathcal{D}[\hat{F}]\hat{\rho} + \gamma_D\mathcal{D}[\hat{v}]\hat{\rho}_J(t), \tag{6.1}$$

as explained in Section 1.6. However, to describe the stochastic dynamics of a single realisation of the system (e.g., through numerical simulations) one would need Eq. (1.62) which we recall to be

$$\begin{aligned}\mathrm{d}\hat{\rho}_J(t) =& \mathrm{d}t\left\{-\mathrm{i}\left[\hat{H}, \hat{\rho}_J(t)\right] + \gamma_E\mathcal{D}[\hat{c}]\hat{\rho}_J(t) - \mathrm{i}\gamma_E\left[\hat{F}, \hat{c}\hat{\rho}_J(t) + \hat{\rho}_J(t)\hat{c}^\dagger\right]\right\} \\ &+ \gamma_E\mathcal{D}[\hat{F}]\hat{\rho}_J(t)\mathrm{d}t/\eta + \gamma_E\mathrm{d}W(t)\mathcal{H}[\sqrt{\eta}\hat{c} - \mathrm{i}\hat{F}/\sqrt{\eta}]\hat{\rho}_J(t).\end{aligned} \tag{6.2}$$

We define the feedback operator $\hat{F}$ as depending on an angle ϕ:

$$\hat{F} = \sin(\phi)\hat{\sigma}_z + \cos(\phi)\hat{\sigma}_x. \tag{6.3}$$

We denote a state subject to feedback, measurement and dissipation by $\hat{\rho}_{fb}(t)$. We can then define similar trace distances as in the previous sections :

$$D\left(|0\rangle\langle 0|_{fb}(t), |0\rangle\langle 0|\right) \equiv D_{fb}(|0\rangle\langle 0|)(t), \tag{6.4}$$

$$D\left(|0\rangle\langle 0|_{fb}(t), |0\rangle\langle 0|_n(t)\right) \equiv D_{fbn}(|0\rangle\langle 0|)(t). \tag{6.5}$$

Since the measurement operator depends on θ and the feedback operator depends on ϕ, we set the total time to a fixed value of $\gamma_D t = 2$ so that we can represent the trace distances as functions of the two angles only, as is done in Figure 6.1. These two trace distances have local minimums for some feedback and measurement operators, which means that the feedback can actually reduce the impact of the measurement on the system. In addition, the trace distance $\overline{D}_{fbn}(\theta, \phi)$ reaches 0 at some θ and ϕ values. This is a very important result: applying feedback could completely counter the effects of the homodyne measurement, thus allowing a spy to completely cover its

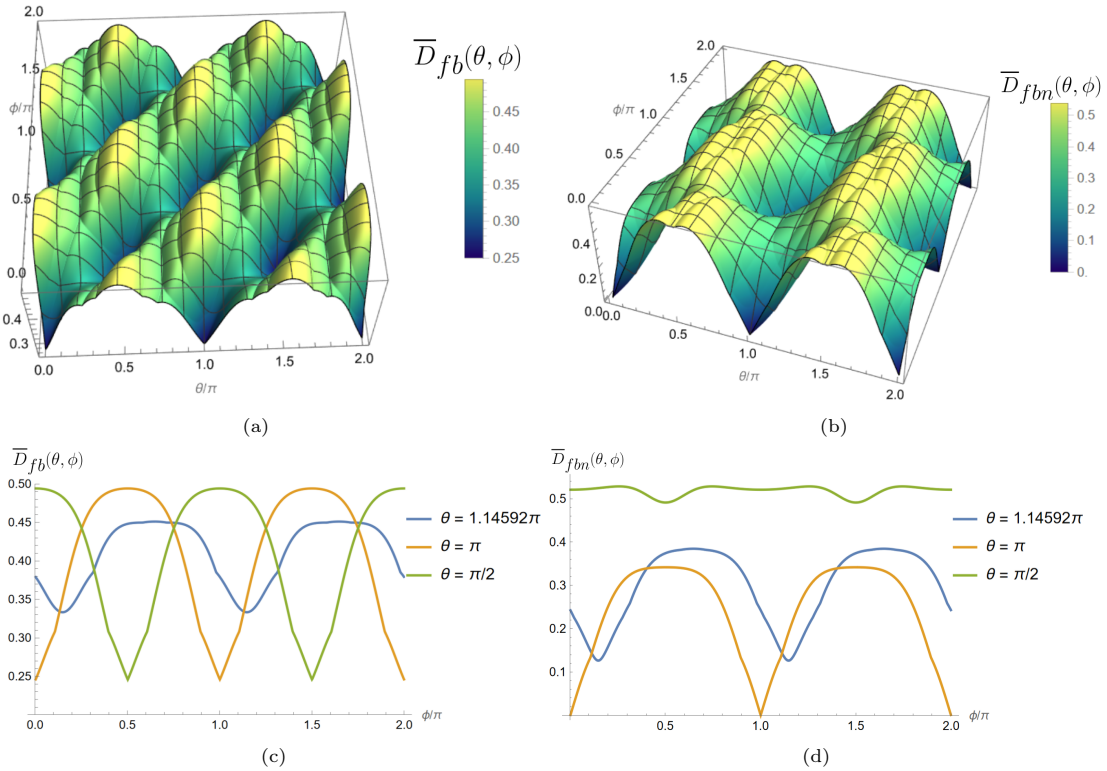


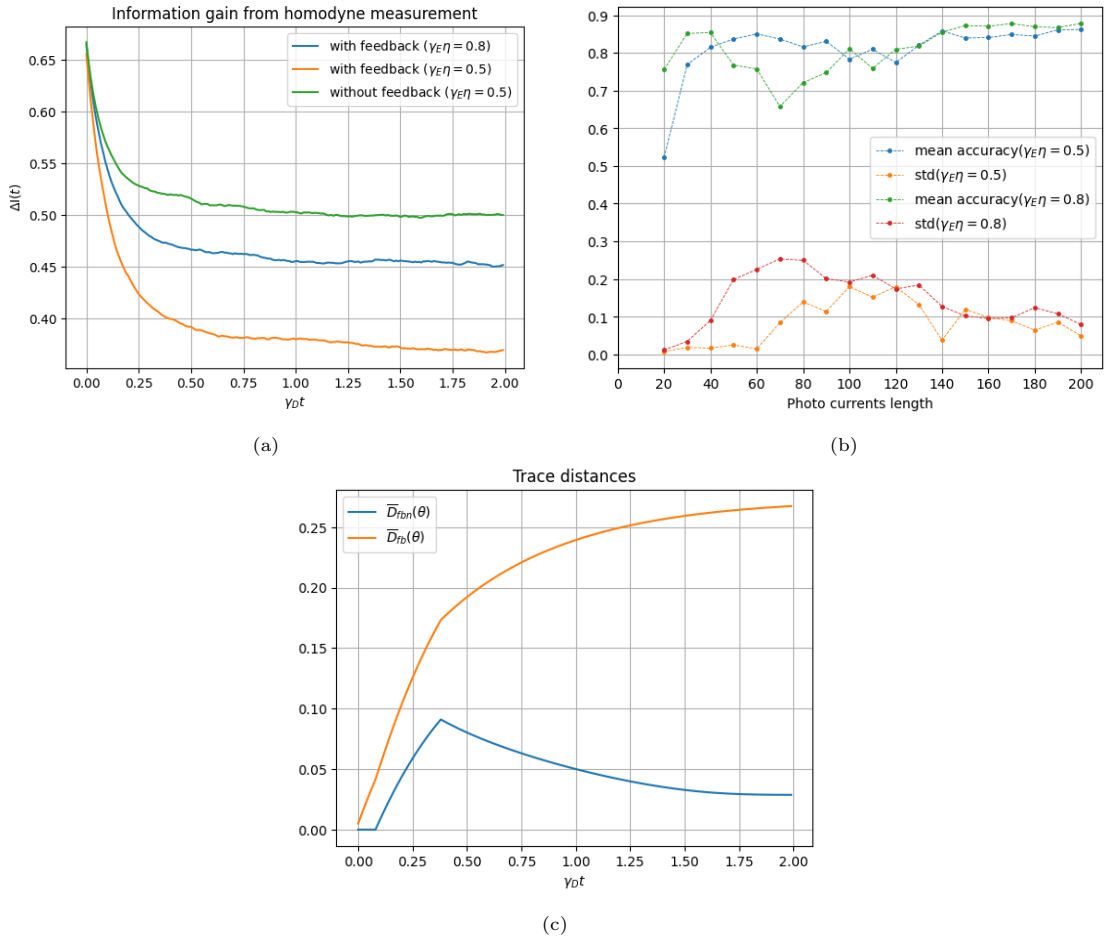
Figure 6.1: (a) and (b) Trace distances $\overline{D}_{fb}(\theta,\phi)$ and $\overline{D}_{fbn}(\theta,\phi)$ respectively, as functions of the measurement angle θ and the feedback angle ϕ. (c) and (d) are the same trace distances for different values of the measurement angle θ, as functions of ϕ. The blue curve is for the optimal angle measurement $\theta = 1,14592\pi$. Other parameters are set to $\omega = \gamma_E = \gamma_D = 1$

tracks. If the measurement is done using the optimal angle $\theta = 1,14592\pi$, the total trace distance can be reduced from 0.45 to 0.35 (Figure 6.1c) and the trace distance $\overline{D}_{fbn}(\theta,\phi)$ from 0.39 to 0.126, if the feedback is correctly engineered. The feedback angle values minimizing the trace distances for the optimal measurement are $\phi = 0,14642\pi$ and $1,14592\pi$. Therefore, using the optimal measurement angle for the feedback basis maximizes its effect.

For some angles, the maximums of the traces distances are due to the third term in Eq. (6.1), which represents the noise introduced in the system by the feedback. Indeed, if the feedback is not correctly engineered it causes more perturbations to the photons than it prevents, thus increasing the trace distance.

Let us look at the information gained from the measurement while applying feedback. Indeed, even if the homodyne measurement is the same as before, the introduction of feedback could introduce some unwanted noise and change the photon state, therefore modifying the extracted current. The information gains ΔI with and without feedback, for the measurement and feedback angle $\theta = \phi = 1.14592\pi$, are displayed in Figure 6.2a. The information gain with feedback is below the one without it, even though the maximum difference is 0.18, which means that the introduction of feedback does in fact have an impact on the extracted information. This result is logical since, as shown in Figure 6.1b, the feedback cannot completely erase the impact of measurement on the state of the system even when perfectly engineered for $\theta = 1.14592\pi$, and

thus perturbs the states.



(a) (b)

(c)

Figure 6.2: (a) Information gains ΔI without feedback (green curve) and with feedback, for $\gamma_E\eta = 0.5$ (orange curve) and 0.8 (blue curve). The measurement and feedback angle are set to $\theta = \phi = 1.14592\pi$. (b) Mean accuracy (blue) and standard deviation (orange) of the neural network as a function of the photo current lengths (in time steps) used. The starting time step of the measurement is 10. (c) Trace distances for a measurement interval going from 10 to 50 time steps, $\gamma_E\eta = 0.5$ and $\theta = \phi = 1.14592\pi$. Other parameters are set to $\omega = \gamma_E = \gamma_D = 1$.

We now try to optimize the measurement time similarly to what we did in Section 5.3 for the homodyne measurement without feedback. The results are displayed in Figure 6.2b. The network achieves a mean accuracy 83.66% with a standard deviation of 0.023 for a measurement length of 40 time steps, still considering that the measurement starts on the tenth time step.

So far in this work, we set the parameter $\gamma_E\eta$, efficiency of the homodyne measurement, to the value of 0.5. However, a greater efficiency, for example 0.8, would be a realistic consideration, and could increase the extracted information and the tomography accuracy, while increasing the impact of measurement. We computed the extracted information with feedback and $\gamma_E\eta = 0.8$ (blue curve) in Figure 6.2a, and the mean accuracy of the network for this same efficiency value in Figure 6.2b. The information gain is increased by almost 0.15 and is much closer the gain without feedback (for $\gamma_E\eta = 0.5$). However, the accuracy is increased only for measurement lengths of $10, 20$ and 30 time steps, but with a much larger standard deviation. We can thus

keep the efficiency to 0.5 and set the measurement to start at 10 time steps and end at 50. The final metric to evaluate is the impact of this optimized measurement and feedback scheme. We do so using the two trace distances $\overline{D}_{fbn}(t)$ and $\overline{D}_{fb}(t)$, which are displayed in Figure 6.2c. We see that the total trace distance goes up to 0.25, but the trace distance with respect to the state evolution under noise only is small (0.02875). In addition, we see that after the measurement this trace distance decreases. Such a low trace distance means that Alice and Bob have a very small probability of detecting Eve, and thus that she could potentially break the BB84 protocol security.

6.1 Summary

In summary, we implemented an eavesdropping scheme composed of homodyne measurement, quantum unconditional feedback, and recurrent neural network tomography, on the BB84 protocol. The results we obtained after optimizing the measurement operator and duration, the feedback angle, and the efficiency, are displayed in Table 6.1.

This last result shows the theoretical possibility to break the BB84 protocol by using the measurement and feedback scheme imagined in this master thesis.

θ	ϕ	Measurement interval (time steps)	$\overline{D}_{fbn}$ ($\gamma_D t = 2$)	Mean accuracy	$\lambda(\theta)$
1.14592π	1.14592π	10 - 50	0.029	83.659%	1.2299

Table 6.1: Mean accuracy of the neural network and corresponding standard deviations for the values of theta maximizing the test accuracy.

Conclusion

The aim of this master thesis was to study the security of the BB84 quantum key distribution protocol subjected to dissipation and eavesdropping. More specifically, the main objective was to assess the theoretical feasibility of recovering photon states (i.e., recovering the private key) for an eavesdropper using weak measurements, and using feedback to minimize the impact of such a measurement scheme.

In Chapter 1 of this work, we introduced the density operator formalism to describe open quantum systems and recalled the quantum mechanics postulates using it. We then detailed the GKSL master equation to describe a photon traveling in an optical fiber either subject to dissipation, eavesdropping or both. Finally we derived the evolution equations describing a system monitored via homodyne measurement with and without feedback.

In Chapter 2 we introduced the definition of qubits from linearly polarized photons used in the BB84 protocol. We have explained the latter step by step in an idealized case first, without dissipation nor eavesdropping, then in a realistic case where privacy amplification and information reconciliation can be implemented.

In Chapter 3, we first analyzed the impact on the protocol security of noise and eavesdropping, separately and then jointly. We obtained that the probabilities of Bob measuring the photon state Alice initially sent were $\frac{3}{4}$ with Eve performing projective measurements at the middle of the channel, $\frac{e^{-2\gamma t}}{4} + \frac{3}{4}$ with a $\hat{\sigma}_x$-modeled noise only, and $\frac{e^{-2\gamma t}}{8} + \frac{5}{8}$ with both noise and eavesdropping. Still in Chapter 3, we discussed the possible implementation of a homodyne measurement scheme by the spy, the output of such measurement (namely the homodyne photo current) and its characteristics (i.e., mean, variance, auto-correlation). The last Section consisted in a analysis of such measurement scheme impact on the photons using different trace distances as metrics. The results were that the angles minimizing the impact on the photons of homodyne measurements, when the corresponding operator was defined as $\hat{c} = \sin\theta\hat{\sigma}_z + \cos\theta\hat{\sigma}_x$, were $k\frac{\pi}{2}$ for the initial states $\{|0\rangle, |1\rangle\}$, and $0 + k\pi$ for $\{|+\rangle, |-\rangle\}$.

In Chapter 4, we detailed the impossibility to retrieve the initial state, sent by Alice, using standard tomography on a single realization of the photo current. This result motivated the use of machine learning and, in particular recurrent neural networks, the theoretical aspects of which we described in the same Chapter. We then implemented an LSTM neural network to perform initial state tomography, achieving an accuracy of 88.265% for some specific measurement angle $\theta = 1.14592\pi$.

In Chapter 5, we introduced the concept of expected information gain which formally justified the higher accuracy yielded by this measurement angle. Then, using the average trace distance on the four initial states, we computed the λ coefficient to optimize the trade-off between information extraction and state perturbation. This new quantity revealed $\theta = 1.14592\pi$ as one of its minimizers, angle which was used as the optimal measurement angle in the remaining sections. This result is the central point of this work since it constitutes a way for Eve to potentially

retrieve enough information to compute the private key of Alice and Bob. The measurement scheme was also corrected to be more realistic by setting its starting point to 10 time steps, and optimized by decreasing its length to 40 time steps.

Finally, in the last Chapter we considered the introduction of unconditional feedback by the spy to minimize its impact. We discovered that for same measurement and feedback angles $\theta = \phi = 1.14592\pi$ and a measurement and feedback duration of 40 time steps (10 to 50), the spy could decrease the impact of its measurement from 0.22645 to 0.02875, while achieving an accuracy of 83.659%.

As we just detailed, this Master's thesis contains several results on the impact of attacks in quantum cryptography that have not been reported in the literature, such as the QST accuracy achieved using recurrent neural networks and homodyne measurements, or the reduction of measurement impact on the photons, induced by the introduction of unconditional quantum feedback. A manuscript gathering those results is under preparation.

A question that arises directly from this work is "Is a measurement scheme such as homodyne detection and feedback possible to implement in practice on the BB84 protocol?". This is of course something important to be investigated, and should constitute the direct continuation of this Master's thesis. Even though very few researches are present in the literature, one could adapt this work for QKD implementation in space, between satellites [47, 48]. It's important to note that the practical implementation of QKD technologies has been a major topic of interest in both academic and private sectors for the past few years and continues to be.

To go further we could look at conditional feedback, where the measurement result is taken into account before applying the feedback on the system. However, This would constitute too great an obstacle within the scope of this work, as conditional feedback involves non-Markovianity. Indeed, since the measurement result (i.e., the photo current in the case of homodyne detection) must be processed before being fed back into the system, and since this processing cannot be done instantaneously, this feedback must include non-Markovianity in the sense that the current would be injected back into the system after some finite time.
Furthermore, during all the calculations and analysis we have done (e.g., to derive the GKSL master equation), one of the assumptions was that the dynamics of the system is Markovian, which means there is no memory effect involved, either within the system itself or in its interaction with the environment. Consequently, exploring the non-Markovian field of feedback would need to detail and derive too many mathematical and physical tools.

One could also consider the case of the photon having a non-nul probability to be absorbed in the optical fiber, as in [25]. This consideration amounts to spanning the Hilbert space of the system on three basis vectors instead of two (i.e. adding a vacuum base vector $|0\rangle$), which would change the mathematical description of the system, along with its physical interpretation.

Bibliography

[1] P. Botsinis, D. Alanis, Z. Babar, H. V. Nguyen, D. Chandra, S. X. Ng, and L. Hanzo, IEEE Communications Surveys I& Tutorials **21**, 1209 (2019).

[2] O. D. Okey, S. S. Maidin, R. Lopes Rosa, W. T. Toor, D. Carrillo Melgarejo, L. Wuttisittikulkij, M. Saadi, and D. Zegarra Rodríguez, Sustainability **14** (2022), 10.3390/su142315901.

[3] R. L. Rivest, A. Shamir, and L. Adleman, Commun. ACM **21**, 120–126 (1978).

[4] P. W. Shor, SIAM Journal on Computing **26**, 1484–1509 (1997).

[5] R. Renner, *Security of Quantum Key Distribution*, Ph.D. thesis, ETH Zurich (2006).

[6] C. H. Bennett and G. Brassard, Theoretical Computer Science **560**, 7 (2014), theoretical Aspects of Quantum Cryptography – celebrating 30 years of BB84.

[7] T. Metger and R. Renner, Nature Communications **14** (2023), 10.1038/s41467-023-40920-8.

[8] R. Kumar, F. Mazzoncini, H. Qin, and R. Alléaume, Scientific Reports **11** (2021), 10.1038/s41598-021-87574-4.

[9] S. R. M and C. M. B, "Comprehensive analysis of bb84, a quantum key distribution protocol," (2023), arXiv:2312.05609 [quant-ph] .

[10] A. Adu-Kyere, E. Nigussie, and J. Isoaho, Sensors **22** (2022), 10.3390/s22166284.

[11] M. L. Bellac, *Physique quantique, 2e édition* (CNRS édition, 2007).

[12] J. Martin, T. Bastin, and P. Schlagheck, "Mécanique quantique avancée, phys3021-1," (2022).

[13] G. Lindblad, Communications in Mathematical Physics **48**, 119 (1976).

[14] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Journal of Mathematical Physics **17**, 821 (1976).

[15] D. Manzano, AIP Advances **10**, 025106 (2020).

[16] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, 2009).

[17] Q. Xu, *Optical Homodyne Detection and Applications in Quantum Cryptography*, Ph.D. thesis, Institut Polytechnique de Paris (2009).

[18] H. Carmichael, *An Open Systems Approach to Quantum Optics: Lectures Presented at the Université Libre de Bruxelles, October 28 to November 4, 1991*, vol. 18 (Springer Berlin Heidelberg, 1993).

[19] H. M. Wiseman and G. J. Milburn, Phys. Rev. A **49**, 1350 (1994).

[20] H. M. Wiseman, Phys. Rev. A **49**, 2133 (1994).

[21] N. D. Mermin, *Quantum Computer Science: An Introduction* (Cambridge University Press, 2007).

[22] N. Datta, "Lecture notes in quantum information and computation," (2019-2020).

[23] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[24] A. Chiuri, V. Rosati, G. Vallone, S. Pádua, H. Imai, S. Giacomini, C. Macchiavello, and P. Mataloni, Phys. Rev. Lett. **107**, 253602 (2011).

[25] A. Kozubov, A. Gaidash, and G. Miroshnichenko, Phys. Rev. A **99**, 053842 (2019).

[26] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Physical Review A **71** (2005), 10.1103/physreva.71.062310.

[27] G. M. D'Ariano and H. P. Yuen, Phys. Rev. Lett. **76**, 2832 (1996).

[28] O. Alter and Y. Yamamoto, Phys. Rev. Lett. **74**, 4106 (1995).

[29] Y. Aharonov, J. Anandan, and L. Vaidman, Phys. Rev. A **47**, 4616 (1993).

[30] M. Ueda and M. Kitagawa, Phys. Rev. Lett. **68**, 3424 (1992).

[31] A. Imamoglu, Phys. Rev. A **47**, R4577 (1993).

[32] A. Royer, Phys. Rev. Lett. **73**, 913 (1994).

[33] N. Cerf, G. Leuchs, and E. Polzik, *Quantum Information With Continuous Variables of Atoms and Light* (2007).

[34] R. Schmied, Journal of Modern Optics **63**, 1744–1758 (2016).

[35] J. Altepeter, E. Jeffrey, and P. Kwiat (Academic Press, 2005) pp. 105–159.

[36] J. C. Chapman, J. M. Lukens, B. Qi, R. C. Pooser, and N. A. Peters, Optics Express **30**, 15184 (2022).

[37] N. Mosco and L. Maccone, Physics Letters A **449**, 128339 (2022).

[38] W. S. McCulloch and W. Pitts, The bulletin of mathematical biophysics **5**, 115 (1943).

[39] F. Rosenblatt, Psychological Review **65**, 386 (1958).

[40] G. Louppe, "Deep learning, info8010-1," (2024).

[41] L. Geurts, Pierre. Wehenkel, "Introduction to machine learning, elen0062-1," (2023).

[42] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).

[43] S. Hochreiter, *Untersuchungen zu dynamischen neuronalen Netzen*, Master's thesis, Institut für InformatikTechnische Universität München (1991).

[44] S. Hochreiter and J. Schmidhuber, Neural Computation **9**, 1735 (1997).

[45] K. Jacobs, *Quantum Measurement Theory and its Applications* (Cambridge University Press, 2014).

[46] H. J. Groenewold, International Journal of Theoretical Physics **4**, 327 (1971).

[47] S.-K. Liao, W.-Q. Cai, W.-Y. Liu, L. Zhang, Y. Li, J.-G. Ren, J. Yin, Q. Shen, Y. Cao, Z.-P. Li, F.-Z. Li, X.-W. Chen, L.-H. Sun, J.-J. Jia, J.-C. Wu, X.-J. Jiang, J.-F. Wang, Y.-M. Huang, Q. Wang, Y.-L. Zhou, L. Deng, T. Xi, L. Ma, T. Hu, Q. Zhang, Y.-A. Chen, N.-L. Liu, X.-B. Wang, Z.-C. Zhu, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, Nature **549**, 43–47 (2017).

[48] M. Avesani, L. Calderaro, M. Schiavon, A. Stanco, C. Agnesi, A. Santamato, M. Zahidy, A. Scriminich, G. Foletto, G. Contestabile, M. Chiesa, D. Rotta, M. Artiglia, A. Montanaro, M. Romagnoli, V. Sorianello, F. Vedovato, G. Vallone, and P. Villoresi, npj Quantum Information **7** (2021), 10.1038/s41534-021-00421-2.