Efficient Rational Agents for Quantum Tomographies and Inferences

Thiago de Oliveira Maciel

January 2016

Efficient Rational Agents for Quantum Tomographies and Inferences

Thiago de Oliveira Maciel

Orientador:

Prof. Dr. Reinaldo Oliveira Vianna

Orientador (doutorado sanduíche):

Prof. Dr. David Gross

Tese apresentada à UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, como requisito parcial para a obtenção do grau de DOUTOR EM FÍSICA.

Belo Horizonte
Brasil
Janeiro de 2016

Dedication

Aos meus pais: Ademar e Nara.

Abstract

We will elaborated on correct inferences, or sometimes plausible ones, when inferring quantum states and processes. Making correct inferences can be considered part of being a rational agent, because one way to act rationally is to reason logically to the conclusion that a given action will achieve our objective and then to act on that conclusion. Yet, correct inference is not all of rationality; in some situations, there is no provably correct thing to do, especially in possession of incomplete information and noise data, but something must still be done. The aspects of quantum mechanics are maturing in both theory and experiment and we still need higher-dimensional systems to test them. Here, we will explore procedures to infer unknown mixed states and unknown quantum dynamics, which can be experimentally viable and efficiently processed. The heart of our methods are hierarchies of semi-definite programs and its flexible formulations. We presented numerous schemes to achieve the optimal results in the paradigm of complete and incomplete information. We tested our methods in different experiments, such as quantum optics and nuclear magnetic resonance (NMR) systems and presented the results here.

Keywords

quantum state tomography, quantum process tomography, quantum information, quantum maps, positive maps, completely positive maps, positive operators, rational agents, semi-definite programs, SDP, Kraus operators, Choi operators, Choi map, detecting entanglement, convex sets, density matrices, SIC-POVM.

Resumo

Iremos desenvolver sobre inferências corretas e, às vezes, plausíveis ao inferir estados e processos quânticos. Fazer uma inferência correta pode ser considerado parte de ser um agente racional, porque uma forma de agir racionalmente é pensar logicamente até chegar a conclusão que uma determinada ação irá atingir nossos objetivos e, então, agir de acordo com essa conclusão. No entanto, para uma inferência correta não basta somente racionalidade; em alguns casos, não há nenhuma evidência da melhor coisa a se fazer, especialmente quando temos informações incompletas e dados ruidosos, mas alguma coisa ainda deve ser feita. Os aspectos da mecânica quântica estão amadurecendo tanto na teoria quanto nos experimentos e ainda precisamos de sistemas de maiores dimensões para testá-los. Aqui, desenvolvemos vários métodos a fim de inferir estados quânticos desconhecidos e processos quânticos também desconhecidos com pouca, ou toda informação, que podem ser experimentalmente viáveis e eficientemente processados. O coração dos nossos métodos são hierarquias de programas semi-definidos e suas formulações flexíveis. Apresentaremos uma enormidade de receitas para atingir nossos no paradigma de informação completa e incompleta. Testamos nossos métodos em diferentes experimentos, como sistemas em óptica quântica e ressonância magnética nuclear (RMN) e apresentamos os resultados aqui.

Palavras-chave

tomografia de estado quântico, tomografia de processo quântico, informação quântica, operações quânticas, mapas positivos, mapas completamente positivos, operadores positivos, agentes racionais, programação semi-definida, SDP, operadores de Kraus, operadores de Choi, mapa de Choi, matriz densidade, SIC-POVM.

Agradecimentos

Não há como não começar a agradecer, primeiramente, à minha família: a meu pai Ademar e à minha mãe Nara — a quem dedico esta tese —; às minhas irmãs Velise e Taísa; à minha familia que produziu todo o meu contexto familiar (tios, tias, primos, primas e agregados, sejam presentes ou ausentes) e que me ajudam ser quem sou e onde as bases para o que ainda posso ser foram construídas.

Em especial, agradeço ao pessoal do Infoquant[1] (Reinaldo, Tanus, Debarba, Iemini, e Diego) e todos aqueles que um dia foram dignos de um convite para entrar no cafofo e ainda estão vivos[2], sintam-se merecedores deste agradecimento[3].

Agradeço, também, ao meu professor e amigo Mário Mazzoni: grande parte do que aprendi sobre física fora ensinado por ele. Também, ao Sebastião de Pádua e ao C. H. Monken, que — devido às grandes discussões e ensinamentos — entendo, hoje, que o que acontece no laboratório é muito diferente de calcular traços de produtos de operadores hermiteanos.

E, claro, agradeço ao meu orientador, professor e amigo Reinaldo Oliveira Vianna: sem a oportunidade de trabalhar com ele, nenhuma palavra estaria escrita nesta tese. Obrigado por me dar a oportunidade de enxergar a mecânica quântica dessa belíssima forma e que, hoje, gera este texto.

For the ones in Freiburg/Köln, especially David Gross, Rafael Chaves, Richard Küng. Alexandre Lopes and Johan Aberg, my sincere acknowledgements! Thank you!

Esta tese teve o apoio financeiro direto e indireto da **CAPES**, **FAPEMIG** e **CNPq**.

[1] Vulgo, *cafofo*: um lugar inóspito, frio, barulhento, fedido e cheio de agressividade.

[2] Ou não, *e.g.*, a ilustríssima professora M.C. Nemes†.

[3] Enumerar pessoas é sempre uma tarefa inglória, perdoem-me por não citar nominalmente.

Contents

List of Figures

List of Tables

List of Publications

Thesis' publications:

- Variational quantum tomography with incomplete information by means of semidefinite programs [45];
- Optimal estimation of quantum processes using incomplete information: variational quantum process tomography [46];
- Experimental implementation of a NMR entanglement witness [19];
- Quantum state tomography with incomplete data: maximum entropy and variational quantum tomography [22];
- Minimum tomography of two entangled qutrits using local measurements of one-qutrit symmetric informationally complete positive operator-valued measure [57];
- Quantum process tomography with informational incomplete data of two J-coupled heterogeneous spins relaxation in a time window much greater than T_1 [48].

Publications as co-author

- Witnessed entanglement and the geometric measure of quantum discord [17];
- Quantifying quantum correlations in fermionic systems using witness operators [30];
- Inferring latent structures via information inequalities [10];
- Entanglement of indistinguishable particles as a probe for quantum phase transitions in the extended Hubbard model [29].

CHAPTER 1

Introduction

Working at a problem, we would like to guess the whole solution, but, if we fail, sometimes we satisfy ourselves guessing some feature of the solution. Wrangling over data, testing, and restating problems are the everyday basis of an inferrer.

As Pólya said in [58], to be a good inferrer you must be a good guesser; to be a good guesser you should be able to examine your guesses, compare them with the facts, modify them if necessary and, so, acquire an extensive experience with guesses that failed and guesses that came true. With such an experience, you may be able to rationalize more competently about possible solutions.

For example, we usually attain a discovery in physics in two steps: first, we note a certain regularity in the observed data; then, we explain this regularity as a consequence of some general law. Different scientists might take these two steps and, possibly, with a long interval of time between them, *e.g.*, the observation of photoelectric effect by Hertz in his experiments in 1887 [26] and the explanation of the phenomenon by Einstein in 1905 [18].

The first mental model of a scientist is usually the *"Simplex sigillum veri"*, or "Simplicity is the seal of truth", but here we are confining ourselves to *"Try the simplest thing first!"*.

Trying the simplest thing first is part of an attitude which is reasonable and advantageous in face of little or great problems. This first examination may serve useful when applying the *Ockham's razor*, if we need to return and give a closer look in more complex possibilities.

Physicists are usually inclined to imagine that Nature obtains the greatest effect with the least effort. We succeed in forming lots of principles of this sort in all areas of physics. In this thesis, we will find a great number of methods written as maxima or minima problems. Especially, we will attack the problem of inferring quantum states and maps in cases where little or no information is available.

The problem of inferring probability distributions is old as the theory of probability itself. Laplace gave us the first well-known principle to supply a criterion of choice, which reads,

Principle 1 (Laplace Principle of Insufficient Reason)**.** *One should assign equal probabilities when no prior information is available about the events and if there is no reason to think otherwise.*

However, as Jaynes said in [32], except in cases where there is an evident element of symmetry that clearly renders the events *equally possible*, this assumption may appear just as arbitrary as any other. Due to the lack of any constructive principle which would give us a reason for preferring one probability distribution over another (when both agree equally well with the available information), the scientific community practically abandoned this way of formulating problems from Laplace up to Jaynes in 1957 [32, 33].

Then, Shannon [66], with the birth of classical information theory, gave the missing ingredient, allowing Jaynes to formulate his principle [32], namely,

Principle 2 (Jaynes' Principle of Maximum Entropy)**.** *In making inferences based on partial information, we must use the probability distribution which has maximum entropy subject to any prior knowledge.*

Now, we can extend the principle of insufficient reason with Jaynes' principle of maximum entropy (MaxEnt) with the following difference: we assert for the positive reason that it is uniquely determined as the one which is maximally noncommittal regarding missing information using the maximum entropy distribution, instead of the negative one that there is no reason to think otherwise. We use this principle often in statistical mechanics, as the maximum entropy gives the estimate of thermodynamical properties.

In quantum scenarios, the subject of estimating density matrices based on measurement results, or simply Quantum State Tomography (QST), is valuable on its own. It is essential in Quantum Computation; Quantum Information Processing [54]; Quantum Process Tomography (QPTs) [46]; and in validation of quantum dynamics. However, the number of required measurements for a complete tomography grows exponentially in the dimension of the system. In this case, an informationally complete context of measurements becomes impractical in real experiments, requiring proper inference schemes.

If we try to apply Principle 2 to infer quantum properties, *e.g.*, entanglement, Horodecki *et al.* [27] showed that Jaynes principle fails, suggesting it is not a proper inference scheme in quantum information processing when entanglement plays the lead role.

The lack of scalable methods to deal with measurements errors and reliable quantum inference schemes justifies this doctoral work. We are in pursuit of a principle which relies on semi-definite programming (SDPs), instead of nonlinear optimization problems like MaxEnt and other formulations. We chose SDPs knowing that modern interior-point algorithms can solve them with polynomial computational complexity [53].

Inferences of this kind are highly interdisciplinary. Dialoging with concepts like Rational Agents from the field of Artificial Intelligence and Decision Theory is one of the many milestones of this thesis.

We will enter in realm of the quantum inferences using this interdisciplinary language, starting with Rational Agents in Chapter 2; Quantum inference on informationally complete contexts in Chapter 3; Estimation on informationally incomplete contexts in Chapter 4; Estimation of quantum processes in informationally incomplete contexts in Chapter 5; One application of our methods in Quantum process tomography with informational incomplete data of two J-coupled heterogeneous spins relaxation in a time window much greater than T_1 in Chapter 6; and concluding our journey in Chapter 7.

CHAPTER 2

Rational Agents

The idea of this thesis is offering mechanisms to construct a *Rational Agent* capable of dealing with our experimental data and produce *proper inferences* about our physical systems. In this chapter, we concentrate on general principles of rational agents and on components for constructing them, following Russell *et al.*'s work [63].

In [63], Russell defines an *agent* as something that acts: it operates autonomously, perceives their environment, adapts to change, persists over a prolonged time, creates, and pursues goals. A *rational agent* is one that acts to achieve the best outcome or the best *expected* outcome considering limitations, *e.g.*, when there is uncertainty involved, when there is not enough time, or memory to do all the computations. We will pursue an *efficient rational agent*, where we limit ourselves in using polynomial algorithms[1] in our inferences.

2.1 Agents and environments

To be more precise, we can say:

Definition 2.1 (Russell *et al.* [63])**.** *An* agent *is anything that can perceive its environment through* sensors *and can act upon that* environment *through* actuators*.* We illustrated this definition in Fig. 2.1.

For example, our agent could be an inferrer that perceives what happens in an optical table via some data generated by the photocounting devices (sensors) and could act sending signals to the table's controller board (actuator) to realign some components.

Russell used the term *percept* to refer to the agent's perceptual inputs at any given instant. He also defined the agent's *percept sequence* as the complete history of everything the agent has ever perceived. In general, an agent's choice of action at any given instant can depend on the entire percept sequence

[1] *I.e.*, we can solve them by a deterministic Turing machine using a polynomial number of operations, or with polynomial complexity.

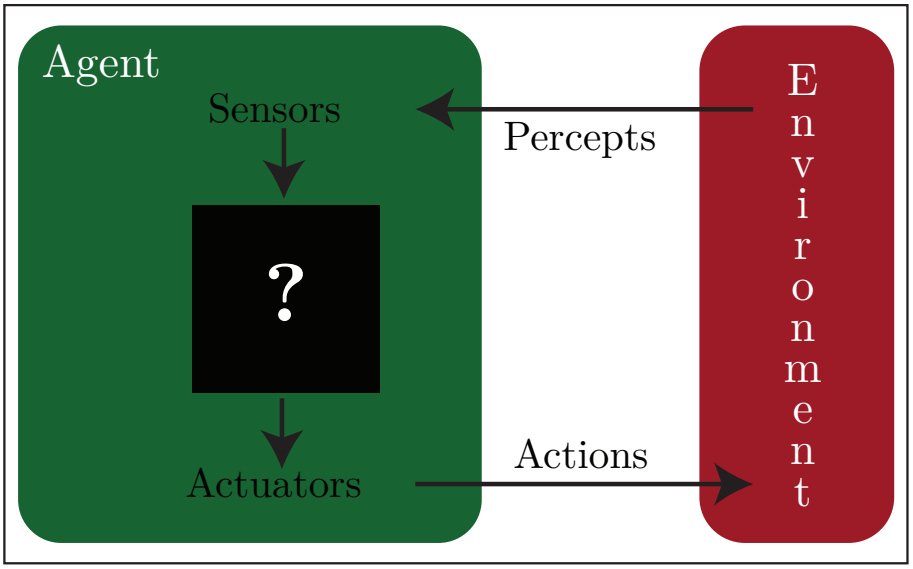

Figure 2.1: Agents interact with environments through sensors and actuators.

observed, *but not on anything it has not perceived*. In other words, we will infer *only* with given information.

With that, we are now in place to choose how the agent will behave, *i.e.*:

Definition 2.2 (Russell *et al.* [63])**.** *An agent function that maps any given percept sequence to an action defines the agent's behavior.*

How can we choose between different behaviors is a natural question after this definition. We need our agent to think rationally.

2.2 Being rational

Definition 2.3 (Russell *et al.* [63])**.** *When the agent* always *do the* right thing*, he achieves perfect rationality.*

This definition seams simple at first glance, but faces two major challenges while implementing the agent's rationality: first, defining the right thing to do, and, second, always doing this right thing. Let us start with the first, what does it mean to do the right thing?

This is a decision problem, where *Decision Theory* is a field of its own. A complete description of the theory lies outside of the scope of this thesis, but we will elucidate our arguments following the definitions in Young *et al.*'s work in [74], keeping in mind that we want to create an inferrer agent.

Central to decision theory, is the notion of a set of *decision rules* for an inference problem. We can compare different decision rules examinating the *risk functions* (or *performance measures*). The risk function describes the *expected loss* in use of the rule, under hypothetical repetition of the sampling experiment giving rise to some data x, as function of the *parameter* of interest, say θ.

We can describe formally a statistical decision problem using the following elements (*cf.* Young *et al.* [74]):

(I) : A parameter space Θ, which will usually be a subset of $\mathbb{R}^d$ for some $d \geq 1$ so that we have a vector of d unknown parameters. This represents the set of possible unknown states of nature. The unknown parameter value $\theta \in \Theta$ is the quantity we wish to make an inference about;

(II) : A *sample space* $\mathcal{X}$, the space in which the observed data (percept sequences) x lie. Typically, we have n preceived observations, so the data are of the form $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$;

(III) : An *action space* $\mathcal{A}$. This represents the set of all actions, or decisions, available to the inferrer, *e.g.*, in a hypothesis testing problem, where it is necessary to decide between two hypotheses H_0 and H_1, there are two possible actions corresponding to "accept H_0" and "accept H_1". So here $\mathcal{A} = \{a_0, a_1\}$, where a_0 represents accepting H_0 and a_1 represents accepting H_1;

(IV) : A *loss function* $L : \Theta \times \mathcal{A} \to \mathbb{R}$. This function links the action to the unknown parameter. If we take action $a \in \mathcal{A}$ when the true state of nature is $\theta \in \Theta$, then we incur a loss $L(\theta, a)$;

(V) : A set $\mathcal{D}$ of *decision rules*. An element $d : \mathcal{X} \to \mathcal{A}$ of $\mathcal{D}$ is such that we associate a specific action $d(x) \in \mathcal{A}$ with each point x in $\mathcal{X}$, *e.g.*, with hypothesis testing, we can adopt the rule: "Accept H_0 if $\bar{x} \leq \pi$, otherwise accept H_1", where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. This corresponds to a decision rule

$$d(x) = \begin{cases} a_0, & \text{if } \bar{x} \leq \pi, \\ a_1, & \text{if } \bar{x} > \pi. \end{cases}$$

(VI) : The *risk* associated with a decision rule d based on random data X. It is the expected loss with respect to the distribution of X for the particular parameter value θ.

$$\begin{aligned} R(\theta, d) &= \mathbb{E}_\theta \, L(\theta, d(X)) \\ &= \sum_{x \in \mathcal{X}} L(\theta, d(X)) f(x; \theta), \quad \text{for discrete } X; \end{aligned} \tag{2.1}$$

These six elements define what we call a complete decision problem.

Obviously, there is not one fixed loss function for all tasks and agents. We will use various artificial ones while deciding the element (IV). In a point estimation problem, a typical example is use of the squared error loss function, namely,

$$L(\theta, a) = (\theta - a)^2. \tag{2.2}$$

Using Eq. (2.2), the risk $R(\theta, d)$ of a decision rule is just the mean squared error of $d(X)$ as an estimator of θ, *i.e.*, $\mathbb{E}_\theta\{d(X) - \theta\}^2$. In this context, our agent could aim for a decision rule d that minimize this mean squared error.

Other commonly used loss functions are the absolute error,

$$L(\theta, a) = |\theta - a|, \tag{2.3}$$

and

$$L(\theta, a) = \begin{cases} 0, & \text{if } |\theta - a| \leq \delta, \\ 1, & \text{if } |\theta - a| > \delta, \end{cases} \tag{2.4}$$

where δ is some prescribed tolerance limit.

Typically, an inferrer will devise one appropriate to the circumstances. This is not as easy as it sounds. Usually, after choosing a loss function $L(\theta, a)$, there is no way to find a decision rule $d \in \mathcal{D}$ which makes the risk function $R(\theta, d)$ uniformly smallest for all values of θ. Instead, it is necessary to consider several criteria, which help to narrow down the class of decision rules we consider in the hope of end up with a unique decision rule for the given inference problem. This optimal decision rule is what we call the right thing to do.

With a rational agent you will get what you asked for. Consider, for example, a quantum state tomographer agent. We can choose to infer states maximizing entropy while minimizing the mean squared error of the frequencies as performance measure. We will see this when we present our different inferece schemes in further chapters. *As a general rule, it is better to design performance measures according to what one actually wants in the environment, rather than according to how one thinks the agent should behave*.

With this in mind, we can define a rational agent properly:

Definition 2.4 (Rational Agent)**.** *For each possible given data (percept sequence), a rational agent should select a decision rule that minimize the risk (or maximize its performance measure), given the evidence provided by the given data (percept sequence) and whatever built-in knowledge the agent has.*

Unfortunately, the world is a highly demanding place. The computational requirements to always do the right thing at any given time can increase rapidly. This led us to deal with the issue of limited rationality, *i.e.*, acting appropriately when there is not enough time, or memory, to do all the computations one would like. That is why we choose to pursue the matter with *efficient rational agents*, limiting ourselves with polynomial algorithms.

2.3 Omniscience, learning, and autonomy

Rationality is different from perfection. Rationality maximizes *expected* performance, while perfection maximizes *actual* performance. Yet, if we expect an agent to do what turns out to be the best action after the fact, it will be impossible to design an agent to fulfil this specification.

That is why we need to be careful while distinguishing between rationality and *omniscience*. An omniscient agent knows the *actual* outcome of its actions after the fact and can act accordingly, but omniscience is impossible[2].

[2]Unless we improve the performance of our foretelling.

Our definition of rationality in Def. 2.4 does not require omniscience: our rational choice depends only on the percept sequence *to date* (*cf.* Def. 2.2).

The agent's initial configuration could reflect some prior knowledge of the environment with the possibility of acquiring new data. There are extreme cases in which the environment is completely known *a priori*. In such cases, the agent need not perceive or learn: it simply acts correctly. For example, suppose we have all the information we need to infer one state and this information has no errors of any kind. In this case, there is no inference, the agent should just do a linear inversion. We will have more to say about this scenario in Chapter 3.

To the extent that an agent relies on the prior knowledge of its designer rather than on its own percepts, we say that the agent lacks *autonomy*. A rational agent should be autonomous, *i.e.*, it should learn what it can to compensate for partial or incorrect prior knowledge. As a practical matter, one rarely requires complete autonomy from the start: when the agent has had little or no information, it could act randomly or it could give some non-informative inference (in our scenario could infer the normalized identity as result). After enough perceived information of its environment, the behavior of a rational agent can become effectively independent of its prior knowledge. Hence, the inclusion of learning allows one to design a single rational agent that will succeed in a vast variety of environments.

Now that we have a definition of rationality, we are almost ready to think about building rational agents. One way of summarizing what we talked about in this chapter is thinking about task environments, which are essentially the "problems" to which rational agents are the "solutions".

2.4 Specifying the task environment

We begin by showing how to specify a task environment. In our discussion on the rationality of the quantum state inferrer agent, we had to specify the performance measure, the environment, and the agent's actuators and sensors. We group all these under the heading of the task environment, and call it PEAS (Performance, Environment, Actuators, Sensors) description.

In designing an agent, the first step must always be to specify the task environment as fully as possible. First, what is the performance measure to which we would like our quantum state inferrer to aspire? Desirable qualities include goodness of fit of the measured probabilities; trace distance between the estimated state and the reference one; amount of entanglement and so on. Obviously, some of these goals conflict, so it can require some tradeoffs.

We have sketched the basic PEAS elements for our quantum state inferrer in Tab. 2.1.

Next, what is the inferring environment that the agent will face? In our case, our environment consists of a laboratory, which our data come from: it can be an optical table, quantum antenna, or some other physical system where we perform our measurements. The agent could also interact with the laboratory, for instance, we can have preliminary measurements to align our apparatus

Agent type	Performance measure	Environment	Actuators	Sensors
Quantum state inferrer	Goodness of fit, trace distance, amount of entanglement	laboratory, optical table, quantum antenna	screen output, write files, send packets, control apparatus	CCD cameras, photon counting devices, voltmeter

Table 2.1: PEAS description of the task environment for a quantum state inferrer.

autonomously. Obviously, the more restricted the environment, the easier the design problem.

The actuators for our inferrer include those available to a software agent: display something in the screen, write files, send network packets, control the electronical and mechanical apparatus and so on. The sensors for our agent can be CCD cameras, photon counting devices, voltmeter, and other physical measurement devices.

This concludes the ideas involved when looking at the problem of constructing a rational agent. With this skeleton, we will be able to elaborate separate aspects of our agent to infer quantum states, maps, and properties of physical systems.

CHAPTER 3

Quantum state inference on informationally complete contexts

The task of our quantum inferrer agent is receive, analyze the percept sequence and infer ρ.

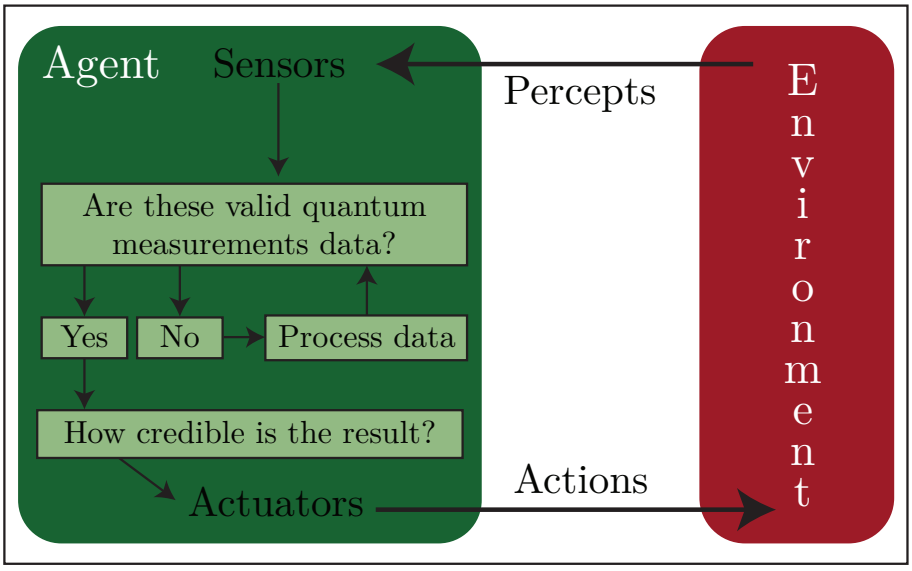

Figure 3.1: First interaction of our quantum inferrer agent: data processing in order to describe a valid quantum experiment.

In this chapter, we will elaborate our agent to execute the task in Fig. 3.1. First, we must check whether our data satisfy our working hypotheses. Then,

depending on the result, we process it to make it so; finally, we apply our performance measures, which will inform us how credible are these data.

3.1 Working hypotheses

Our agent will rationalize based on five hypotheses:

- H_0 : Quantum mechanics is a valid physical theory;
- H_1 : The data came from a quantum experiment;
- H_2 : We prepare the same state ρ, or *effectively* ρ;
- H_3 : We trust in our measurement apparata;
- H_4 : *Born's rule* is the quantum mechanical rule for associating measurement probabilities with states and effects, namely,

$$p(E_i = x_i \,|\, \rho,\, E) = \mathrm{tr}(\rho\, E_i), \tag{3.1}$$

where the set $\{E_i\}$ is a POVM (see Def. 3.3).

Let us explain our hypotheses: H_0 express the fact our agent is not concerned with the validity of quantum mechanics itself, so it will consider it valid and sound; H_1 is a possible ground were, as we will see later, we will rely on (the positive semidefiniteness of ρ, see Def. 3.2); our objective is one proper inferred ρ, so, we assume H_2 also in order to guarantee the agent's work reliability; H_3 is necessary to guarantee that we are measuring a given E_i, but we can relax H_3 to an ϵ-good approximation of E_i in the trace-norm (see Def. 3.1), namely,

$$\left\| E_i^{\text{ideal}} - E_i^{\text{actual}} \right\|_1 \leq \epsilon, \quad \text{for some small } \epsilon > 0; \tag{3.2}$$

and, finally, H_4 connects the measured data perceived in the sensors with the assumed quantum mechanical theory. We can read the left-hand side of Eq. (3.1) as "the probability of measuring the effect E_i and obtaining the outcome x_i, *given* we prepare the system in a state ρ and E_i is a possible event in the set $\{E_i\}$.

Now that we explained our working assumptions, we are in place to define some terms extensively used in this thesis:

Definition 3.1 (Schatten p-norm)**.** *For any operator $A \in \mathcal{L}(\mathcal{H})$ and any real number $p \geq 1$, one defines the Schatten p-norm of A as*

$$\|A\|_p = \left[tr\Big((A^\dagger A)^{p/2} \Big) \right]^{1/p}.$$

We also define

$$\|A\|_\infty = max\{\|A|u\rangle\| \,:\, |u\rangle \in \mathcal{H},\, \||u\rangle\| = 1\},$$

which coincides with $\lim_{p\to\infty} \|A\|_p$.

Definition 3.2 (Density Matrix)**.** *We can represent a quantum state in finite dimensions by a* Density Matrix ρ*, a* $d \times d$ *(with* $d \in \mathbb{N}$*,* $d \geq 1$*) linear operator in* $\mathcal{L}(\mathcal{H})$*;* ρ *is positive semi-definite (having non-negative eigenvalues); it has trace equals one; and* $n = d^2 - 1$ *linearly independent parameters.*

Definition 3.3 (Positive Operator Valued Measure)**.** *We call a set* $\{E_i\}$ *a* Positive Operator Valued Measure (POVM) *when the elements* E_i *are a positive semi-definite operators with eigenvalues smaller than one and* $\sum_i E_i = \mathbb{1}$*. A POVM describes a measurement with i labelling the possible events.*

Definition 3.4 (Projective Measurement)**.** *We call a set of orthogonal projectors* $\{P_i\}$ *a* ProjectiVe Measurement *(PVM) when* $\sum_i P_i = \mathbb{1}$ *and* $P_i^2 = P_i$*. A PVM describes a measurement it is a special case of a POVM.*

Definition 3.5 (Informationally Incomplete POVM)**.** *We call a POVM* informationally incomplete *when the set* $\{E_i\}$ *has less than* $d^2 - 1$ *linearly independent operators.*

Definition 3.6 (Informationally Complete POVM)**.** *We call a POVM* informationally complete *when the set* $\{E_i\}$ *has exactly* $d^2 - 1$ *linearly independent operators.*

Definition 3.7 (Symmetric Informationally Complete POVM [61])**.** *We call a POVM* symmetric informationally complete *when the set* $\{E_i\}$ *is an IC-POVM (according to Def. 3.6) and* $tr\left(E_i E_j\right) = c, \forall\, i \neq j$ *for the same constant c.*

Definition 3.8 (Super-complete POVM)**.** *We call a POVM* super-complete *when the set* $\{E_i\}$ *has more than* $d^2 - 1$ *linearly independent operators and other* $m \in \mathbb{N}$ *and* $m > 0$ *linearly dependent ones.*

3.2 The (im)perfect experiment

In an ideal world, experiments with ideal measurement apparata and perfect preparation schemes give us a set of noiseless data. In this case, we do not need to discuss about noise sources, faulty detectors, blind counts and so forth. Our agent will use Born's rule (3.1) and solve one linear system to handle the tomography scheme.

We need to be careful to distinguish between complete and noiseless data. We relate the completeness of a given information only with the number of linearly independent parameters collected in the percept sequence. Nevertheless, one can have enough noise in the system to render the complete information useless, but we will treat them as separate entities in this thesis.

Suppose we have a percept sequence with actual probabilities p_i and we chose to measure an Informationally Complete POVM (IC-POVM) defined in Def. 3.6. We can write the prepared state $\rho \in \mathcal{L}(\mathcal{H}^d) \cong \mathcal{L}(\mathbb{C}^d)$ as

$$\rho = \sum_{i=1}^{d^2-1} \theta_i E_i, \tag{3.3}$$

where E_i form a basis in Hilbert-Schmidt space and θ_i is the solution to the following linear system [54]:

$$G^{-1}\vec{p} = \vec{\theta}, \quad G_{ij} = \mathrm{tr}\big(E_i E_j\big). \tag{3.4}$$

Here, we kept the notation from classical statistics concerning $\vec{\theta} \in \Theta$ (*cf.* Section 2.2). The *parameter space* Θ usually is a subset of $\mathbb{R}^n$ for some $n \geq 1$ (in this case, $n = d^2 - 1$). Depending on the basis, p_i can be equal to θ_i, but it is not the general case. One should note that the choice of measuring an IC-POVM also implies G having an inverse.

Yet, the world is a dirty place and we have limited resources. Typically, instead of having the actual probabilities p_i, we possess N observations of different events, corresponding to the set $\{E_i\}$ (still assumed to be an IC-POVM). Therefore, the perceived data (a generic element of the sample space $\mathcal{X}$) has the form $\vec{x} = (x_1, \ldots, x_N)^T \in \mathbb{R}^N$.

The first real-world task of our agent is choosing a decision rule (*cf.* Section 2.2) to process $\vec{x}$ into $\vec{\theta}$, considering we do not have direct access to $\vec{p}$. Our agent can decide as following:

$$d(x_i) \equiv \frac{\text{Number of times that we observed } x_i}{\text{Total number of observations}} \equiv f_i. \tag{3.5}$$

We call Eq. (3.5) the *frequentist approach*. The frequentism is an interpretation of probability and defines an event's probability as the limit of its relative frequency f_i in a large number of trials, *i.e.*,

$$p_i \equiv \lim_{N\to\infty} \frac{\#x_i}{N}, \tag{3.6}$$

where $\#x_i$ counts the number times that we observed x_i.

There are different interpretations of probability leading to different choices of $d(x_i)$, but we will use the frequentist approach for the sake of simplicity and operationality. Thus, the first implementation of our agent could be the following:

(i) : Colect x_i;

(ii) : Use the decision rule $d(x_i)$ in Eq. (3.5) to produce f_i;

(iii) : Assume that $f_i \approx p_i$;

(iv) : Infer θ using the linear system in Eq. (3.4);

(v) : Construct ρ using Eq. (3.3) and display it.

This implementation has some flaws: the assumption in the step (iii) is very strong, we can have a small number of observations, rendering the approximation unreasonable, and it also fails to consider that sources of noise can disturb the measurements; the method in step (iv) does not take into account the losses and, consequently, the risk of choosing the decision rule in Eq. (3.5); and, finally, we cannot guarantee the positive semi-definiteness of ρ in step (v). Therefore, we need better methods connecting $d(x_i)$ with valid quantum probability distributions.

3.3 Maximum likelihood estimation

The well-known statistically based approach employed in quantum tomography is the Maximum Likelihood Estimation (MaxLik) [40, 28, 60, 31]. MaxLik searches for an estimate of the density matrix such that the observed data are most likely:

$$\begin{aligned} \max_{\rho} \quad & \mathcal{L}(\rho \,|\, d(X), E) \equiv q(d(X) \,|\, \rho, E) \\ \text{subject to} \quad & \rho \succeq 0\,; \\ & \operatorname{tr}(\rho) = 1, \end{aligned} \tag{3.7}$$

where $q(d(X) \,|\, \rho, E)$ is the joint probability density function of the observed data X given ρ over the set $\{E_i\}$, considering some decision rule $d(\cdot)$ upon X.

The MaxLik approach depends on the parametric model assumed for $\mathcal{L}(\rho \,|\, d(X), E)$ (cf. [74]). We may use for example a multinomial distribution, as considered in [28]:

$$\mathcal{L}(\rho \,|\, d(X), E) = \frac{N!}{\prod_j \#x_j} \prod_i p(E_i = x_i \,|\, \rho,\, E)^{\#x_i} = \frac{N!}{\prod_j \#x_j} \prod_i \operatorname{tr}(\rho E_i)^{\#x_i}, \tag{3.8}$$

where ρ is a variable, $\#x_i$ counts the number times that we observed the outcome x_i and N is the total number of observations $N = \sum_i \#x_i$.

In photocounting experiments [1], we have $\#x_i$ photon count rates, and considering linear measurement model, we have

$$N\operatorname{tr}(\rho E_i) - \#x_i = N\Delta_i, \quad \forall\, i = 1, \ldots, m, \tag{3.9}$$

(or frequencies $\operatorname{tr}(\rho E_i) - f_i = \Delta_i$) where Δ_i are independent, identically distributed (i.i.d.) measurement errors.

While seeing Eq. (3.9), one might remember Section 2.2, where we defined an abstract loss function $L(\theta,\, d(X)) = |\theta - d(X)|$ in Eq. (2.3). This is not a coincidence, since the error Δ_i captures the idea of an incurred loss when we choose to represent $\operatorname{tr}(\rho E_i)$ — which is our parameter of interest θ_i — by f_i (in the case of frequentist decision rule in Eq. (3.5)).

There is another model used in photonic tomography, proposed by James *et al.* [31],

$$\mathcal{L}(\rho \,|\, d(X), E) = \frac{1}{N_{norm}} \prod_i \exp\left(-\frac{1}{2} \frac{(N \operatorname{tr}(\rho E_i) - \#x_i)^2}{\sigma_i^2} \right), \tag{3.10}$$

where $\sigma_i = \sqrt{\#x_i}$ is the standard deviation in the Poissonian distribution. This model assumes that the photon detection counting is well approximated by a Gaussian (when the count numbers are large enough).

In practice, instead of maximizing $\mathcal{L}(\rho \,|\, d(X), E)$, we maximize the logarithm of it (also called *log-likelihood function* [74]), in case of Eq. (3.10), reads:

$$\ell(\rho \,|\, d(X), E) \propto -\sum_i \left(\frac{(N \operatorname{tr}(\rho E_i) - \#x_i)^2}{\sigma_i^2} \right), \tag{3.11}$$

i.e., Eq. (3.11), resembles the squared error loss function $L(\theta,\, d(X)) = (\theta - d(X))^2$ defined in Eq. (2.2) of Section 2.2.

Moreover, we may interpret *any* penalty function approximation problem

$$\min_{\rho} \quad \sum_{i} \phi(N \operatorname{tr}(\rho E_i) - \#x_i) \tag{3.12}$$

as a maximum likelihood estimation problem with noise density

$$\mathcal{L}(\rho \,|\, d(X), E) = \frac{e^{-\phi(d(X))}}{\int e^{-\phi(\mu)} d\mu}, \tag{3.13}$$

measurements X (*cf.* [7] for more details), and some decision rule $d(\cdot)$ upon X. One may see this observation as a statistical interpretation of the penalty function approximation problems.

For a general non-linear $\mathcal{L}(\rho \,|\, d(X), E)$, solving (3.7) is a challenge on its own. Usually, numerical methods for specific likelihoods are still computationally inefficient and expensive. In case of (3.8), the $R\rho R$ algorithm and variants [28, 60] have shown practical behavior. For the likelihood (3.10) a common method is to parametrize the density matrix as follows:

$$\rho = \frac{T^\dagger T}{\operatorname{tr}(T^\dagger T)}, \tag{3.14}$$

where T is an upper triangular matrix, to fall into an unconstrained optimization problem in the entries of T. Although unconstrained, this new problem in T is non-convex and plagued with many local maxima, but Gonçalves *et al.* [21] proved that these local maxima are all global.

3.4 Quantum tomographies via SDPs

As we said in Chapter 2, we will pursue an *efficient rational agent*, where we limit ourselves in using polynomial algorithms.

In [45], we introduced our first scheme: the Variational Quantum State Tomography (VQT). We formulated VQT as an SDP (Semi-Definite Programming). An SDP is a class of optimization problems characterized by having a linear objective function, constrained by linear matrix inequalities (LMIs). One advantage of SDPs is that we can solve them efficiently using modern interior-point algorithms. For more details about SDPs and its efficient solutions, *cf.* [35, 7, 25, 53].

Suppose we are measuring an IC-POVM (Def. 3.6) and using the frequentist decision rule in Eq. (3.5). We can write the VQT scheme presented in [45] as

$$\begin{aligned} \min_{\rho,\, \vec{\Delta}} \quad & \sum_{i=1}^{d^2-1} \Delta_i \\ \text{subject to} \quad & \rho \succeq 0\,; \\ & \operatorname{tr}(\rho) = 1\,; \\ & |\operatorname{tr}(\rho E_i) - f_i| \leq \Delta_i f_i\,, \quad i = 1, \ldots, d^2 - 1; \\ & 0 \leq \Delta_i \leq 1\,. \end{aligned} \tag{3.15}$$

Let us analyze the problem in (3.15) line by line. The first line has our objective function: where we are minimizing our expected losses (or errors). In the second and third lines, we are constraining ourselves with positive semi-definite matrices of trace one, this will play the role of our hypothesis H_2 in Section 3.1. We define our losses in the last two lines: in this case, each individual loss cannot be greater than the observed relative frequency f_i.

Sometimes, modeling our losses as in (3.15) can be too restrictive to the point of being infeasible, but SDPs are flexible enough to allow different loss functions. For example, we saw in the last section that we can relate Gaussian noise with least squares minimization. So, if we want a solution like (3.11), we can write the following:

$$\begin{aligned} \min_{\rho,\,\vec{\Delta}} \quad & \sum_{i=1}^{d^2-1} \Delta_i \\ \text{subject to} \quad & \rho \succeq 0\,; \\ & \operatorname{tr}(\rho) = 1\,; \\ & \frac{(\operatorname{tr}(\rho E_i) - f_i)^2}{\sigma_i^2} \leq \Delta_i\,, \quad i = 1, \ldots, d^2-1; \\ & \Delta_i \geq 0\,. \end{aligned} \tag{3.16}$$

We can also try

$$\begin{aligned} \min_{\rho,\,\Delta} \quad & \Delta \\ \text{subject to} \quad & \rho \succeq 0\,; \\ & \operatorname{tr}(\rho) = 1\,; \\ & \left\| M\,vec(\rho) - \overrightarrow{f} \right\|_2 \leq \Delta \\ & \Delta \geq 0\,, \end{aligned} \tag{3.17}$$

where $vec(\rho)$ is the vector representation of ρ (in Dirac's notation, this results in flipping a bra to a ket, $vec(|a\rangle\langle b|) \equiv |a\rangle|b\rangle$, with $|a\rangle$,$|b\rangle$ being standard basis elements) and M is a measurement matrix, with each row being $vec(E_i)^T$. The problem (3.17) is a conic program and uses the ℓ_2-norm for vectors (we could use ℓ_1-norm as well).

There is also another scheme in [3] using Bayesian estimates, namely,

$$\begin{aligned} \min_{\rho,\,\tau} \quad & \tau \\ \text{subject to} \quad & \rho \succeq 0\,; \quad \operatorname{tr}(\rho) = 1; \quad \tau \geq 0\,; \\ & \left(M\,vec(\rho) - \overrightarrow{\mu}\right)^T \Sigma^{\dagger} \left(M\,vec(\rho) - \overrightarrow{\mu}\right) \leq \tau\,, \end{aligned} \tag{3.18}$$

where $\vec{\mu}$ is a Bayesian estimate of the mean, $\Sigma^{\dagger}$ is the Moore-Penrose pseudoinverse of the Bayesian estimate of the covariance matrix and τ plays the role of the ℓ_2-distance. We are not entering in details of Bayesian estimation.

3.5 Agent's design

Now, we have everything to design our agent, again, assuming we are measuring an IC-POVM (Def. 3.6) and using the frequentist decision in Eq. (3.5):

(i) : Colect x_i;

(ii) : Use the decision rule $d(x_i)$ in Eq. (3.5) to produce f_i;

(iii) : Execute programs (3.15), (3.16) and (3.17);

(iv) : Exclude unfeasible solutions and compare the expected losses;

(v) : Choose and display the best estimate of ρ.

We have a better design compared to our first one in Section 3.2. Now we have methods that minimize expected losses while respecting our working hypotheses in Section 3.1. We illustrated this in Fig. 3.2.

Choosing the best estimate of ρ will depend on what we want. One possible choice is selecting the state with the maximum entropy, or, sometimes, the state with the least entanglement according to some appropriate measure.

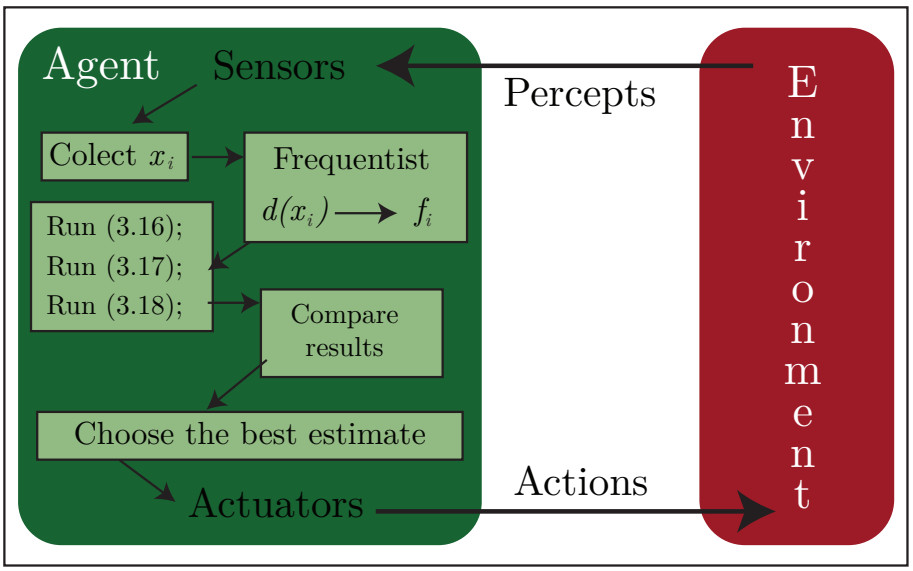

Figure 3.2: Last design of our agent for the informationally complete context.

CHAPTER 4

Estimation on informationally *in*complete contexts: the need for inference

In this chapter, we will relax the assumption of measuring an informationally complete POVM, *i.e.*, we are now using Def. 3.5 for our set $\{E_i\}$. Being in this context means that we possess less than $d^2 - 1$ linearly independent coefficients. This lack of information has important consequences and we need to address them carefully.

Whenever we have an incomplete set of measurements, the inferred state may not be uniquely determined by the available data. Therefore, we need suitable estimation methods for correct interpretation and quantum diagnostics, which will not bias the undetected components.

The quantum version of Jaynes' Principle [32] of Maximum Entropy estimation (MaxEnt) [9] is usually the tool of choice in this scenario. However, this is not an easy task due to the nonlinearities in the reconstruction algorithm.

Another possible approach, which avoids nonlinear difficulties, is using Compressed Sensing [24], our Variational Quantum Tomography (VQT) [45] and our revised version of VQT in [22].

Here, we analyze the similarities and differences between VQT and MaxEnt for quantum tomography with incomplete data, from the first method [45] towards the revised version [22]. First, we examine the behavior of these two methods in setting the probabilities related to the missing information. Then, we show that our revised version captures the idea behind the Maximum Entropy approach, in the sense that it sets the unmeasured probabilities as uniform as possible. This modified VQT is still an SDP, so it is still an efficient formulation for our agent.

4.1 Variational Quantum Tomography

We introduced the Variational Quantum Tomography in Section 3.4 in the context of complete data. Although still assuming the frequentist decision rule in Eq. (3.5), we are now measuring an informationally incomplete POVM (Def. 3.5) and we need to add an extra term in our objective function to deal with this lack of information, namely

$$\begin{aligned} \min_{\rho,\,\vec{\Delta}} \quad & \sum_{i\in\mathcal{I}} \Delta_i + \sum_{i\notin\mathcal{I}} \mathrm{tr}(E_i\rho) \\ \text{subject to} \quad & \rho \succeq 0\,; \\ & \mathrm{tr}(\rho) = 1\,; \\ & |\mathrm{tr}(\rho E_i) - f_i| \leq \Delta_i f_i\,, \quad i \in \mathcal{I}; \\ & 0 \leq \Delta_i \leq 1\,. \end{aligned} \tag{4.1}$$

The difference from Eq. (3.15) is the extra term in our objective function and the set of indexes of measured data $\mathcal{I}$ (*cf.* Section 3.4 for other details of the formulation).

With this extra term in the objective function,

$$\sum_{i\notin\mathcal{I}} \mathrm{tr}(E_i\rho),$$

we can choose the solution that also minimizes the expectation values of the unmeasured observables. Minimizing the unknown subspace is a reasonable model, and we successfully applied this inference scheme in an actual experiment [43, 19].

However, the problem (4.1) does not have unique solution, *i.e.*, we can have more than one estimate that minimizes our expected losses and satisfies the constraints. This motivated us to revise our method and we will give more details in the next sections.

4.2 Maximum Entropy estimation

Buzek *et al.* [9] introduced the Maximum Entropy approach in the context of quantum state tomography with incomplete data. The idea is to take as an estimate for ρ the density matrix that maximizes the von Neumann entropy,

$$S(\rho) = -\mathrm{tr}(\rho \ln \rho), \tag{4.2}$$

and is compatible with the available observed data. This estimate is the solution of the following optimization problem:

$$\begin{aligned} \max_{\rho} \quad & -\mathrm{tr}(\rho \ln \rho) \\ \text{subject to} \quad & \rho \succeq 0, \\ & \mathrm{tr}(\rho) = 1, \\ & \mathrm{tr}(E_i\rho) = f_i, \quad i \in \mathcal{I}. \end{aligned} \tag{4.3}$$

Although we have to deal again with constraints in the semidefinite positive cone, it is possible to devise an explicit solution for (4.3), using the first order optimality conditions.

Applying the KKT first order optimality conditions [34, 37], we obtain

$$\begin{aligned} \ln(\rho) + \mathbb{1} + \lambda_0 \mathbb{1} + \sum_{i\in\mathcal{I}} \lambda_i E_i &= S, \\ \operatorname{tr}(E_i\rho) &= f_i,\ i \in \mathcal{I} \\ \operatorname{tr}(\rho) &= 1, \\ \rho \succeq 0, &\quad S \succeq 0 \\ \rho S &= 0, \end{aligned} \tag{4.4}$$

where $\mathbb{1}$ is the identity operator and λ_i are Lagrange multipliers. Assuming that $\rho \succ 0$, we get

$$\begin{aligned} \ln(\rho) + \mathbb{1} + \lambda_0 \mathbb{1} + \sum_{i\in\mathcal{I}} \lambda_i E_i &= 0, \\ \operatorname{tr}(E_i\rho) &= f_i,\ i \in \mathcal{I} \\ \operatorname{tr}(\rho) &= 1. \end{aligned} \tag{4.5}$$

From the first equation in (4.5), we obtain

$$\rho = \exp\left(-\sum_{i\in\mathcal{I}} \lambda_i E_i - \lambda_0 \mathbb{1} - \mathbb{1} \right) \succ 0, \tag{4.6}$$

and defining $1/\exp(-\lambda_0 - 1) = \mathcal{N} = \operatorname{tr}(\exp \sum_i -\lambda_i E_i)$, we have

$$\rho_{ME} = \frac{1}{\mathcal{N}} \exp \sum_{i\in\mathcal{I}} -\lambda_i E_i. \tag{4.7}$$

Clearly, $\rho_{ME} \succ 0$ and $\operatorname{tr}(\rho_{ME}) = 1$ due to the normalization constant $\mathcal{N}$. We can determine the Lagrange multipliers solving the non-linear system of equations

$$\operatorname{tr}\left(E_j \rho_{ME}\right) = f_j, \quad j \in \mathcal{I}, \tag{4.8}$$

that is,

$$\operatorname{tr}\left(E_j \exp\left(\sum_{i\in\mathcal{I}} -\lambda_i E_i \right) \right) = \mathcal{N} f_j, \quad j \in \mathcal{I}. \tag{4.9}$$

Since we usually have noisy data $\{f_i\}$, we solve the following non-linear least-squares problem

$$\min_{\lambda} \sum_{j\in\mathcal{I}} \left[\operatorname{tr}\left(E_j \exp\left(\sum_{i\in\mathcal{I}} -\lambda_i E_i \right) \right) - \mathcal{N} f_j \right]^2, \tag{4.10}$$

instead of the non-linear equations (4.9).

4.3 VQT, Eigenbasis Measurements and Maximum Entropy

Since we are considering quantum tomography with incomplete measurements, an important point is how the methods assign the probabilities associated to unmeasured POVM elements. To simplify our analysis, we will consider first the case of eigenbasis measurements in an ideal experiment (free of noise). We can write the true state ρ using spectral decomposition:

$$\rho = \sum_{i \in \mathcal{I}} c_i P_i + \sum_{i \notin \mathcal{I}} c_i P_i, \tag{4.11}$$

where P_i's are d orthonormal projectors onto the eigenspace of ρ. The task of tomography now is to determine the coefficients c_i's (eigenvalues) based on observed data. Assuming we have $m < d$ projectors (d is the dimension of the Hilbert space), and that $\{P_i\}$ is an orthonormal set, it is easy to show that

$$c_i = \mathrm{tr}(P_i \rho), \ \ \forall\, i, \tag{4.12}$$

and since we have an ideal experiment, we also obtain

$$\mathrm{tr}(P_i \rho) = f_i, \ \ \forall i \in \mathcal{I}, \tag{4.13}$$

where f_i's are the measured data. Due to the normalization constraint, we have that

$$\sum_{i \notin \mathcal{I}} c_i = 1 - \sum_{i \in \mathcal{I}} c_i = 1 - \sum_{i \in \mathcal{I}} f_i, \tag{4.14}$$

and the constraint $\rho \succeq 0$ implies that $c_i \geq 0, \forall i$.

Now let us consider the MaxEnt solution, given by (4.7):

$$\begin{aligned} \rho_{ME} &= \frac{1}{\mathcal{N}} \exp\left(\sum_{i \in \mathcal{I}} -\lambda_i P_i - \sum_{i \notin \mathcal{I}} 0 P_i \right) \\ &= \sum_{i \in \mathcal{I}} \frac{e^{-\lambda_i}}{\mathcal{N}} P_i + \frac{1}{\mathcal{N}} \sum_{i \notin I} P_i, \end{aligned} \tag{4.15}$$

where λ_i are the Lagrange multipliers related to the constraints (4.13). Since ρ_{ME} must satisfy those constraints, we have

$$f_j = \mathrm{tr}\left(P_j \left(\sum_{i \in \mathcal{I}} \frac{e^{-\lambda_i}}{\mathcal{N}} P_i + \frac{1}{\mathcal{N}} \sum_{i \notin I} P_i \right) \right) = \frac{e^{-\lambda_j}}{\mathcal{N}}. \tag{4.16}$$

Thus, $e^{-\lambda_i}/\mathcal{N} = f_i = c_i, \ \ \forall i \in \mathcal{I}$, as we expected. Moreover, as $\mathrm{tr}(\rho_{ME}) = 1$, we obtain

$$\sum_{i \in \mathcal{I}} f_i + \frac{d - m}{\mathcal{N}} = 1, \tag{4.17}$$

which implies that, for the unmeasured coefficients,

$$c_i = \frac{1}{\mathcal{N}} = \frac{1 - \sum_{j \in \mathcal{I}} f_j}{d - m}, \ \ \forall\, i \notin \mathcal{I}. \tag{4.18}$$

In other words, the MaxEnt solution

$$\rho_{ME} = \sum_{i\in\mathcal{I}} f_i P_i + \sum_{i\notin\mathcal{I}} \left(\frac{1 - \sum_{j\in\mathcal{I}} f_j}{d - m} \right) P_i, \tag{4.19}$$

uniformly distributes the remainder $1 - \sum_{j\in\mathcal{I}} f_j$ among the other coefficients $c_i, \forall\, i \notin \mathcal{I}$.

Now let us compare this solution with the VQT solution. Considering the formulation (4.1), and applying equations (4.12),(4.13),(4.14), we obtain

$$\begin{aligned} \min_{c_i, \forall i\notin\mathcal{I}} \quad & \sum_{i\notin\mathcal{I}} c_i \\ \text{subject to} \quad & \sum_{i\notin\mathcal{I}} c_i = 1 - \sum_{i\in\mathcal{I}} f_i, \\ & c_i \geq 0,\ \forall\, i \notin \mathcal{I}. \end{aligned} \tag{4.20}$$

Since we are assuming an ideal experiment, $c_i = \mathrm{tr}(P_i\rho) = f_i,\ \forall\, i \in \mathcal{I}$, we have $\Delta_i = 0, \forall\, i \in \mathcal{I}$. Furthermore, the variables of the problem (4.20) are $c_i, \forall\, i \notin \mathcal{I}$, and any feasible solution of (4.20) is also optimal, because the objective function is the same as the left hand side of the first constraint. Thus we can express the solution of VQT as

$$\rho_{VQT} = \sum_{i\in\mathcal{I}} f_i P_i + \sum_{i\notin\mathcal{I}} c_i P_i. \tag{4.21}$$

Therefore, there is no constraint or penalty in the objective function that forces $c_i, i \notin \mathcal{I}$, to agree with those of MaxEnt solution.

To guarantee that the VQT solution agrees with the MaxEnt solution, at least in the ideal eigenbasis case, we propose a change in the VQT formulation (4.1). Let us define the vector $\tilde{c}$, of size $d^2 - m$, with components

$$\tilde{c}_i = \mathrm{tr}(E_i\rho),\ \ \forall\, i \notin \mathcal{I}. \tag{4.22}$$

Considering that E_i are usually POVM elements (or projectors), we will assume that $\tilde{c}_i \geq 0$. Thus we have that

$$\|\tilde{c}\|_1 = \sum_{i\notin\mathcal{I}} |\mathrm{tr}(E_i\rho)| = \sum_{i\notin\mathcal{I}} \mathrm{tr}(E_i\rho). \tag{4.23}$$

Our proposal consists in using

$$\|\tilde{c}\|_\infty = \max_{i\notin\mathcal{I}} |\mathrm{tr}(E_i\rho)| = \max_{i\notin\mathcal{I}}\ \mathrm{tr}(E_i\rho), \tag{4.24}$$

instead of $\|\tilde{c}\|_1$ in the objective function of the problem (4.1). Whenever we fix the sum of the components of $\tilde{c}$, minimizing $\|\tilde{c}\|_\infty$ promotes a more uniform distribution of these coefficients.

In this case, the VQT$_\infty$ (VQT with $\|.\|_\infty$) formulation becomes

$$\begin{aligned} \min_{\rho,\,\Delta} \quad & \sum_{i\in\mathcal{I}} \Delta_i + \max_{i\notin\mathcal{I}}\ \mathrm{tr}(E_i\rho) \\ \text{subject to} \quad & |\mathrm{tr}(E_i\rho) - f_i| \leq \Delta_i f_i \qquad i \in \mathcal{I} \\ & \Delta_i\ \geq 0, \\ & \mathrm{tr}(\rho)\ = 1, \\ & \rho\ \succeq 0, \end{aligned} \tag{4.25}$$

and for the ideal eigenbasis case, the equivalent of (4.20) is

$$\begin{aligned} \min_{c_i, \forall i \notin \mathcal{I}} \quad & \max_{i \notin \mathcal{I}} c_i \\ \text{subject to} \quad & \sum_{i \notin \mathcal{I}} c_i = 1 - \sum_{i \in \mathcal{I}} f_i, \\ & c_i \geq 0, \ \forall i \notin \mathcal{I}. \end{aligned} \tag{4.26}$$

Notice that the problem (4.26) has a unique solution that corresponds to the uniform distribution of the remainder $1 - \sum_{i \in \mathcal{I}} f_i$ among the coefficients $c_i, \forall i \notin \mathcal{I}$, which coincides with the MaxEnt solution in this case.

Despite the agreement with MaxEnt in the eigenbasis case, the proposed modification does not turn the problem (4.25) harder than (4.1). This is true because minimizing $\|\tilde{c}\|_\infty$ is equivalent to minimizing some auxiliary variable δ subject to $|\tilde{c}_i| \leq \delta, \ \forall\, i \notin \mathcal{I}$. Therefore, we have again a linear SDP:

$$\begin{aligned} \min_{\rho,\, \Delta,\, \delta} \quad & \delta + \sum_{i \in \mathcal{I}} \Delta_i \\ \text{subject to} \quad & \rho \ \succeq 0, \\ & \operatorname{tr}(\rho) \ = 1, \\ & |\operatorname{tr}(E_i \rho) - f_i| \leq \Delta_i f_i \qquad i \in \mathcal{I}, \\ & \operatorname{tr}(E_i \rho) \leq \delta \qquad\qquad\ \ i \notin \mathcal{I}, \\ & 0 \leq \ \Delta_i \ \leq 1, \\ & \delta \ \geq 0. \end{aligned} \tag{4.27}$$

Although the equivalence between the solution of the VQT formulation (4.27) and the one of MaxEnt is true only for the eigenbasis case, the relationship between these two problems is clear: both of them, each in its own way, try to set the unmeasured probabilities the most uniformly as possible.

4.4 Noisy data and the MaxLik-MaxEnt approach

To assist or MaxEnt approach in case of noisy data, we will use with Maximum Likelihood (MaxLik) [40, 28, 60, 31], see Section 3.3. Regardless the method we use to solve (3.7), whenever we do not have a tomographically complete set of measurements or some data is missing (incomplete data), there is a convex set of maximizers of $\mathcal{L}(\rho \,|\, d(X), E)$, that is, a convex set of matrices for which the likelihood achieves its maximum value. Then the intersection of this set with the set of density matrices may be a non-empty convex set and, in this case, we have more than one solution for (3.7). Each of these solutions is compatible with the data, in the sense that minimizes some relative distance between the observed data f_j and the probabilities predicted by quantum mechanics $\operatorname{tr}\big(E_j \rho\big)$. However, these solutions differ on how to fit the unmeasured probabilities.

One way to choose a unique solution from the set of ML (Maximum Likelihood) solutions, in this case, is to apply the MaxEnt method constrained to the set of ML solutions. To accomplish this task, after we have one ML solution, say ρ_{ML}, we know that

$$\operatorname{tr}(E_i \rho) = \bar{p}_i \equiv \operatorname{tr}(E_i \rho_{ML}), \ \ \forall\, i \in \mathcal{I}, \ \ \forall\, \rho \in \mathcal{S}_{ML}, \tag{4.28}$$

where $\mathcal{S}_{ML}$ is the ML solution set. Therefore, we can obtain the Maximum Entropy solution among the ML solutions solving the problem (4.3) switching f_i by $\bar{p}_i$ for all $i \in \mathcal{I}$. This is the Maximum Likelihood - Maximum Entropy (MaxLik-MaxEnt) [59] estimate for ρ. Other possible approach is the joint maximization of likelihood and entropy through a Lagrangian function [68].

4.5 Numerical simulations: VQT$_\infty$ and MaxLik-MaxEnt

To compare the VQT$_\infty$ and the MaxEnt approach, for each rank, we sampled 1000 uniformly distributed density matrices[1] representing four-qubits states, and we fixed a SIC-POVM basis [61] (see Def. 3.7) to measure.

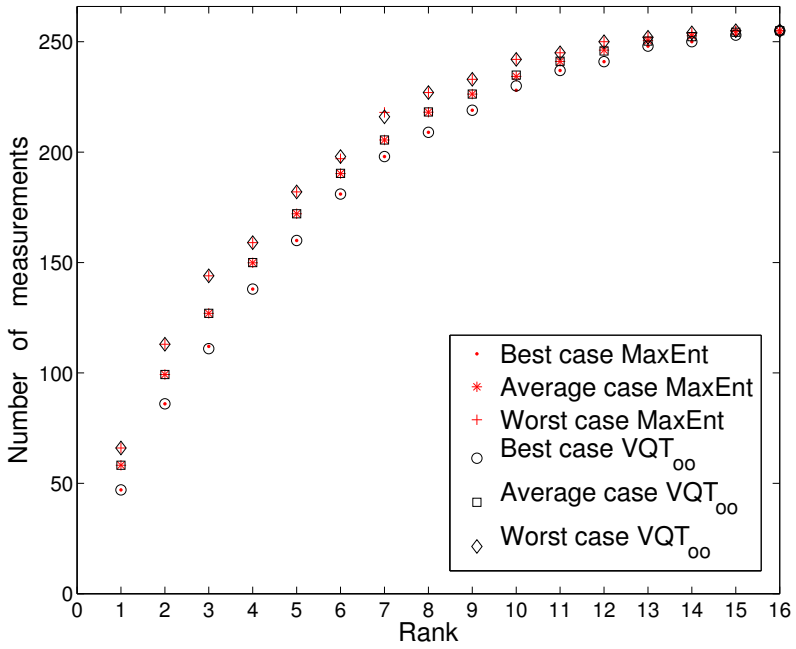

Figure 4.1: Convergence of VQT$_\infty$ and MaxEnt increasing the number of measurements.

For each rank, we computed the worst, the average and the least number of measurements needed for each method to converge to the reference state, depicted in Fig. 4.1. We considered that a method converges if the trace distance,

$$D(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_1, \quad \|\cdot\|_1 \text{ is the Schatten } p\text{-norm, see Def. 3.1,} \tag{4.29}$$

to the real state is less than 10^{-4}.

[1] According to the Haar measure [75] when we have a pure state and according to Hilbert-Schmidt measure [75] otherwise.

Firstly, we consider noise-free data. As one can see, the required number of measurements to converge is almost the same for MaxEnt and VQT$_\infty$ methods. Fig. 4.1 also shows that the worst case for each method does not exceeds the $\mathcal{O}(rd\log(d)^2)$ number of measurements (r is the rank) mentioned in compressed sensing works [24] in quantum tomography.

Afterwards, we compared the convergence of both methods in terms of average trace distance and average entropy for rank one states as the measurements increase, illustrated in Fig. 4.2. Again, one can see a similar behavior between MaxEnt and VQT$_\infty$, the former with a slightly smaller distance and greater entropy than the second.

The impact of the proposed modification (4.27) comes up when we compare the distance of $\tilde{c}$, the vector of remaining probabilities, to the uniform vector whose entries we define in (4.18). Fig. 4.4 shows, for ranks one and six, the average Kullback-Leibler divergence,

$$D_{KL}(\vec{p}\,||\,\vec{q}) \equiv -\sum_i \log\left(\frac{p_i}{q_i}\right), \quad p_i, q_i \geq 0, \quad \sum_i p_i = \sum_i q_i, = 1, \tag{4.30}$$

to the uniform distribution. As one can see, the VQT$_\infty$ is closer to the MaxEnt than the original VQT. When the number of measurements is sufficient to determine uniquely the state, then we observe the agreement of the methods.

To assess the convergence properties in a more realistic scenario we introduced two kinds of error in the true probabilities, one using a Gaussian perturbation with zero mean and standard deviation 10^{-6} and the other considering values uniformly distributed in an interval of 5% deviation of the true probabilities.

We used the MaxLik-MaxEnt method in comparison with VQT$_\infty$. For the likelihood function, we used a variant of (3.11). In Fig. 4.5, we plot the average trace distance depending on the number of measurements for random rank one states for the two kinds of error.

These numerical simulations corroborate the relationship between VQT$_\infty$ and MaxEnt methods proved in Section 4.3 for eigenbasis measurements. The new proposed formulation (4.25), as the MaxEnt approach, tries to fit the unmeasured probabilities the most uniformly as possible. We see that the results for these methods are quite close and the VQT$_\infty$ has the advantage of the linear SDP programming against non-linear optimization problems of MaxEnt. Each optimization using VQT takes no more than five seconds whereas MaxLik-MaxEnt tomography spends about twelve seconds in the worst case. We did the simulations on a Intel Core 2 Duo, 2 GB RAM computer, in MATLAB, using Yalmip/SEDUMI [44, 67] to model and solve the SDP problems.

4.6 Agent's design

Summarizing, we proposed a variant of the VQT method for quantum tomographies with incomplete information, namely VQT$_\infty$, that tries to fit the unmeasured probabilities to be the most uniformly as possible. Therefore,

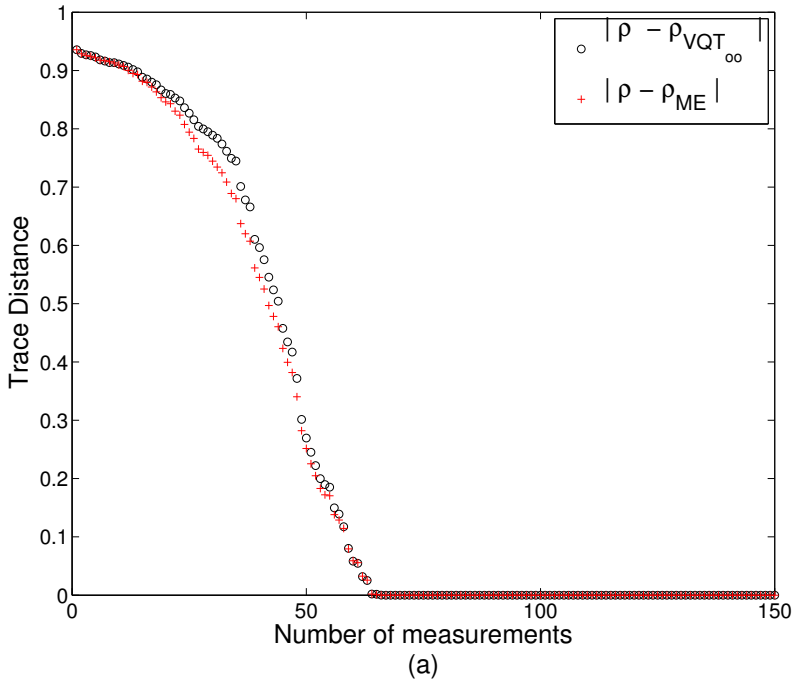

Figure 4.2: (a): Trace distance to the real rank-1 state for VQT_∞ and MaxEnt. (b): von Neumann entropy of VQT_∞ and MaxEnt estimates.

VQT_∞ has a similar behavior to the well-known MaxEnt approach in the noise-free case, and to the MaxLik-MaxEnt in the presence of noise. We confirmed this claim theoretically for the case of eigenbasis measurements, in our numerical simulations, and we implemented it successfully in actual experiments, *cf.* [57, 48].

With this work, we achieved the main goal of this thesis: now we can design a rational agent that can infer quantum states properly in the spirit of Jaynes' Principle (see Principle 2) using algorithms of polynomial complexity solely. The final design is the following (see Fig. 4.3):

(i) : Colect x_i;

(ii) : Use the decision rule $d(x_i)$ in Eq. (3.5) to produce f_i;

(iii) : Check the completeness of the measured $\{E_i\}_{i\in\mathcal{I}}$;

(iv) : If $\{E_i\}_{i\in\mathcal{I}}$ is informationally complete, execute (3.15), otherwise, execute (4.27) (here we chose one loss function for the probabilities in each case, but we can choose others like we did in Section 3.5);

(v) : Exclude unfeasible solutions and compare the expected losses;

(vi) : Choose and display the best estimate of ρ.

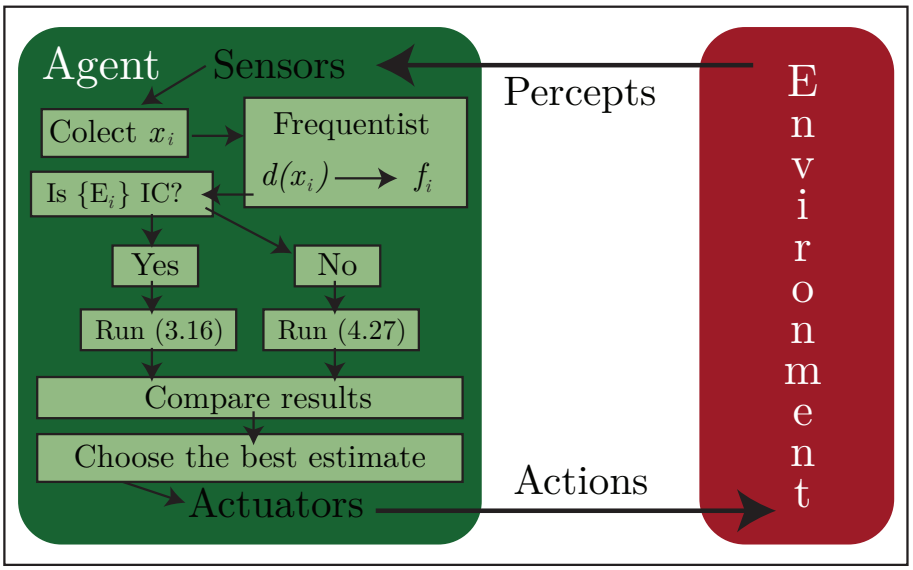

Figure 4.3: Final design of our quantum inferrer agent for the informationally incomplete context.

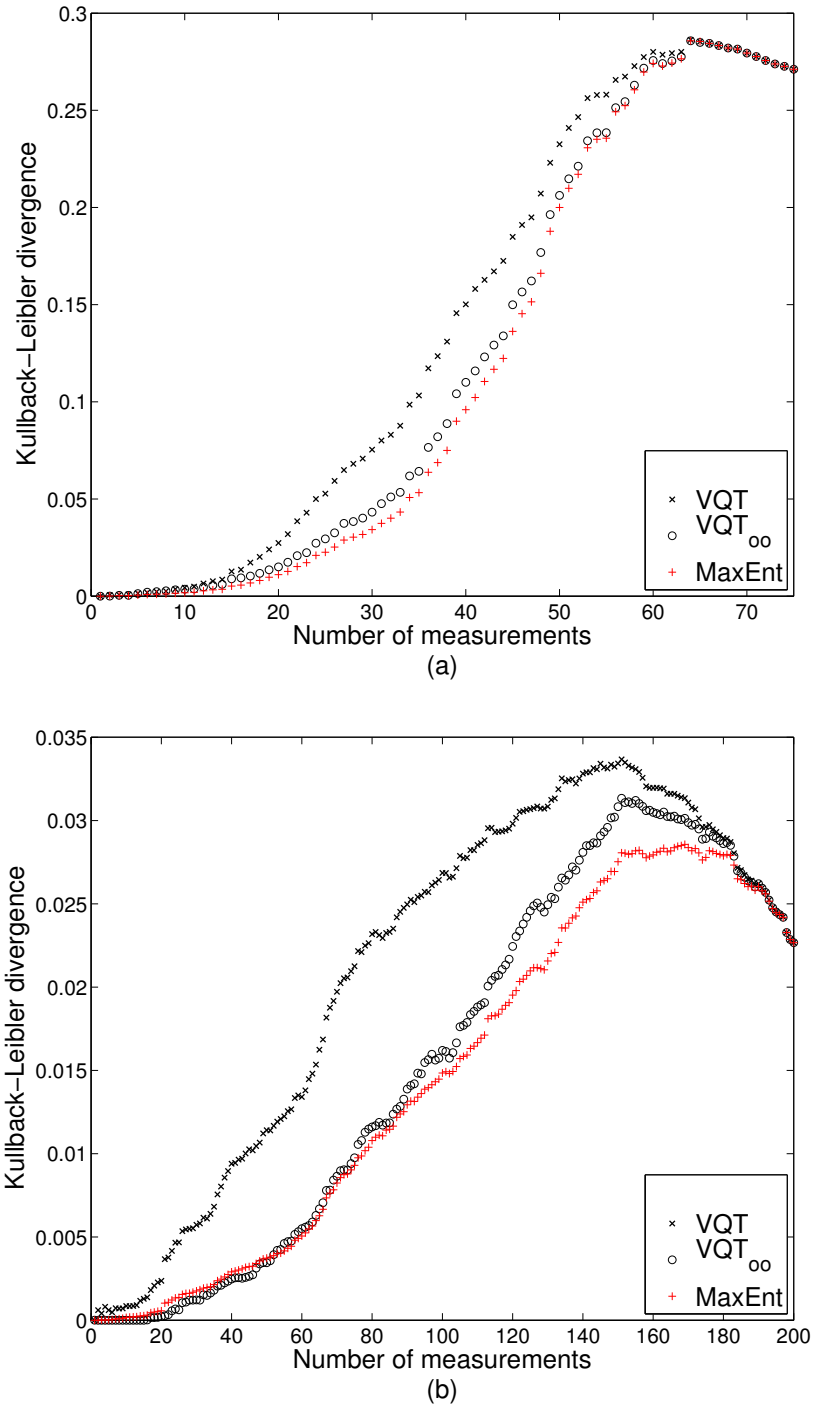

Figure 4.4: Average Kullback-Leibler divergence of $\tilde{c}$ to the uniform distribution. (a) rank one, (b) rank six.

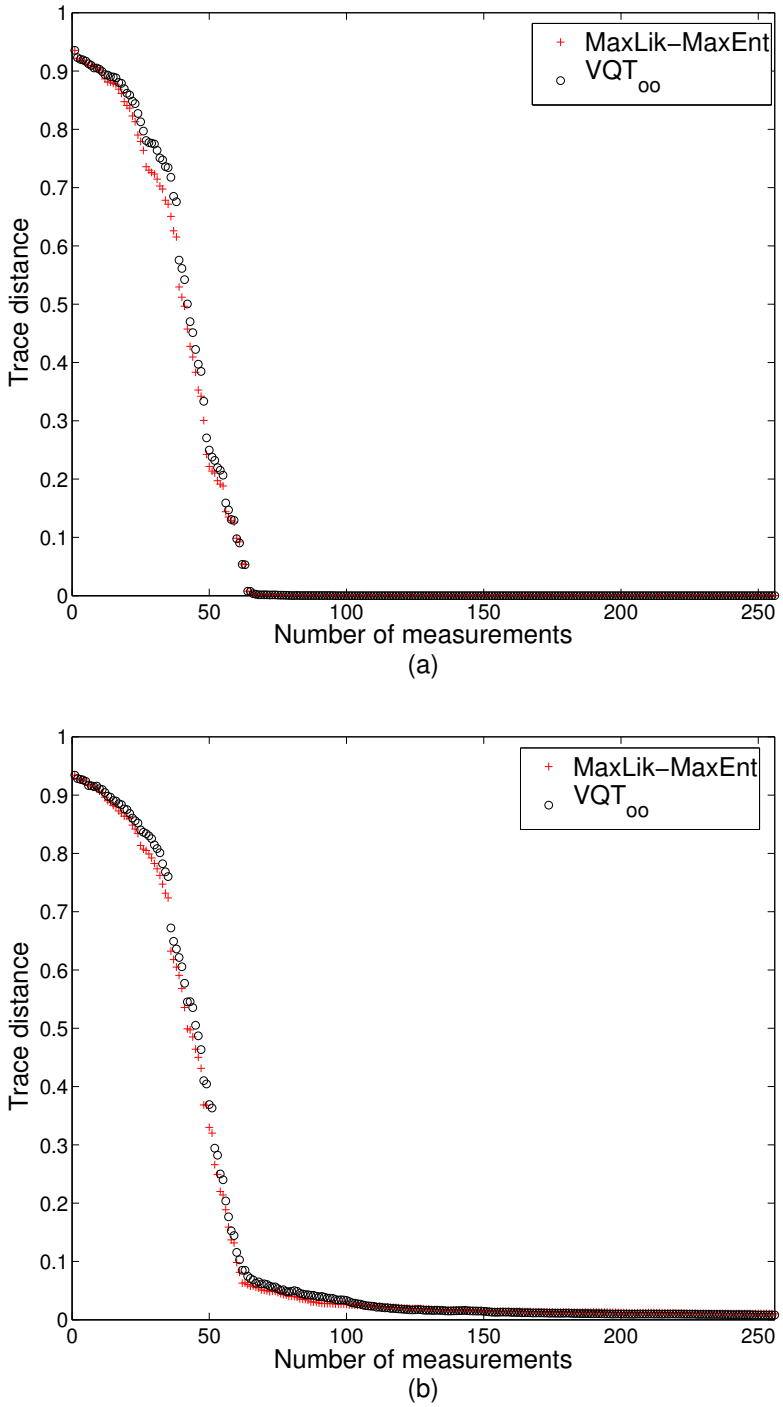

Figure 4.5: (a): Trace distance to the reference rank-1 state with probabilities perturbed by a Gaussian noise. (b): Trace distance to the reference state with 5% uniform noise.

CHAPTER 5

Estimation of quantum processes in informationally incomplete contexts

After dealing with inferences of quantum states in Chapter 4, we will proceed to a more complex scenario: characterization of quantum dynamics.

5.1 Linear mappings and quantum channels

We describe these dynamics with linear mappings of the form $\Phi : \mathcal{L}(\mathcal{H}^I) \to \mathcal{L}(\mathcal{H}^O)$, where $\mathcal{H}^I$ and $\mathcal{H}^O$ are the Hilbert spaces of input and output respectively and, here, $dim(\mathcal{H}^I) = dim(\mathcal{H}^O) = d$. This abstract representation can be useful in some situations, but here we prefer to work with more operational forms of quantum channels (see Watrous' lecture notes in [72] for an extesive discussion on quantum channels and its representations).

Suppose we have finite and non-empty sets of operators

$$\{A_a : a \in \Gamma\}, \{B_a : a \in \Gamma\} \subset \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$$

for which the equation

$$\Phi(\rho) = \sum_{a \in \Gamma} A_a \, \rho B_a^\dagger \tag{5.1}$$

holds for all $\rho \in \mathcal{L}(\mathcal{H}^I)$. We call the expression in Eq. (5.1) the *Kraus representation* [36] of the map Φ and it is never unique . We usually reserve this term for the case where $A_a = B_a$ for each $a \in \Gamma$. In this case, we call the operators A_a *Kraus operators*, and they exist if, and only if, Φ is completely positive [36], *i.e.*, $\Phi \otimes \mathbb{1}_{\mathcal{L}(\mathcal{H}^K)}$ is positive for all possible extensions in $\mathcal{L}(\mathcal{H}^K)$ and $dim(\mathcal{H}^K) \geq 1$.

To show an example of the non-uniqueness of the Kraus representation, we can expand A_a and B_a using an IC-POVM (Def. 3.6),

$$\begin{aligned}\Phi(\rho) &= \sum_{a\in\Gamma} A_a\,\rho B_a^\dagger = \sum_{a\in\Gamma}\left(\sum_{i=1}^{d^2} c_i^a E_i\right)\rho\left(\sum_{j=1}^{d^2} d_j^a E_j\right)^\dagger \\ &= \sum_{i,j=1}^{d^2}\chi_{ij}E_i\,\rho E_j^\dagger,\end{aligned} \tag{5.2}$$

where $\chi_{ij} = \sum_{a\in\Gamma} c_i^a \overline{d_j^a}$ and the bar above means the complex conjugate. The set of coefficients $\{\chi_{ij}\}$ defines the super-operator $\chi \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ [54] , which has all the information about the process. It is a $d^2 \times d^2$ positive semi-definite operator, and we can think of it as a density matrix in the Hilbert-Schmidt space with d^4 independent real parameters (or $d^4 - d^2$ in the trace preserving case).

From A_a and B_a (or χ and E), we can also obtain the *Choi-Jamiołkowski representation* [11] of Φ,

$$J(\Phi) = \sum_{a\in\Gamma} vec(A_a)vec(B_a)^\dagger = \sum_{i,j=1}^{d^2}\chi_{ij}vec(E_i)vec(E_j)^\dagger, \tag{5.3}$$

where $J(\Phi) \in \mathcal{L}(\mathcal{H}^O \otimes \mathcal{H}^I)$, and the operation $vec(\cdot)$ transforms ρ into a column vector (*cf.* Section 3.4). To recover the action of Φ using this representation, we use the following relation:

$$\Phi(\rho) = \mathrm{tr}_{\mathcal{H}^I}\Big(J(\Phi)(\mathbb{1}_{\mathcal{H}^O}\otimes\rho^T)\Big). \tag{5.4}$$

Differently from Kraus representation, $J(\Phi)$ is unique, we can write it into its spectral decomposition,

$$J(\Phi) = \sum_{i=1}^{r}\lambda_i vec(\tilde{K}_i)vec(\tilde{K}_i)^\dagger, \tag{5.5}$$

and bring it back to Kraus form

$$\Phi(\rho) = \sum_{i=1}^{r} K_i\,\rho K_i^\dagger, \tag{5.6}$$

with $vec(K_i) = \sqrt{\lambda_i}vec(\tilde{K}_i)$ and $r = rank(J(\Phi))$ is the *Kraus rank* of Φ, *i.e.*, the minimum number of Kraus operators needed to represent Φ. This number will play the same role as the rank of inferred states in Chapter 4.

We can also obtain interesting properites of Φ using $J(\Phi)$:

(i): Φ is completely positive if, and only if, $J(\Phi)$ is a positive semi-definite operator;

(ii): Φ is trace-preserving ($\mathrm{tr}(\Phi(\rho)) = \mathrm{tr}(\rho)$) when

$$\mathrm{tr}_{\mathcal{H}^O}(J(\Phi)) = \mathbb{1}_{\mathcal{H}^I}; \tag{5.7}$$

(iii): Φ is unital ($\Phi(\mathbb{1}) = \mathbb{1}$) when

$$\mathrm{tr}_{\mathcal{H}^I}(J(\Phi)) = \mathbb{1}_{\mathcal{H}^O}. \tag{5.8}$$

See reference [72] for a complete treatise on quantum maps. We can also define what we mean by a quantum channel, namely

Definition 5.1 (Quantum Channel)**.** *We call* Φ *a* quantum channel *when it is both completely positive and trace-preserving.*

5.2 Variational Quantum Process Tomography (VQPT)

The first question which arises when Φ is unknown is what do we need to fully characterize the dynamics. To answer this question, we need to choose a *quantum process tomography* (QPT) scheme. Each procedure demands different resources and operations.

There are four general types of QPT procedures: (i) *standard quantum process tomography* (SQPT) [12]; (ii) *ancilla-assisted process tomography* (AAPT) [14, 42, 2]; (iii) *direct characterization of quantum dynamics* (DCQD) [50, 51, 52]; (iv) *selective and efficient quantum process tomography* (SEQPT) [4, 5, 65].

In (i), we obtain the information indirectly, performing a set of *quantum state tomographies* (QSTs) [15, 16, 60, 47, 45, 43] of linear independent states, which spans the Hilbert-Schmidt space of interest, after the action of the unknown map. The second scheme (ii) – also an indirect procedure – makes use of an auxiliary system. The information is then extracted by means of QST of the joint space (system and ancilla). The third one (iii) obtains the dynamical information directly – using *quantum error detection* (QED) [54] concepts – measuring stabilizers and normalizers. Finally, the last method (iv) – which also measures the parameters directly – consists in estimating averages over the entire Hilbert space of products of expectation values of two operators. For the special case of one-parameter quantum channels, there is also an interesting method developed by Sarovar and Milburn [64].

We showed we can infer quantum states propely out of incomplete and noisy information with VQT and VQT$_\infty$ in Chapter 4. Here, we extend VQT to the tomography of quantum processes [46]. The new method inherits all the advantages of the VQT and opens the door to the characterization of maps in larger systems, where lots of measurements could be necessary. We will present our method recasting both SQPT(i) and AAPT(ii) – which rely on tomography of states – in the paradigm of incomplete information, as we did in Chapter 4 with quantum states.

Suppose we prepare ρ^k, with $k = 1, \ldots, m$, input states. After the action of the map, we can write the output states as

$$\tilde{\varrho}^k \equiv \Phi(\rho^k) = \sum_{i,j=1}^{d^2} \chi_{ij} E_i \rho^k E_j^\dagger. \tag{5.9}$$

Suppose also that we measured n^k, with $n^k \leq d^2$, elements of an IC-POVM (Def. 3.6) in the k^{th} output state, which reads

$$\text{tr}\left(E_\lambda \tilde{\varrho}^k\right) = f_\lambda, \quad \lambda(i,k) \in \mathcal{I}, \tag{5.10}$$

where $\mathcal{I}$ is the set of indexes of measured data (same as in Chapter 4) and $\lambda(i,k)$ is a fused index of i and k (we will write just λ from now on to avoid cumbersome notation). Note we also assumed the frequentist decision rule in Eq. (3.5) to map our measurements onto f_λ.

The SQPT method demands quantum state tomography in all linearly independent states which span the Hilbert-Schmidt space. With n^k measurements in $k = 1, \ldots, m$ different states, the variational SQPT reads:

$$\begin{aligned}
\min_{\chi,\, \vec{\Delta}} \quad & \sum_{\lambda\in\mathcal{I}} \Delta_\lambda + \sum_{\lambda\notin\mathcal{I}} \text{tr}\left(E_\lambda \tilde{\varrho}^k\right) \\
\text{subject to} \quad & \chi \succeq 0\,; \\
& \text{tr}\left(\tilde{\varrho}^k\right) \leq 1\,; \\
& |\text{tr}\left(\tilde{\varrho}^k E_\lambda\right) - f_\lambda| \leq \Delta_\lambda f_\lambda\,, \quad \lambda \in \mathcal{I}; \\
& 0 \leq \Delta_\lambda \leq 1\,; \\
& k = 1, \ldots, m\,,
\end{aligned} \tag{5.11}$$

whit $\tilde{\varrho}^k$ defined in Eq. (5.9). We can also write Eq. (5.11) in terms of our VQT$_\infty$ inference scheme as

$$\begin{aligned}
\min_{\chi,\, \vec{\Delta},\, \delta} \quad & \delta + \sum_{\lambda\in\mathcal{I}} \Delta_\lambda \\
\text{subject to} \quad & \chi \succeq 0\,; \\
& \text{tr}\left(\tilde{\varrho}^k\right) \leq 1\,; \\
& |\text{tr}\left(\tilde{\varrho}^k E_\lambda\right) - f_\lambda| \leq \Delta_\lambda f_\lambda\,, \quad \lambda \in \mathcal{I}; \\
& |\text{tr}\left(\tilde{\varrho}^k E_\lambda\right)| \leq \delta\,, \qquad\qquad \lambda \notin \mathcal{I}; \\
& 0 \leq \Delta_\lambda \leq 1\,; \\
& \delta \geq 0\,; \\
& k = 1, \ldots, m\,.
\end{aligned} \tag{5.12}$$

The SDPs in Eqs. (5.11) and (5.12) return the operator χ, which we can use to infer the unknown process Φ using the representations in Eqs. (5.2) and (5.3). Note that, at the same time, we were able to infer $\tilde{\varrho}^k$ optimally.

We can analyze the convergence of these methods in function of the Kraus rank of Φ (similarly to the analysis in Section 4.5). In Fig. 5.1, we simulated the reconstruction of two-qubit processes with $d^4 = 256$ linearly independent parameters using the scheme in Eq. (5.11). For each rank, we randomly generated[1] 100 Choi matrices $J(\Phi)$; fixed a SIC-POVM basis [61] (see Def. 3.7); and chose the sequence of measured E_λ at random.

[1] According to the Haar measure [75] when we have a pure map and according to Hilbert-Schmidt measure [75] otherwise.

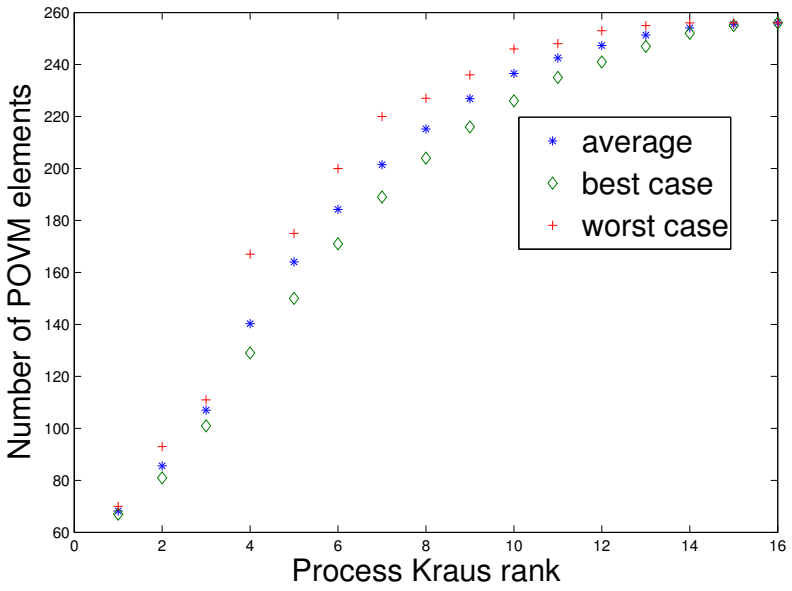

Figure 5.1: Reconstruction of two-qubit processes of all ranks. We plot the number of linearly independent POVM elements necessary to reconstruct the map against the process rank. For each rank, we randomly generated 100 processes.

We can also recast the AAPT method. In this case, we need to add an ancillary system with the same dimension of the first one. Then the quantum process takes place in half subspace, and we perform a quantum state tomography in the whole space, ancilla plus main system.

The output state now reads

$$\tilde{\varrho} = (\Phi \otimes \mathbb{1})(\rho) = \sum_{i,j=1}^{d^2} \chi_{ij} (E_i \otimes \mathbb{1}) \rho (E_j \otimes \mathbb{1})^\dagger . \tag{5.13}$$

With $n \leq d^4$ measurements performed in this scheme, AAPT can be recast as

$$\begin{aligned} \min_{\chi,\, \vec{\Delta}} \quad & \sum_{i \in \mathcal{I}} \Delta_i + \sum_{i \notin \mathcal{I}} \mathrm{tr}(E_i \tilde{\varrho}) \\ \text{subject to} \quad & \chi \succeq 0 \,; \\ & \mathrm{tr}(\tilde{\varrho}) \leq 1 \,; \\ & |\mathrm{tr}(\tilde{\varrho} E_i) - f_i| \leq \Delta_i f_i \,, \quad i \in \mathcal{I}; \\ & 0 \leq \Delta_i \leq 1 \,, \end{aligned} \tag{5.14}$$

now with $\tilde{\varrho}$ defined by Eq. (5.13). In terms of our VQT$_\infty$ inference scheme, we

have

$$\begin{aligned} \min_{\chi,\,\vec{\Delta}} \quad & \delta + \sum_{i\in\mathcal{I}} \Delta_i \\ \text{subject to} \quad & \chi \succeq 0\,; \\ & \mathrm{tr}(\tilde{\varrho}) \leq 1\,; \\ & |\mathrm{tr}(\tilde{\varrho}E_i) - f_i| \leq \Delta_i f_i\,, \quad i \in \mathcal{I}; \\ & |\mathrm{tr}(\tilde{\varrho}E_i)| \leq \delta\,, \qquad\quad i \notin \mathcal{I}; \\ & 0 \leq \Delta_i \leq 1\,; \\ & \delta \geq 0\,. \end{aligned} \tag{5.15}$$

5.3 Agent's design

Comparing with Eqs. (4.1) and (4.27), we see that the programs in Eqs. (5.11), (5.12), (5.14) and (5.15) are roughly quantum state tomographies in larger spaces.

So, to design a rational agent to infer quantum maps, we can refer to our previous design in Section 4.6, while choosing how to model the output states: using Eq. (5.9) in case of SQPT procedure or Eq. (5.13) in case of AAPT.

CHAPTER 6

Quantum process tomography with informational incomplete data of two J-coupled heterogeneous spins relaxation in a time window much greater than T_1

In this chapter, we will present an application of quantum process tomography (*cf.* Chapte 5) as written in [48]. Throughout this chapter, it will be clear the paramount effort needed to properly infer quantum states and processes in complex scenarios.

6.1 Introduction

During the last few years, we have seen proposals and experimental implementations of quantum information processors for various quantum systems [38]. While there is still no ultimate physical platform, nor an underlying information processing model, most of the known candidates must deal with adverse quantum effects. Nuclear Magnetic Resonance (NMR) techniques have been successfully used to implement quantum protocols, *e.g.*, quantum teleportation [55] and Shor's [70] algorithm. NMR remains indispensable in quantum computing as it continues to provide new insights, new methods and

new techniques, as well as allowing for testing interesting quantum computing tasks, although there are still huge difficulties in initialization, error-correction and scalability [54].

In the NMR Ensemble Quantum Computing (EQC) introduced by Cory *et al.* [13], a key observation is that, compared to the ideal case where we have one hundred percent of polarized molecules, the proportion of molecules effectively contributing to the magnetization in a pseudo-pure state is about one part in a million, nonetheless one still has a macroscopic detectable signal. On the other hand, when performing an algorithm with EQC, one has to deal with two intrinsic limitations of the liquid state NMR, namely: (i) the natural relaxation process the spins are subjected to is non-unital, which means that the state simulated by a pseudo-pure state has a non-linear time evolution, what can compromise the quantum computation. (ii) Related to this fact, there is the recovery of the magnetization towards equilibrium, characterized by the spin-lattice relaxation time T_1, which diminishes the amplitude of the component corresponding to the simulated state along the time. This means that expectation values measured in different rounds of the experiment will not be comparable, unless one takes into account the reduction of their magnitudes along the time. When the quantum computation lasts about the transverse relaxation time T_2, the typical coherence time, these drawbacks are negligible, and one can ignore both the non-linear evolution and the degradation of the simulated state, and also some particular expectation value can be taken as a reference in order to normalize the other measurements in the computation. Notwithstanding, the spins natural relaxation is an ever present process (Φ_{relax}) superimposed to the map of the particular computation one is performing. In the present work, our goal is to characterize the relaxation process in the same way one would reconstruct the map corresponding to a quantum computation, *i.e.*, by means of quantum process tomography with the least possible assumptions. We will be able to quantify, along the time, how much the relaxation process departs from being unital and trace preserving, and we will also characterize how the relaxation process affects the normalization of expectation values. This way we will learn in which time window it is safe to assume the relaxation process is approximated unital or trace preserving, complementing other previous studies [6, 23, 20]. Thus we hope to shed some light in how one could compensate for the two drawbacks aforementioned, and consequently increase the time window for quantum simulations in liquid state NMR. The present work also serves the purpose to test, in a real NMR experiment, and in a non trivial situation, the efficiency, robustness, flexibility and simplicity of implementation of our quantum tomography techniques [45, 46, 22].

Robust techniques for quantum process tomography in NMR EQC have been developed in the case the process is assumed to be Markovian and trace preserving [6] , or when some model is assumed for the Hamiltonian governing the process [23]. In both cases, the process tomography was performed in a time window of about T_2, and therefore the process could be safely assumed to be unital. We, on the contrary, will assume the worst case scenario, meaning the process may be non-unital, non-Markovian, and not trace preserving. Besides that, we want to follow the quantum process for very long times, much more

than the spin-lattice relaxation time T_1, which raises the problem of how to compare measurements taken at different times, which is what we call the *normalization problem*. A pedagogical discussion of why the state simulated by a pseudo-pure state evolves non-linearly under the relaxation process, and an *ad hoc* phenomenological method to circumvent the *normalization problem* were presented by Gavini-Viana *et al.* [20]. The basic idea is that a pseudo-pure state, according to Cory *et al.* [13], has the form

$$\rho_{pp} = [(1-\alpha)\mathbb{1}_{2^n} + 2\alpha|\psi\rangle\langle\psi|]/[(1-\alpha)2^n + 2\alpha], \tag{6.1}$$

where n is the number of spin-$\frac{1}{2}$ particles and $\mathbb{1}_{2^n}$ is the $2^n \times 2^n$ identity matrix[1]. In a quantum computing simulation, we are concerned with unitary evolutions, which of course affect only $|\psi\rangle\langle\psi|$, the simulated state. However, if Φ is a non-unital map, $\Phi(\mathbb{1}_{2^n}) \neq \mathbb{1}_{2^n}$, rewriting $\Phi(\rho_{pp})$ as a pseudo-pure state is equivalent to a non-linear evolution of $|\psi\rangle\langle\psi|$. Note also that the new pseudo-pure state $\Phi(\rho_{pp})$, differently from the case of unital evolutions, has a new "α". Therefore, the *normalization problem* is equivalent to account for the time evolution of "α" under non-unital processes, such that we can always track the evolution of the resulting pseudo-pure state.

To avoid confusion, it is important to remember that a quantum process is a complete positive map, that can be expressed in the usual Kraus sum representation (see Section 5.1), namely:

$$\Phi(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{6.2}$$

where the K_i are the Kraus operators. A map is unital if and only if

$$\sum_i K_i K_i^\dagger = \mathbb{1}_{2^n}, \tag{6.3}$$

a condition that does not imply trace preservation. To be trace preserving, a map must satisfiy

$$\sum_i K_i^\dagger K_i = \mathbb{1}_{2^n}. \tag{6.4}$$

The state simulated by a pseudo-pure state evolves linearly under unital and non-trace preserving maps, and therefore these maps pose no particular challenge to usual process tomography techniques in liquid state NMR EQC. In particular, the quantum process tomography method we developed in [46] (see Chapter 5), and experimentally tested in Quantum Optics by Vianna *et al.* [71], could be employed without any modification to handle these maps, as well as non-unital ones. A caveat about our basic assumption of the complete positiveness (CP) of the map is that the tomographed map could eventually be non-CP due to noisy measurements [73], or in the case of correlations between the initial prepared state and the environment [62].

The choice of the objective linear function can relate our methods [46, 45, 22] (see Chapters 3 and 4) to other approaches. For example, by choosing

[1] One should remember that the pseudo-pure state is not an actual quantum state of the NMR system, but rather the result of *spatial* or *temporal labelling* [54].

the nuclear (trace) norm of the positive semidefinite matrix of interest as the objective function, we have a matrix completion approach equivalent to compressed sensing [24]. Note however, that nuclear norm minimization is also a good heuristic for quantum tomography in general, including the case of informationally complete measurements. Detailed comparison of our methods with maximum likelihood and maximum entropy was reported in [22] (see Chapter 4). Our methods were tested for state and process tomography in Quantum Optics experiments [43, 57, 71], and in liquid state NMR it was tested for state tomography [56].

6.2 The NMR system

The NMR experiment was performed on a liquid-state enriched carbon-13 Chloroform sample, at room temperature, in a Varian 500 MHz shielded spectrometer. This sample exhibits two qubits encoded in the ^{1}H and ^{13}C spin-$\frac{1}{2}$ nuclei. The heteronuclear interaction of the two spins is characterized by the magnitude of the J-coupling being much smaller than the resonant frequencies of the two nuclei, 500 MHz and 125 MHz, for ^{1}H and ^{13}C, respectively. This system is useful for the simulation of quantum information processing in a Hilbert space of two qubits.

The quantum model adopted in room temperature liquid-state NMR experiments is of an ensemble of N ($\mathcal{O}(10^{15})$) non-interacting independent molecules. Therefore the Hamiltonian can be written as

$$\begin{aligned} H_{\{N\}} &= H \otimes \mathbb{1}^{\otimes N-1} + \mathbb{1} \otimes H \otimes \mathbb{1}^{\otimes N-2} + \dots \\ &\quad \dots + \mathbb{1}^{\otimes N-2} \otimes H \otimes \mathbb{1} + \mathbb{1}^{\otimes N-1} \otimes H \\ H_{\{N\}} &= \textstyle\sum_{i=1}^{N} \bigotimes_{j=1}^{N} \big(\delta_{ij} H + (1-\delta_{ij})\mathbb{1}\big), \end{aligned} \tag{6.5}$$

where $\mathbb{1}$ is the 2×2 identity matrix, δ_{ij} is the Kronecker's delta, and H is an *effective* Hamiltonian for a single molecule. The wave function and the density matrix corresponding to the N-particle Hamiltonian are assumed to be separable, namely:

$$\Psi_{\{N\}} = \bigotimes_{i=1}^{N} \psi_i \quad \text{and} \quad \rho_{\{N\}} = \bigotimes_{i=1}^{N} \rho_i, \tag{6.6}$$

where ψ_i and ρ_i are the wave function and density matrix for a single molecule, respectively, where we shall work only in the spin degrees of freedom.

In an NMR experiment, one does not have access to the particular state of a distinct molecule in the preparation procedure, all one has is the average state of a representative molecule of the ensemble. Therefore, we can write:

$$\rho_{\{N\}} = \bigotimes_{i=1}^{N} \rho_i = \rho^{\otimes N}, \tag{6.7}$$

where ρ stands for this *representative state*, which we shall discuss further later.

Our observables are related to the resultant macroscopic magnetization induced by an external static magnetic field, and we assume they are operators

of the form:

$$\begin{aligned} A_{\{N\}} &= A \otimes \mathbb{1}^{\otimes N-1} + \mathbb{1} \otimes A \otimes \mathbb{1}^{\otimes N-2} + \ldots \\ &\quad \cdots + \mathbb{1}^{\otimes N-2} \otimes A \otimes \mathbb{1} + \mathbb{1}^{\otimes N-1} \otimes A, \\ A_{\{N\}} &= \textstyle\sum_{i=1}^{N} \bigotimes_{j=1}^{N} (\delta_{ij} A + (1-\delta_{ij})\mathbb{1}). \end{aligned} \tag{6.8}$$

Therefore, we have expectation values of the kind:

$$\begin{aligned} \langle A_{\{N\}} \rangle &= \textstyle\sum_{i=1}^{N} \mathrm{tr}\Big(\rho^{\otimes N} \Big[\bigotimes_{j=1}^{N} (\delta_{ij} A_i + (1-\delta_{ij})\mathbb{1})\Big]\Big) \\ &= N \,\mathrm{tr}(\rho A). \end{aligned} \tag{6.9}$$

This rationale, assuming Eqs. (6.5-6.8) as premisses, formalizes a model in which one has equally prepared, but independent, black boxes and the same measurement "buttons" are pressed in all of them when the measurement occurs. Thus, we have a *reducible* to a single box model and we can generalize and specialize our arguments back and forth, *i.e.*, what happens in a single box, will happen in the other ones and *vice versa*. We shall use this freely. This model fits the common knowledge that measurements in NMR have contributions of all the molecules in the ensemble [54, 13, 56], justifying the detected macroscopic signals.

The observables of interest in the NMR experiment are generally the components of the macroscopic nuclear magnetization, which are proportional to the ensemble average values of the nuclear spin operators. The matrix representation of these operators in the $|S,m\rangle$ basis (with $S = 1/2$) reads $S_x = \sigma_x/2$, $S_y = \sigma_y/2$, $S_z = \sigma_z/2$, with the usual Pauli matrices given by:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}; \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

For example, one can calculate the magnetization in the z-direction of an ensemble of nuclear spins using the following relation:

$$\langle M_z \rangle \equiv \hat{I}_z \propto \mathrm{tr}(\rho \sigma_z), \tag{6.10}$$

where $\hat{I}_z$ is the empirical value associated with the observation of the magnetization in the z-direction and ρ is the density matrix of the system of a single specimen. Remember that the left-hand side of this equation represents the statistical average over the entire ensemble of N molecules, and not the expectation value for a given particular member of the ensemble.

The relevant effective two-spin Hamiltonian in Eq. (6.5) is given by

$$H = -\hbar\omega_H S_z \otimes \mathbb{1} - \hbar\omega_C \mathbb{1} \otimes S_z + 2\pi\hbar J S_z \otimes S_z, \tag{6.11}$$

where the labels H and C stand for hydrogen and carbon, respectively, as the first and second particles. The resonant frequencies ω_H and ω_C include the effects of isotropic chemical shift σ_{iso} in acetone solution for each nucleus, *i.e.*, $\omega = \omega_L(1 - \sigma_{iso})$. The Larmor frequency is given by $\omega_L \equiv \gamma_n B_0$, where γ_n

is the gyromagnetic ratio of the nucleus, and B_0 is the static magnetic field in z-direction.

At thermal equilibrium, the density matrix obtained from the Hamiltonian in Eq. (6.11) is given by the maximum entropy principle [32]:

$$\rho_{eq} = \frac{e^{-H/k_BT}}{\mathrm{tr}\left(e^{-H/k_BT}\right)}, \tag{6.12}$$

where k_B is the Boltzmann constant and T is the temperature. In the high temperature limit, one can expand this exponential as

$$\rho_{eq} \approx \frac{1}{4}\mathbb{1}\otimes\mathbb{1} + \frac{1}{4}\frac{\hbar\omega_H}{k_BT}S_z\otimes\mathbb{1} + \frac{1}{4}\frac{\hbar\omega_C}{k_BT}\mathbb{1}\otimes S_z + \frac{1}{2}\frac{\hbar\pi J}{k_BT}S_z\otimes S_z. \tag{6.13}$$

Note that the last term of Eq. (6.13) is usually neglected (see for example [54, 56]), for it is actually very small compared to the other terms. Notwithstanding, as we will show, this term contains crucial information about the time window for quantum computing simulations.

Now, we are able to calculate the total magnetization of the system in the z-direction, using the equilibrium state in Eq. (6.13), the relation in Eq. (6.10) and Curie's law, which leads to

$$M_z^{eq} = \frac{n_H\gamma_H^2\hbar^2B_0}{4k_BT} + \frac{n_C\gamma_C^2\hbar^2B_0}{4k_BT} + \frac{n_J\pi^2\iota_J^2\hbar^2B_0}{k_BT}, \tag{6.14}$$

where n_H and n_C are the number of hydrogen and carbon nuclei per unit volume, respectively, which are equal and correspond to the number of chloroform molecules per unit volume ($n_H = n_C = n$); n_J stands for the number of *responsive* pairs of J-coupled nuclei per unit volume, and ι_J its "gyromagnetic ratio". These responsive pairs of spins correspond to a small fraction of chloroform molecules ($N_J/N \ll 1$), for which the spins of ^{1}H and ^{13}C are in a chosen coherent state prepared by a certain sequence of radiofrequency pulses (RF). The first issue on normalizing the experimental data arises due to the fact that neither N_J nor its density n_j, are straightforwardly revealed by the magnetization measurements. We shall discuss this later.

Let us focus on the experiment now. We know that interesting NMR quantum information processing experiments happen in non-equilibrium, where the responsive J-coupled pairs of spins play an important role. In our experiment, we prepare the system in a desired non-equilibrium magnetization state at time $t = 0$ and allow it to relax. First, at $t = 0$, a certain sequence of RF pulses prepares the system in some desired non-equilibrium state, where we have the maximum number of responsive J-coupled spin pairs. This results in a small macroscopic non-equilibrium magnetization. The second part of the experiment happens in the time window which we can use the small fraction of responsive molecules in the non-equilibrium state to simulate the dynamics of quantum parameters (*e.g.*, [70]). The system is allowed to relax, the molecules that responded to the preparation start to recover their equilibrium state, diminishing N_J, and finally ($t \to \infty$) the system reaches the equilibrium with vanishing N_J.

The *relaxation* is such that, after some time has elapsed, the magnetization returns almost completely to the initial z-direction, satisfying again the thermal equilibrium requirements. It is worth emphasising that the exact meaning of "some time" is largely dependent on the details of each particular nuclear spin system and its environment, ranging typically from microseconds to several hours. The relaxation consists of two different processes, occurring simultaneously but, in general, independently, namely: the *transverse relaxation* and the *longitudinal relaxation* [56]. We shall characterize, by the quantum process tomography, a net effect of these two phenomena, that is paramount to the problem of NMR simulation of quantum information processing. For reasons to become clear soon, we shall refer to this net phenomenon as the *J-coupling relaxation*.

The longitudinal relaxation is the process that leads the longitudinal component of the nuclear spin magnetization to recover its equilibrium value. The recovery of the M_z component of the magnetization is related to transitions between the nuclear spin levels. The natural tendency is the system to give up its excess of energy by effecting transitions from the upper to the lower energy level. After some time, which is commonly named T_1, the Boltzmann distribution is almost reestablished.

Simultaneously to the longitudinal relaxation, the transverse relaxation is the process that leads to the disappearance of the components of the nuclear magnetization $\mathbf{M}$ that are perpendicular to the static field $\mathbf{B}_0$. The origin of the transverse relaxation relies on the loss of coherence in the precession motion of the spins (or dephasing of the spins), caused by the existence of spread in precession frequencies for the collection of nuclear spins. This spread progressively results in a reduction of the resultant transverse components (*e.g.*, M_x and M_y). After some time, which is commonly named T_2, the spins distribute randomly in a precession cone around $\mathbf{B}_0$ and the transverse magnetization tends to zero.

The J-coupling relaxation, which we introduce here, can be considered as the net effect of both longitudinal and transverse relaxations. In NMR quantum information processing simulation experiments, correlated spins in non-equilibrium states are indispensable to *simulate* quantum algorithms. We name T_1^J and T_2^J the characteristic times in which the coupled magnetizations M_{ij} ($\forall\, i,j \in \{x,y,z\}$) decay, T_1^J for the longitudinal, or M_{zz} component, and T_2^J for the transverse components. Note that what we call *coupled-magnetization* refers specifically to expectation values of two-spin operators $\left(\mathrm{tr}\left(\rho\sigma_i^H \otimes \sigma_j^C\right)\right)$, while *uncoupled-magnetization* is related to expectation values of one-spin operators $\left(\mathrm{tr}\left(\rho\sigma_i^H \otimes \mathbb{1}\right)\right.$ or $\left.\mathrm{tr}\left(\rho\mathbb{1} \otimes \sigma_j^C\right)\right)$. Our motivation for introducing T_1^J and T_2^J was the following. (*i*) These parameters can be extracted from the usual NMR data; (*ii*) the M_{zz} component behaves distinctly from all the other coupled magnetization components; (*iii*) and finally, these parameters characterize an important change in the behavior of the relaxation process map, as we will show.

In order to reconstruct the density matrix of the system along the time, we need to normalize the experimental data, such that we can relate the measured

non-equilibrium magnetizations to expectation values. Here, we are not concerned with the simulated states, we deal only with the real non-equilibrium NMR states, although it is worth to emphasise that, in the preparation procedure, we also assume *ergodicity*, as we collect the whole set of measurements using *temporal labeling* [54, 56, 41].

6.3 The normalization problem

The normalization issue is a daunting problem in NMR experiments. Firstly, by the data acquisition itself, since it involves taking the intensity of the peaks from a Fourier transform spectra of an oscillating signal. Secondly, because we are forced to assume the representative state ρ in Eq. (6.7). It would be preferable if we could address each molecule individually, such as

$$\rho_{\{N\}}(t) = \rho_1(t) \otimes \rho_2(t) \ldots \otimes \rho_N(t).$$

If it were the case, we would be able to count how many spins are still coherent and responding accordingly, therefore it would be possible to know exactly the number of responding molecules as in Eq. (6.9) (*e.g.*, if we were interested in knowing N_J). However such a case is not possible and we shall stick to the weaker premise of the representative state, losing track of this counting. Thirdly, as consequence of the traceless nature of the operators corresponding to the magnetization components, we do not possess any direct experimentally measured data related to the trace of the non-equilibrium density matrix. If we were measuring the eigenprojectors (P_k) of these operators ($A = \sum_{k=1}^{d} \lambda_k P_k$), this problem would not appear, as the completeness relation ($\mathbb{1}_d = \sum_k P_k$) would give us a trivial recipe for normalization, *i.e.*, we would impose $\sum_{k=1}^{d} \mathrm{tr}(\rho P_K) = 1$. However, what we have is $\langle A \rangle = \mathrm{tr}(\rho A)$, and for this measurement to be useful, we need some reference expectation value $\left\langle A_{ref} \right\rangle$. Of course, the reference expectation value cannot vary in time. For example, in the EQC NMR experiment reported in [56], which employed the same experimental setup and molecule of the present work, the reference expectation value was $\langle \sigma_Z^H \otimes \sigma_Z^C \rangle$ in a pseudo-pure state simulating the two-qubit state $|\uparrow\uparrow\rangle$. It was imposed then that $\langle \sigma_Z^H \otimes \sigma_Z^C \rangle / N_{ref} = 1$, and all other expectation values were divided by N_{ref}. As we shall see, the quantum tomography techniques we employ here [45, 46, 22], allow us to circumvent this problem, yielding proper inferences of the normalized non-equilibrium quantum states and quantum maps.

We have three different kinds of signals corresponding to our measurements, which we have to deal with in order to normalize the experimental data. One may grasp this in Fig. 6.1. We have, responding to RF pulses in the resonant frequency ω_H of the hydrogen, both coherently coupled and uncoherently coupled hydrogen nuclei; the same happens to carbon, for RF pulses in the resonant frequency ω_C. As the gyromagnetic ratio of the hydrogen is almost 4 times larger than that of the carbon, the hydrogen species is expected to have a larger magnetization in equilibrium. The last kind of signal, which is very small compared to the first two, corresponds to that coherently coupled magnetization of the N_j chloroform molecules. To remedy the lack of some relation

tying the intensity of measured magnetization components, like the completeness relation we have mentioned in the previous paragraph, we normalize all measured signals [41] corresponding to non-equilibrium magnetizations backwards, such that the expected value of the magnetization in equilibrium is equal to one. Therefore, the equilibrium magnetization plays the role of our reference expectation value that does not vary in time, as discussed in the previous paragraph. This rescaling of the magnetization signals is also convenient in order to perform our numerical optimizations with good precision. For the hydrogen and carbon signals, we have

$$\hat{I}_i^H(t) = \frac{1}{M}\frac{\hbar\omega_H}{4k_bT}\frac{\hat{I}_i^{H_{exp}}(t)}{\hat{I}_z^{H_{eq}}} \text{ and } \hat{I}_j^C(t) = \frac{1}{M}\frac{\hbar\omega_C}{4k_bT}\frac{\hat{I}_j^{C_{exp}}(t)}{\hat{I}_z^{C_{eq}}}, \tag{6.15}$$

and for the coupled signals we have

$$\hat{I}_{ij}^J(t) = \frac{1}{M}\frac{\hbar\pi J}{2k_bT}\frac{\hat{I}_{ij}^{J_{exp}}(t)}{\hat{I}_{zz}^{J_{eq}}}, \tag{6.16}$$

where

$$M = \frac{\hbar(\omega_H + \omega_C + 2\pi J)}{4k_bT},$$

and $\hat{I}_{i,j}^{\{H,C,J\}_{exp}}(t)$ are the empirical values associated with the observation of the magnetization in the $i,j \in \{x,y,z\}$ directions relative to the hydrogen, carbon and J-coupled signals respectively. The time evolution of these three kinds of signals is presented schematically in Fig. 6.1. In the figure one can see that at $t = 0$, the hydrogen and carbon have the same magnetization, as a consequence of the state preparation; at $t = T_2^C$, the coherence has almost vanished; at $t = T_1^C$, the system has a magnetization reminiscent of the equilibrium, with the total hydrogen signal greater than that of the carbon; and at $t = 60s$, which we call t_{eq}, the system is practically thermalised, with the total signal equal to one, and the hydrogen signal almost 4 times larger than that of the carbon.

The equilibrium (Eq. (6.13)) and non equilibrium states dealt with in the NMR experiment are of the kind

$$\rho \approx \frac{1}{4}\mathbb{1}\otimes\mathbb{1} + \epsilon\Delta, \tag{6.17}$$

where $\epsilon \approx 10^{-5}$ and Δ is a traceless matrix, known as the deviation matrix, corresponding to the last three terms of Eq. (6.13). Though the experimental magnetization signals come just from the deviation matrix, due to the traceless nature of the spin operators (for example: $\mathrm{tr}\big(\rho\sigma_i\otimes\sigma_j\big) = \epsilon\mathrm{tr}\big(\Delta\sigma_i\otimes\sigma_j\big)$), these are macroscopic signals, for there are still a huge number ($\mathcal{O}(\epsilon\times 10^{23}) = \mathcal{O}(10^{18})$) of molecules contributing to the magnetization.

As discussed in the Introduction, the time evolution of the state simulated by a pseudo-pure state (Eq. (6.1)) is non-linear, if the process is non-unital. It is easy to circumvent this non-linearity. We have simply to tomograph the process acting directly on the NMR state given by Eq. (6.17), which of course evolves linearly.

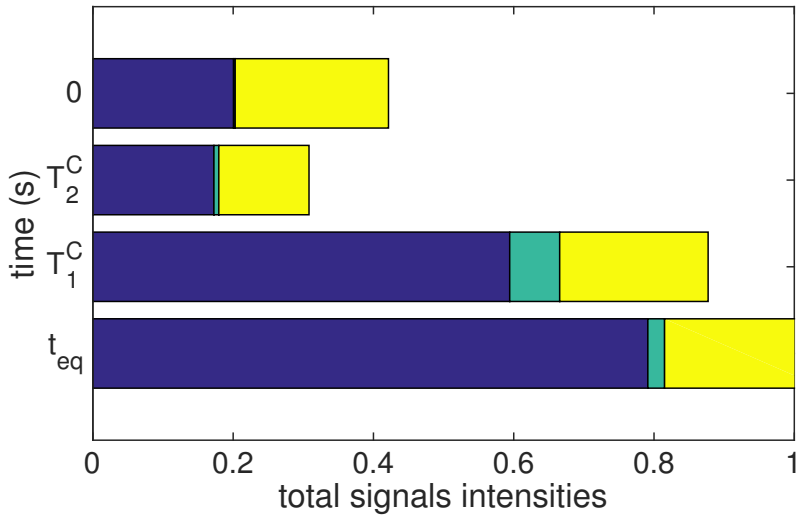

Figure 6.1: Comparison of the three distinct total signals intensities along time in the experiment. (Blue) Total hydrogen signals: $I^H = \sum_{i=x,y,z} |\hat{I}_i^H(t)|$ (*cf.* Eq. (6.15)); (Yellow) total carbon signals: $I^C = \sum_{i=x,y,z} |\hat{I}_i^C(t)|$ (*cf.* Eq. (6.15)); (Green) total coupled signals: $I^J = \sum_{i,j=x,y,z} |\hat{I}_{ij}^J(t)|$ (*cf.* Eq. (6.16)).

In summary, our strategy to keep measurements taken in different times properly normalized and comparable consists of two premisses: (*i*) quantum tomography accounts for the process acting on the pseudo-pure state along the time, rather than the process acting on the simulated state; (*ii*) the process at any finite time must be compatible with the magnetization converging to unity in an infinite time, *i.e.*, the equilibrium magnetization is taken as a reference to normalize all other measurements.

6.4 The quantum tomographies

The *standard quantum process tomography* (SQPT) [12, 54] consists in obtaining information about the map by means of quantum tomography of states modified by the action of the unknown process map. If the density matrices of the tomographed states have dimension d, then we need d^2 linear independent states to reconstruct the quantum map. As we are dealing with two qubits, d is 4. Our experiment consists in preparing 20 different time average pseudo-pure states [54] (ϱ_k^l, $l = 1,2,\ldots,20$), which can be used to simulate 20 different two-qubit pure states, where the density matrices of 16 of them are linear independent, spanning the Hilbert-Schmidt space. We also have 51 observations of these states in the time instants t_k $(k = 0,1,\ldots,50)$. The simulated 20 pure states correspond to 16 states forming the two-qubit Mutually Unbiased Basis (MUB)

described in [47], plus the four canonical states $|\uparrow\uparrow\rangle$, $|\uparrow\downarrow\rangle$, $|\downarrow\uparrow\rangle$, $|\downarrow\downarrow\rangle$.

The overcomplete set of 20 density matrices were preferred in order to increase the quality of the process tomography, following the arguments in [8, 39]. This strategy of enlarging the number of measurements with an overcomplete basis, in order to increase precision, was also used by Vianna *et al.* in the process tomography of a two-qubit map in Quantum Optics [71].

Since we have a large ensemble and sharp peaked measured signals $\hat{I}_{i,j}^{\{H,C,J\}_{exp}}(t)$, we can consider the latter gaussianly distributed with standard deviation $\sigma^2 = 1$. Thus, we can deal with the experimental errors with a simple ℓ_1-norm minimization, which is equivalent to a *log-likelihood* method assuming gaussian noise [7, 3].

The first thing to do is to perform a quantum state tomography in order to have a preliminary estimate of the initial states. We use a variation of [45, 22] also equipped with nuclear norm minimization [24] as follows:

$$\begin{aligned}
\min_{\rho_0^l,\,\Delta} \quad & \left\|\rho_0^l\right\|_* + \textstyle\sum_{l,i,j} \Delta_{i,j,ij}^{l,\{H,C,J\}} \\
\text{subject to} \quad & \rho_0^l \succeq 0 \\
& \left|\mathrm{tr}\left(\rho_0^l\, \sigma_i^H \otimes \mathbb{1}\right) - \hat{I}_i^{l,H}(0)\right| \leq \Delta_i^{l,H} \\
& \left|\mathrm{tr}\left(\rho_0^l\, \mathbb{1} \otimes \sigma_j^C\right) - \hat{I}_j^{l,C}(0)\right| \leq \Delta_j^{l,C} \\
& \left|\mathrm{tr}\left(\rho_0^l\, \sigma_i^H \otimes \sigma_j^C\right) - \hat{I}_{ij}^{l,J}(0)\right| \leq \Delta_{ij}^{l,J} \\
& \forall\, i,j \in \{x,y,z\} \\
& \Delta_i^{l,H}, \Delta_j^{l,C}, \Delta_{ij}^{l,J} \geq 0\,.
\end{aligned} \tag{6.18}$$

where $\|\cdot\|_*$ stands for the *nuclear norm* or *trace norm*. This type of *Semidefinite Problem* (SDP) in the second-order cone can be efficiently solved using well known algorithms [44, 69, 67].

The program in Eq. (6.18) has the following interpretation. We want a point estimate (ρ_0^l) of a density matrix (positive semidefinite constraint), with minimum residuals ($\Delta_{i,j,ij}^{l,\{H,C,J\}}$) in the measured data, according to ℓ_1-norm minimization. The normalization of the states $\{\rho_0^l\}$ is unknown, and that is why it is convenient to adopt the trace norm minimization. The traces of the states $\{\rho_0^l\}$ are constrained by the normalization we imposed on the data, *cf.* Eq. (6.15) and Eq. (6.16), which corresponds to a unity magnetization in equilibrium, *cf.* Eq. (6.14).

After this preliminary estimation, we define a lower bound (N_0) for the trace of the states, namely:

$$N_0 = max\left\{\mathrm{tr}\left(\rho_0^l\right)\right\},\ l = 1,\ldots,20. \tag{6.19}$$

Then we rerun the program in Eq. (6.18) with the additional constraint: $\mathrm{tr}\left(\rho_0^l\right) \geq N_0$. Note that the trace is related to the population in each state. The largest trace yielded by the first run of the state tomography identifies the population of *responsive molecules*. Therefore N_0 is our *reference population*, and

we are interested, at the initial time, in states with the same populations, in order to compare, along the time, the different expectation values.

Now we consider the action of the unknown map $\Phi(t_k)$ upon the initial states determined in the second run of Eq. (6.18), namely (see Eq. (5.4) and Section 5.1 for further details):

$$\tilde{\rho}_k^l \equiv \Phi(t_k)\{\rho_0^l\} = \mathrm{tr}_{\mathcal{H}^I}\Big(J(\Phi(t_k))(\mathbb{1}_{\mathcal{H}^O} \otimes (\rho_0^l)^T)\Big). \tag{6.20}$$

where $J(\Phi(t_k)) \in \mathcal{L}(\mathcal{H}^O \otimes \mathcal{H}^I) = \mathcal{L}((\mathcal{H}^H \otimes \mathcal{H}^C) \otimes (\mathcal{H}^H \otimes \mathcal{H}^C))$.

At each time t_k, quantum process tomography is obtained using a variation of the method in [46], also equipped with a matrix recovery technique like nuclear norm minimization, in order to get the information about the trace:

$$\begin{aligned}
\min_{J(\Phi(t_k)),\,\Delta} \quad & \|J(\Phi(t_k))\|_* + \textstyle\sum_{l,i,j} \Delta_{i,j,ij}^{l,\{H,C,J\}} \\
\text{subject to} \quad & J(\Phi(t_k)) \succeq 0 \\
& |\mathrm{tr}\Big(\tilde{\rho}_k^l\, \sigma_i^H \otimes \mathbb{1}\Big) - \hat{I}_i^{l,H}(t_k)| \leq \Delta_i^{l,H} \\
& |\mathrm{tr}\Big(\tilde{\rho}_k^l\, \mathbb{1} \otimes \sigma_j^C\Big) - \hat{I}_j^{l,C}(t_k)| \leq \Delta_j^{l,C} \\
& |\mathrm{tr}\Big(\tilde{\rho}_k^l\, \sigma_i^H \otimes \sigma_j^C\Big) - \hat{I}_{ij}^{l,J}(t_k)| \leq \Delta_{ij}^{l,J} \\
& \forall\, i,j \in \{x,y,z\} \\
& \forall\, l = 1,\ldots,20 \\
& \Delta_i^{l,H}, \Delta_j^{l,C}, \Delta_{ij}^{l,J} \geq 0\,,
\end{aligned} \tag{6.21}$$

with $\tilde{\rho}_k^l$ defined in Eq. (6.20) and $J(\Phi)$ in Eq. (5.3).

The program in Eq. (6.21) can be interpreted as follows. We want a minimum trace positive semidefinite estimate of the Choi matrix ($J(\Phi(t_k))$), compatible with the data and having minimal losses (see Chapters 3 and 4).

6.5 Results and Discussion

After running the program in Eq. (6.21), we plot in Fig. 6.2 the trace distance between the states ρ_0^l under the action of the map $\Phi(t_k)$ and the rescaled equilibrium state $\frac{1}{M}\rho_{eq}$. We note that the functional form of the curves are like an exponential decay, as expected.

As a byproduct of the process tomography, the map must give us information about the relaxation times $T_1^{C,H,J}$, $T_2^{C,H,J}$. We define the quantity

$$\langle \tilde{M}_{ij} \rangle(t_k) = \frac{\sqrt{\sum_l \Big[\mathrm{tr}\Big(\tilde{\rho}_k^l\, \sigma_i^H \otimes \sigma_j^C\Big) - \mathrm{tr}\Big(\tilde{\rho}_{eq}\, \sigma_i^H \otimes \sigma_j^C\Big)\Big]^2}}{\#l}, \tag{6.22}$$

where $\#l$ stands for the number of summed signals. This quantity is the average of the magnetization in the i,j-direction ($i,j \in \{x,y,z\}$) related to the equilibrium value and it gives us the functional behavior of the magnetization

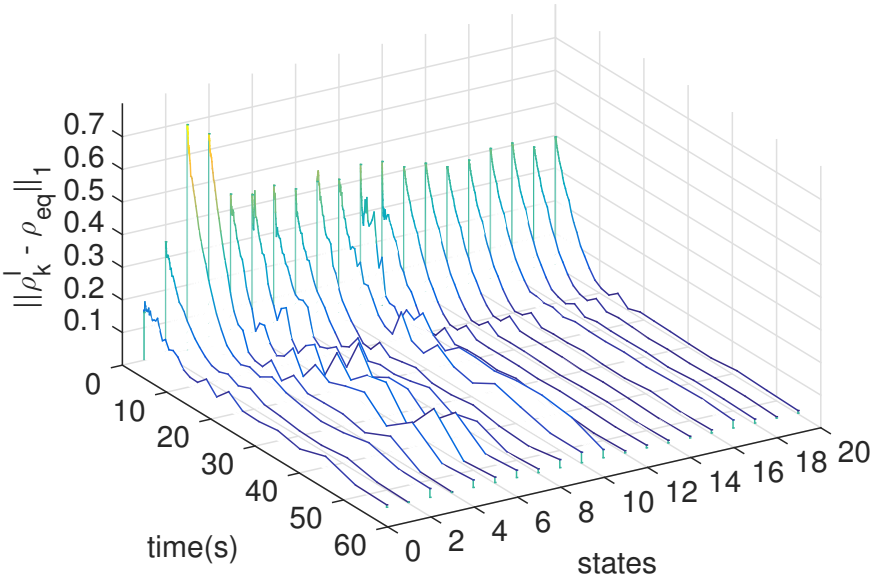

Figure 6.2: The relaxation of the states toward equilibrium. The figure shows the trace distance between the states ρ_0^l under the action of the tomographed quantum map $\Phi(t_k)$ and the equilibrium thermal state. Note that the states in $t = 0$ are perturbations of the thermal state by different sequences of radiofrequency pulses. These perturbed thermal states are the ones useful for quantum information processing simulations.

during the time. As a first attempt to model the functional behavior of Eq. (6.22), one could fit a typical exponential decay like

$$M(t) = M_0 e^{-t/T_*} + c, \tag{6.23}$$

where T_* will be the aimed $T_1^{C,H,J}$, $T_2^{C,H,J}$. This is of course a Markovian approximation of the underlying process. Notwithstanding, in the top panels of Figs. 6.3-6.5, one can see slight deviations of the exponential decay. On the one hand, this could be a signature of non-Markovian behavior, on the other hand, it could also be some artefact caused by the different state preparations necessary for each observation in time. We should emphasise that the tomography itself is independent of this exponential fit, but this simple model yields sensible results, in comparison with previously measured $T_1^{C,H}$, $T_2^{C,H}$. The newly inferred T_1^J, T_2^J is also corroborated by Fig. 6.6, which is independent of the exponential fitting.

The results presented in Figs. 6.3-6.5, and summarized in Table 6.1 with the estimated errors, were computed using nonlinear least squares fitting with a trust-region algorithm [49]. The least squares estimator was chosen to be consistent with the assumption that the experimental errors consist of uniform Gaussian noise, and also consistent with the ℓ_1-norm minimization on the

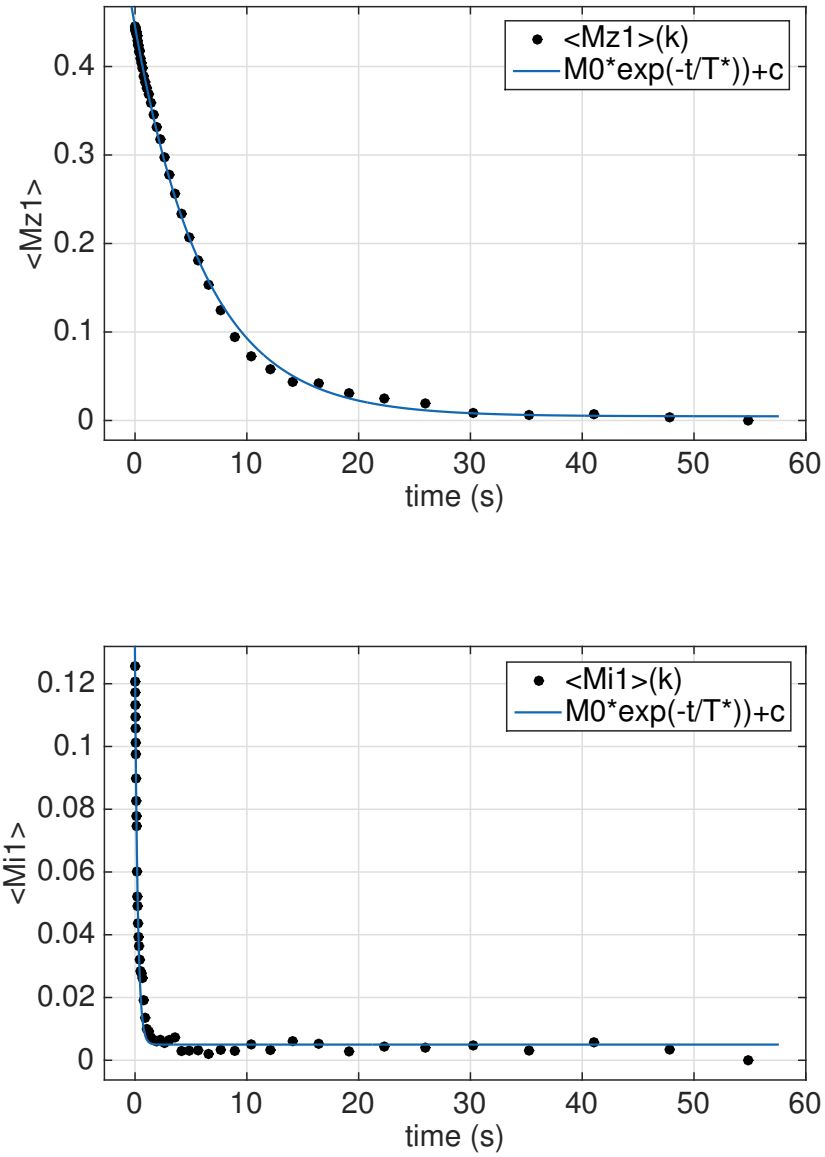

Figure 6.3: The Hydrogen (*uncoupled-*)relaxation. (*Top*): the quantity $\langle \tilde{M}_{z\mathbb{1}} \rangle$ as a function of time. (*Bottom*): the quantity $\frac{\langle \tilde{M}_{x\mathbb{1}} \rangle + \langle \tilde{M}_{y\mathbb{1}} \rangle}{2}$ as function of time. Note that these "magnetizations" are relative to the equilibrium magnetization, according to Eq. (6.22).

observed data in Eq. (6.21).

The pseudo-pure states corresponding to the canonical states ($|\uparrow\uparrow\rangle$, $|\uparrow\downarrow\rangle$,

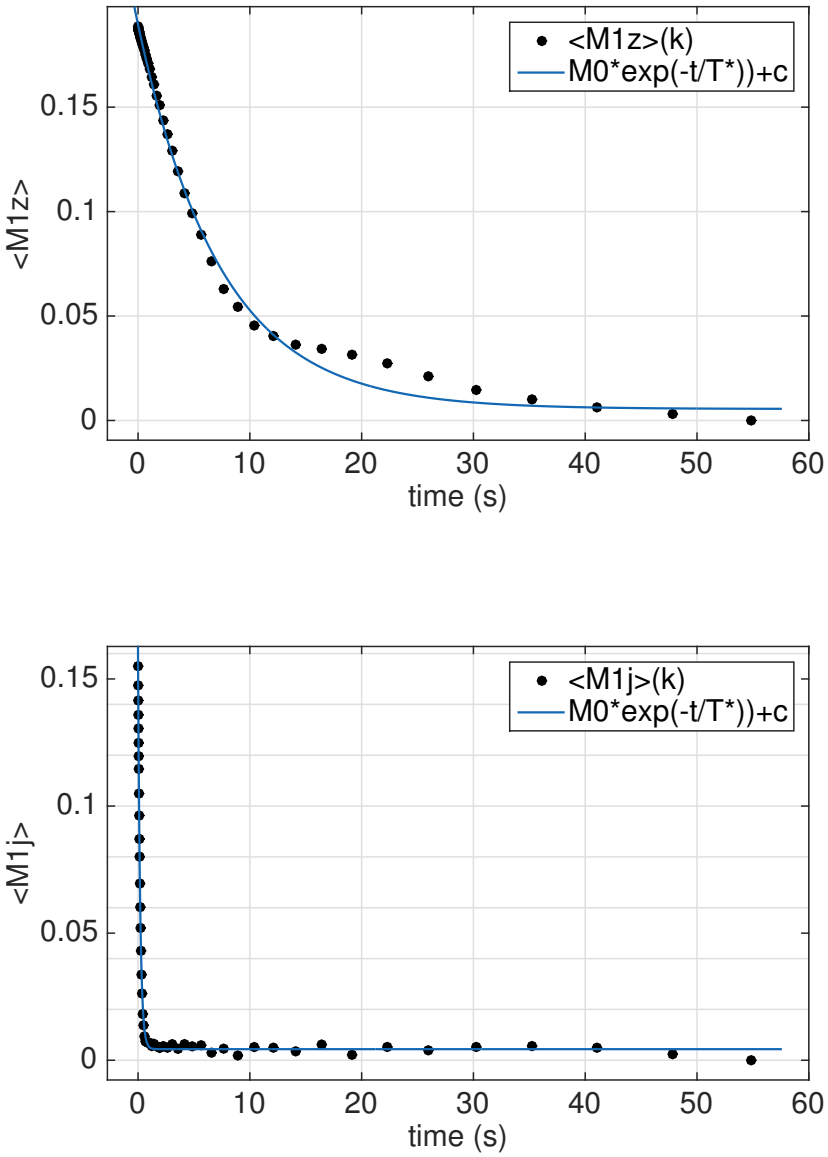

Figure 6.4: The Carbon (*uncoupled-*)relaxation. (*Top*): the quantity $\langle \tilde{M}_{\mathbb{1}z} \rangle$ as a function of time. (*Bottom*): the quantity $\frac{\langle \tilde{M}_{\mathbb{1}x} \rangle + \langle \tilde{M}_{\mathbb{1}y} \rangle}{2}$ as a function of time. Note that these "magnetizations" are relative to the equilibrium magnetization, according to Eq. (6.22).

$|\downarrow\uparrow\rangle$, $|\downarrow\downarrow\rangle$) have magnetization components ($M_{z\mathbb{1}}, M_{\mathbb{1}z}$, M_{zz}) only in the z-direction. From these signals, we can retrieve the $T_1^{C,H,J}$, as we can see in the

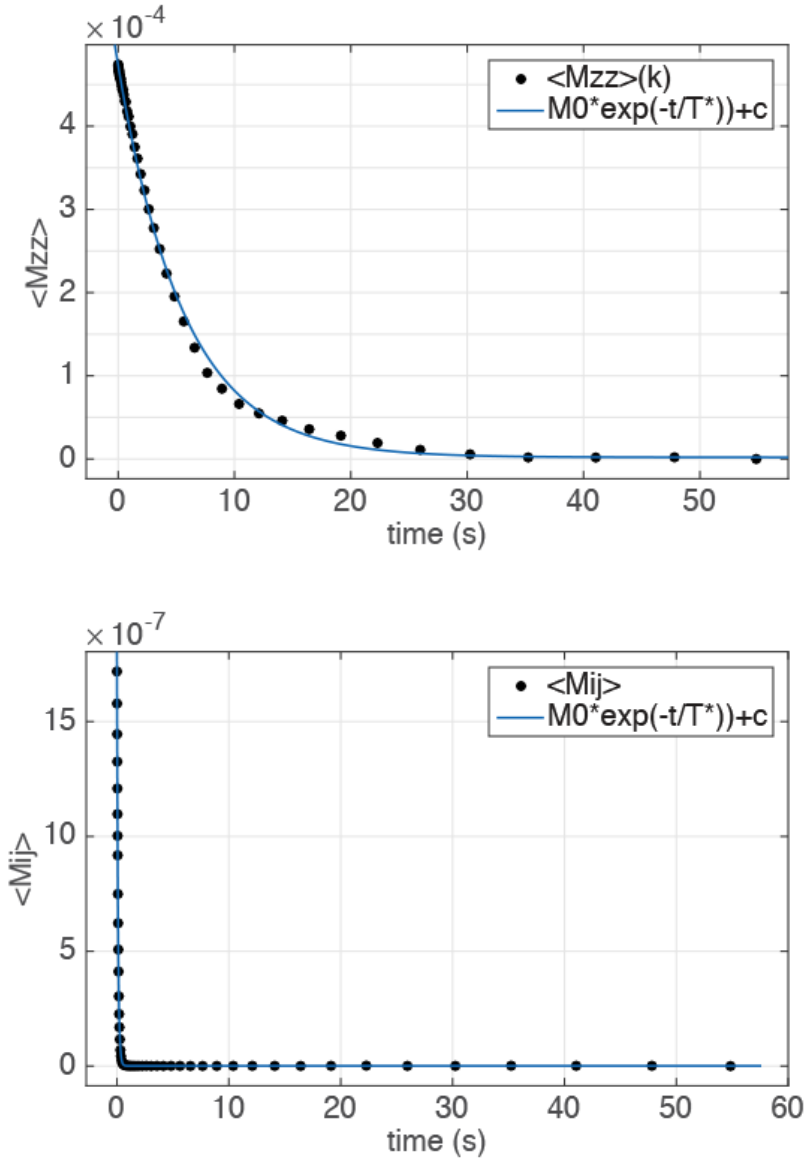

Figure 6.5: The J-coupling relaxation. (*Top*): the quantity $\langle \tilde{M}_{zz} \rangle$ as a function of time. (*Bottom*): the quantity $\frac{\langle \tilde{M}_{xx} \rangle + \langle \tilde{M}_{xy} \rangle + \langle \tilde{M}_{xz} \rangle + \langle \tilde{M}_{yx} \rangle + \langle \tilde{M}_{yy} \rangle + \langle \tilde{M}_{yz} \rangle + \langle \tilde{M}_{zx} \rangle + \langle \tilde{M}_{zy} \rangle}{8}$ as a function of time. Notice how small is this coupled magnetization compared to the uncoupled magnetizations shown in Figs. 3 and 4. Note also that these "magnetizations" are relative to the equilibrium magnetization, according to Eq. (6.22).

top graphics in Figs. 6.3, 6.4, and 6.5.

The other 16 states have both longitudinal and transversal magnetization

	$T_1(s)$	$T_2(s)$
Hydrogen	6.2 ± 0.2	0.24 ± 0.01
Carbon	7.4 ± 0.3	0.192 ± 0.005
J-coupling	5.65 ± 0.07	0.177 ± 0.001

Table 6.1: Relaxation times retrieved from the curve fit in the Figs. 6.3, 6.4, and 6.5 using the model in Eq. (6.23).

components ($M_{*\mathbb{1}}$, $M_{\mathbb{1}*}$, M_{**}) and give us what we need to retrieve the $T_2^{C,H,J}$, as one can see in the bottom graphics in Figs. 6.3, 6.4, and 6.5.

We can conclude from the obtained results that, if one desires to simulate quantum parameters using the chloroform molecule, the time window is $0.177s$, which is the lifetime of the coherent coupling of the spins (T_2^J) in the chloroform molecule. Otherwise, we can consider the system (partially) relaxed and unable to simulate all parameters properly for a quantum computation. In Fig.6.6, we confirm that T_2^J is the characteristic time window for quantum computation simulations by means of the properties of the quantum process relaxation map. One clearly see in that figure that the map is almost perfectly unital and trace-preserving up to $\log(T_2^J = 0.177s) = -1.731$. After T_2^J the map dramatically departs from been both trace-preserving and unital, and therefore the normalization problem comes into play. These results demonstrate the robustness and flexibility of our quantum tomography methodology.

6.6 Agent's design

Summarizing, we performed a quantum process tomography of the relaxation of a two-spin system in an NMR liquid-state experiment. As the experimental data correspond to expectation values of traceless operators, the use of informational incomplete quantum tomography techniques to handle the missing quantum state normalization information was paramount to reconstruct the time dependent non-unital and non trace-preserving quantum map.

The successful approach we employed allowed us to introduce two new parameters (T_1^J, T_2^J) that are relevant for the characterization of the relaxation process. We learned that the coupled magnetizations have a similar behavior to the uncoupled ones, but with different characteristic times. We show that T_2^J characterizes the time window for the simulation of quantum computation with the quantum process tomography, *cf.* Fig. 6.6. For the simulation of quantum maps for times greater than T_2^J, one has to account for the time dependent normalization of the data, where we cannot assume the map associated to the ever present relaxation process being unital and trace-preserving. We also introduced an heuristic for the normalization problem in Eqs. (6.15) and (6.16), which consists in assuming the theoretical equilibrium magnetization as the

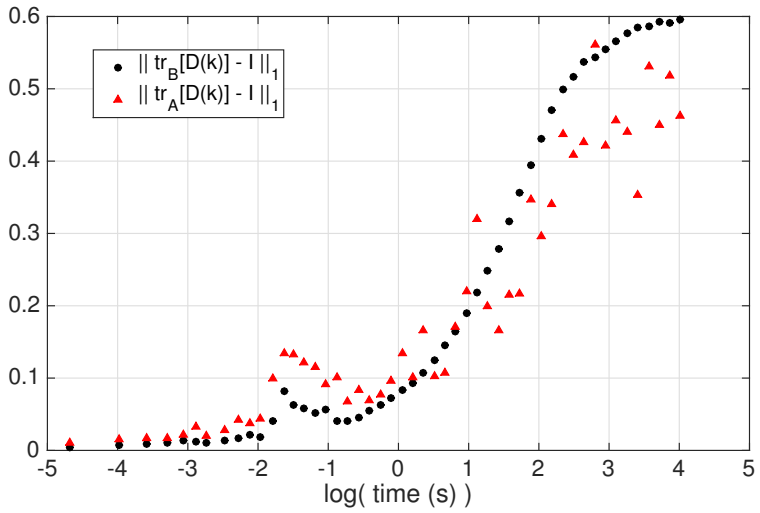

Figure 6.6: Quantification of how much the relaxation process departs from being unital (*cf.* Eq. (5.8)) (black circles), and trace-preserving (*cf.* Eq. (5.7)) (red triangles). The dramatic change in the behavior of the map occurs after $\log\left(T_2^J\right) \approx -1.8$.

reference to compare all the measurements, and the largest trace obtained with the preliminary state tomography (see Eq. (6.18)) as a lower bound for the trace of all states.

So, the final design of our rational agent in this scenario is, for each time step t, do the following (see Fig. 6.7):

(i) : Colect $\hat{I}_i^H(t)$, $\hat{I}_j^C(t)$ and $\hat{I}_{ij}^J(t)$;

(ii) : Our decision rule is normalizing $\hat{I}_i^H(t)$, $\hat{I}_j^C(t)$ and $\hat{I}_{ij}^J(t)$ according to Eqs. (6.15) and (6.16);

(iii) : Run program in Eq. (6.18) and obtain N_0 from Eq. (6.19);

(iv) : Run program in Eq. (6.18) with additional constraint $\mathrm{tr}\left(\rho_0^l\right) \geq N_0$ and collect the set $\{\rho_0^l\}$;

(v) : Run program in Eq. (6.21);

(vi) : Exclude unfeasible solutions and compare the expected losses;

(vii) : Choose and display the best estimate of $J(\Phi(t))$.

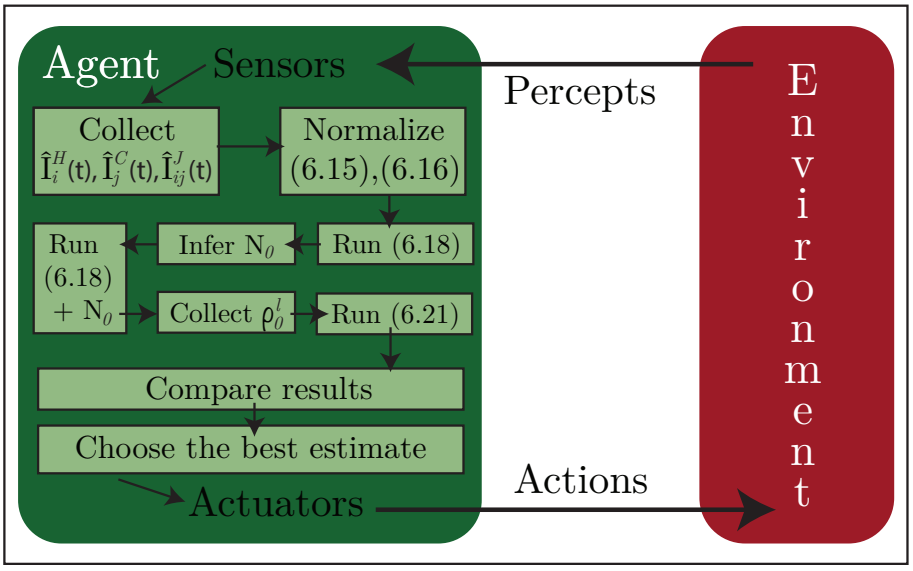

Figure 6.7: Final design of our agent for the quantum process tomography of the relaxation of a two-spin system in an NMR liquid-state experiment.

CHAPTER 7

Conclusion

Every journey has an end and so has this thesis. In Chapter 1, we started this journey discussing what inferences are all about and its possible challenges in the way; in Chapter 2 we borrowed some elements from Artificial Intelligence and Decision Theory to define our rational agents, which gave us a simple, but general framework to include our job as an inferrer; in Chapter 3, we discussed how can this agent infer valid quantum distributions in actual experiments in possession of complete information; in Chapter 4, we upgraded our agent, allowing informationally incomplete contexts; in Chapter 5, we showed how the knowledge of inferring quantum states can help our agent when inferring quantum dynamics; and, finally, we put everything we had in Chapter 6 to model and infer a quantum process tomography in a challenging scenario.

During this PhD, we had opportunities to apply our ideas in real experiments [43, 19, 57, 71, 48], with the last one being the utmost example of how tricky an quantum inference can be, proving that not just efficient methods, but also plausible reasoning, spiced with our background knowledge on the physical system, deserving a separate chapter (see Chapter 6). This opportunity was paramount to mature or analysis and improve our formulations, *e.g.*, our improved VQT_∞ in [22].

Designing rational agents like the ones in Sections 3.5, 4.6, 5.3 and 6.6 help us, not just in the process writing this thesis in a pedagogical way. We expect to inspire this train of thought while deliberating quantum estimates: although it can be easy with our procedures, being that flexible (in formulating loss functions, for example), we must deal with it critically and reason carefully, avoiding overestimations when possible[1].

We also designed our agents in an efficient way in tandem with semi-definite programs (SDP) [7], which are convex linear optimization problems with linear matrix constraints. A great merit of an SDP is the many available numerical methods which converge efficiently with use of polynomial complexity algorithms [7, 69, 67] to find its global minimum.

[1] Remember: purity is a myth!

Bibliography

[1] J. B. Altepeter, E. R. Jeffrey, and P. G. Kwiat. "Photonic State Tomography." In: ed. by P.R. Berman and C.C. Lin. Vol. 52. Advances In Atomic, Molecular, and Optical Physics. Academic Press, 2005, pp. 105 –159. DOI: DOI:10.1016/S1049-250X(05)52003-2.

[2] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White. "Ancilla-Assisted Quantum Process Tomography." In: *Phys. Rev. Lett.* 90 (19 2003), p. 193601. DOI: 10.1103/PhysRevLett.90.193601.

[3] K. M. R. Audenaert and S. Scheel. "Quantum tomographic reconstruction with error bars: a Kalman filter approach." In: *New Journal of Physics* 11.2 (2009), p. 023028. DOI: 10.1088/1367-2630/11/2/023028.

[4] A. Bendersky, F. Pastawski, and J. P. Paz. "Selective and Efficient Estimation of Parameters for Quantum Process Tomography." In: *Phys. Rev. Lett.* 100 (19 2008), p. 190403. DOI: 10.1103/PhysRevLett.100.190403.

[5] A. Bendersky, F. Pastawski, and J. P. Paz. "Selective and efficient quantum process tomography." In: *Phys. Rev. A* 80 (3 2009), p. 032116. DOI: 10.1103/PhysRevA.80.032116.

[6] N. Boulant, T. F. Havel, M. A. Pravia, and D. G. Cory. "Robust method for estimating the Lindblad operators of a dissipative quantum process from measurements of the density operator at multiple time points." In: *Physical Review A* 67.4 (Apr. 2003), pp. 042322–12.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[8] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist. "Choice of measurement sets in qubit tomography." In: *Phys. Rev. A* 78 (5 2008), p. 052122. DOI: 10.1103/PhysRevA.78.052122.

[9] V. Buzek, G. Drobny, R. Derka, G. Adam, and H. Wiedemann. "Quantum State Reconstruction From Incomplete Data." In: *Chaos, Solitons & Fractals* 10.6 (1999), pp. 981–1074. DOI: 10.1016/S0960-0779(98)00144-1.

[10] R. Chaves, L. Luft, T. O. Maciel, D. Gross, D. Janzing, and B. Schölkopf. "Inferring latent structures via information inequalities." In: *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI2014)* (2014).

[11] M. Choi. "Completely positive linear maps on complex matrices." In: *Linear algebra and its applications* 10.3 (1975), pp. 285–290.

[12] I. L. Chuang and M. A. Nielsen. "Prescription for experimental determination of the dynamics of a quantum black box." In: *Journal of Modern Optics* 44.11-12 (1997), pp. 2455–2467. DOI: 10.1080/09500349708231894.

[13] D. G. Cory, A. F. Fahmy, and T. F. Havel. "Ensemble quantum computing by NMR spectroscopy." In: *Proceedings of the National Academy of Sciences of the United States of America* 94.5 (1997), pp. 1634–1639.

[14] G. M. D'Ariano and P. Lo Presti. "Quantum Tomography for Measuring Experimentally the Matrix Elements of an Arbitrary Quantum Operation." In: *Phys. Rev. Lett.* 86 (19 2001), pp. 4195–4198. DOI: 10.1103/PhysRevLett.86.4195.

[15] G. M. D'Ariano, D. F. Magnani, and P. Perinotti. "Adaptive Bayesian and frequentist data processing for quantum tomography." In: *Physics Letters A* 373.12 (2009), pp. 1111–1115.

[16] G. M. D'Ariano and P. Perinotti. "Optimal Data Processing for Quantum Measurements." In: *Phys. Rev. Lett.* 98 (2 2007), p. 020403. DOI: 10.1103/PhysRevLett.98.020403.

[17] T. Debarba, T. O. Maciel, and R. O. Vianna. "Witnessed entanglement and the geometric measure of quantum discord." In: *Phys. Rev. A* 86 (2 2012), p. 024302. DOI: 10.1103/PhysRevA.86.024302.

[18] A. Einstein. "Über einem die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt." In: *Annalen der Physik* 4 (1905).

[19] J. G. Filgueiras, T. O. Maciel, R. E. Auccaise, R. O. Vianna, R. S. Sarthour, and I. S. Oliveira. "Experimental implementation of a NMR entanglement witness." In: *Quantum Information Processing* 11.6 (2012), pp. 1883–1893. DOI: 10.1007/s11128-011-0341-z.

[20] A. Gavini-Viana, A. M. Souza, D. O. Soares-Pinto, J. Teles, R. S. Sarthour, E. R. deAzevedo, T. J. Bonagamba, and I. S. Oliveira. "Normalization procedure for relaxation studies in NMR quantum information processing." In: *Quantum Information Processing* 9.5 (Dec. 2009), pp. 575–589.

[21] D. S. Gonçalves, M. A. Gomes-Ruggiero, C. Lavor, O. J. Farías, and P. H. Souto Ribeiro. "Local solutions of Maximum Likelihood Estimation in Quantum State Tomography." In: *Quantum Information and Computation* 12.9 (2012), pp. 775–790. eprint: http://arxiv.org/abs/1103.3682v1.

[22] D. S. Gonçalves, C. Lavor, M. A. Gomes-Ruggiero, A. T. Cesário, R. O. Vianna, and T. O. Maciel. "Quantum state tomography with incomplete data: Maximum entropy and variational quantum tomography." In: *Phys. Rev. A* 87 (5 2013), p. 052140. DOI: 10.1103/PhysRevA.87.052140.

[23] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory. "Robust online Hamiltonian learning." In: *New Journal of Physics* 14.10 (Oct. 2012), pp. 103013–32.

[24] D. Gross, Y. Liu, S. T. Flammia, S. Becker, and J. Eisert. "Quantum State Tomography via Compressed Sensing." In: *Phys. Rev. Lett.* 105 (15 2010), p. 150401. DOI: 10.1103/PhysRevLett.105.150401.

[25] C. Helmberg, F. Rendl, R. Vanderbei, and H. Wolkowicz. "An interior-point method for semidefinite programming." In: *SIAM Journal on Optimization* 6 (1996), pp. 342–361.

[26] H. Hertz. "Über einen Einfluss des ultravioletten Lichtes auf die electrische Entladung." In: *Annalen der Physik* 267.8 (1887), pp. 983–1000.

[27] R. Horodecki, M. Horodecki, and P. Horodecki. "Entanglement processing and statistical inference: The Jaynes principle can produce fake entanglement." In: *Phys. Rev. A* 59 (3 1999), pp. 1799–1803. DOI: 10.1103/PhysRevA.59.1799.

[28] Z. Hradil, J. Řeháček, J. Fiurášek, and M. Ježek. *Maximum-Likelihood Methods in Quantum Mechanics*. Vol. 649. Lecture Notes in Physics. 2004, pp. 163–172. DOI: 10.1007/978-3-540-44481-7_3.

[29] F. Iemini, T. O. Maciel, and R. O. Vianna. "Entanglement of indistinguishable particles as a probe for quantum phase transitions in the extended Hubbard model." In: *Phys. Rev. B* 92 (7 2015), p. 075423. DOI: 10.1103/PhysRevB.92.075423.

[30] F. Iemini, T. O. Maciel, T. Debarba, and R. O. Vianna. "Quantifying quantum correlations in fermionic systems using witness operators." In: *Quantum Information Processing* 12.2 (2013), pp. 733–746. DOI: 10.1007/s11128-012-0415-6.

[31] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White. "Measurement of qubits." In: *Phys. Rev. A* 64.5 (2001). DOI: 10.1103/PhysRevA.64.052312.

[32] E. T. Jaynes. "Information Theory and Statistical Mechanics I." In: *Physical Review* 106.4 (1957), pp. 620–630.

[33] E. T. Jaynes. "Information Theory and Statistical Mechanics II." In: *Physical Review* 108.2 (1957), pp. 171–190.

[34] W. Karush. "Minima of functions of several variables with inequalities as side conditions." In: *Traces and Emergence of Nonlinear Programming*. Springer, 2014, pp. 217–245.

[35] E. Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Kluwer Academic Publishers, 2002.

[36] E. Krauss. "A representation of arbitrary maximal monotone operators via subgradients of skew-symmetric saddle functions." In: *Nonlinear Analysis: Theory, Methods & Applications* 9.12 (1985), pp. 1381–1399.

[37] H. W. Kuhn and A. W. Tucker. "Nonlinear Programming." In: *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*. 1951, pp. 481–492.

[38] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien. "Quantum computers." In: *Nature* 464.7285 (2010), pp. 45–53. URL: http://dx.doi.org/10.1038/nature08812.

[39] N. K. Langford. "Errors in quantum tomography: diagnosing systematic versus statistical errors." In: *New Journal of Physics* 15.3 (2013), p. 035003. URL: http://stacks.iop.org/1367-2630/15/i=3/a=035003.

[40] E. L. Lehman and G. Casella. *Theory of Point Estimation*. Springer, 1998.

[41] G. M. Leskowitz and L. J. Mueller. "State interrogation in nuclear magnetic resonance quantum-information processing." In: *Phys. Rev. A* 69 (5 2004), p. 052302. DOI: 10.1103/PhysRevA.69.052302.

[42] D. W. Leung. "Choi's proof as a recipe for quantum process tomography." In: *Journal of Mathematical Physics* 44.2 (2003), pp. 528–533. DOI: http://dx.doi.org/10.1063/1.1518554.

[43] G. Lima, E. S. Gómez, A. Vargas, R. O. Vianna, and C. Saavedra. "Fast entanglement detection for unknown states of two spatial qutrits." In: *Phys. Rev. A* 82 (1 2010), p. 012302. DOI: 10.1103/PhysRevA.82.012302.

[44] J. Löfberg. *YALMIP : A Toolbox for Modeling and Optimization in MATLAB*. Taipei, Taiwan, 2004. URL: http://control.ee.ethz.ch/\~{}joloef/yalmip.php.

[45] T. O. Maciel, A. T. Cesário, and R. O. Vianna. "Variational quantum tomography with incomplete information by means of semidefinite programs." In: *International Journal of Modern Physics C* 22.12 (2011), pp. 1361–1372.

[46] T. O. Maciel and R. O. Vianna. "Optimal estimation of quantum processes using incomplete information: variational quantum process tomography." In: *Quantum Information and Computation* 12.5 (2012), pp. 0442–0447.

[47] T. O. Maciel and R. O. Vianna. "Viable entanglement detection of unknown mixed states in low dimensions." In: *Phys. Rev. A* 80 (3 2009), p. 032325. DOI: 10.1103/PhysRevA.80.032325.

[48] T. O. Maciel, R. O. Vianna, R. S. Sarthour, and I. S. Oliveira. "Quantum process tomography with informational incomplete data of two J -coupled heterogeneous spins relaxation in a time window much greater than T 1." In: *New Journal of Physics* 17.11 (2015), p. 113012. URL: http://stacks.iop.org/1367-2630/17/i=11/a=113012.

[49] MATLAB. *version 8.5.0.197613 (R2015a)*. Natick, Massachusetts: The MathWorks Inc., 2015.

[50] M. Mohseni and D. A. Lidar. "Direct Characterization of Quantum Dynamics." In: *Phys. Rev. Lett.* 97 (17 2006), p. 170501. DOI: 10.1103/PhysRevLett.97.170501.

[51] M. Mohseni and D. A. Lidar. "Direct characterization of quantum dynamics: General theory." In: *Phys. Rev. A* 75 (6 2007), p. 062331. DOI: 10.1103/PhysRevA.75.062331.

[52] M. Mohseni, A. T. Rezakhani, J. T. Barreiro, P. G. Kwiat, and A. Aspuru-Guzik. "Quantum process estimation via generic two-body correlations." In: *Phys. Rev. A* 81 (3 2010), p. 032102. DOI: 10.1103/PhysRevA.81.032102.

[53] A. S. Nemirovski and M. J. Todd. "Interior-point methods for optimization." In: *Acta Numerica* 17.-1 (2008), pp. 191–234. DOI: 10.1017/S0962492906370018.

[54] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. 1st ed. Cambridge University Press, 2000.

[55] M. A. Nielsen, E. Knill, and R. Laflamme. "Complete quantum teleportation using nuclear magnetic resonance." In: *Nature* 396.6706 (Nov. 1998), pp. 52–55. URL: http://dx.doi.org/10.1038/23891.

[56] I. Oliveira, R. Sarthour, T. Bonagamba, E. Azevedo, and J.C.C. Freitas. *NMR Quantum Information Processing*. Elsevier Science, 2011.

[57] W. M. Pimenta, B. Marques, T. O. Maciel, R. O. Vianna, A. Delgado, C. Saavedra, and S. Pádua. "Minimum tomography of two entangled qutrits using local measurements of one-qutrit symmetric informationally complete positive operator-valued measure." In: *Phys. Rev. A* 88 (1 2013), p. 012112. DOI: 10.1103/PhysRevA.88.012112.

[58] G. Pólya. *Mathematics and Plausible Reasoning: Patterns of plausible inference*. Mathematics and Plausible Reasoning. Princeton University Press, 1968. ISBN: 9780691025100. URL: https://books.google.com.br/books?id=apw5ZSlnxxsC.

[59] J. Řeháček and Z. Hradil. "Maximum entropy assisted maximum likelihood inversion." In: *Lasers and Electro-Optics Europe, 2005. CLEO/Europe. 2005 Conference on*. 2005, p. 466. DOI: 10.1109/CLEOE.2005.1568244.

[60] J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky. "Diluted maximum-likelihood algorithm for quantum tomography." In: *Physical Review A* 75.4 (2007), p. 042108. DOI: 10.1103/PhysRevA.75.042108.

[61] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves. "Symmetric informationally complete quantum measurements." In: *Journal of Mathematical Physics* 45.6 (2004), pp. 2171–2180. DOI: 10.1063/1.1737053.

[62] M. Ringbauer, C. J. Wood, K. Modi, A. Gilchrist, A. G. White, and A. Fedrizzi. "Characterizing Quantum Dynamics with Initial System-Environment Correlations." In: *Physical Review Letters* 114.9 (Mar. 2015), pp. 090402–5.

[63] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd ed. Prentice Hall, Dec. 2009. ISBN: 0136042597.

[64] M. Sarovar and G. J. Milburn. "Optimal estimation of one-parameter quantum channels." In: *Journal of Physics A: Mathematical and General* 39.26 (2006), p. 8487. URL: http://stacks.iop.org/0305-4470/39/i=26/a=015.

[65] C. T. Schmiegelow, M. A. Larotonda, and J. P. Paz. "Selective and Efficient Quantum Process Tomography with Single Photons." In: *Phys. Rev. Lett.* 104 (12 2010), p. 123601. DOI: 10.1103/PhysRevLett.104.123601.

[66] C. E. Shannon. *A Mathematical Theory of Communication*. Vol. 27. University of Illinois Press, Urbana, 1949, 1948, pp. 379–423, 623–656.

[67] J. F. Sturm. *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*. 1998.

[68] Y. S. Teo, B. Stoklasa, B. Englert, J. Reháček, and Z. Hradil. "Incomplete quantum state estimation: A comprehensive study." In: *Phys. Rev. A* 85 (4 2012), p. 042317. DOI: 10.1103/PhysRevA.85.042317.

[69] *The MOSEK optimization software*. http://www.mosek.com/. URL: http://www.mosek.com/.

[70] L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang. "Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance." In: *Nature* 414.6866 (Dec. 2001), pp. 883–887. URL: http://dx.doi.org/10.1038/414883a.

[71] R. O. Vianna, A. Crespi, R. Ramponi, R. Osellame, L. Sansoni, G. Milani, P. Mataloni, and F. Sciarrino. "Variational quantum process tomography of two-qubit maps." In: *Phys. Rev. A* 87 (3 2013), p. 032304. DOI: 10.1103/PhysRevA.87.032304.

[72] J. Watrous. *Theory of Quantum Information lecture notes*. 2011. URL: http://www.cs.uwaterloo.ca/~watrous.

[73] Y. S. Weinstein, T. F. Havel, J. Emerson, N. Boulant, M. Saraceno, S. Lloyd, and D. G. Cory. "Quantum process tomography of the quantum Fourier transform." In: *The Journal of Chemical Physics* 121.13 (2004), pp. 6117–18.

[74] G.A. Young and R.L. Smith. *Essentials of Statistical Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2005. ISBN: 9780521839716.

[75] K. Zyczkowski, K. A. Penson, I. Nechita, and B. Collins. "Generating random density matrices." In: *Journal of Mathematical Physics* 52.6, 062201 (2011), p. 062201. DOI: 10.1063/1.3595693.