Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Physics

Learning from temporal correlations in quantum data

Marco Radaelli

Supervisor: Prof. Felix C. Binder

2024-25

Submitted in Partial Fullfillment of the requirements

for the Degree of Doctor of Philosophy in Physics

PLAGIARISM DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

LIBRARY DECLARATION

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

JOINT WORK DECLARATION

This thesis contains published and unpublished work carried out in various degrees of collaboration with other researchers, both in Trinity College Dublin and in other international Universities and Research Institutions. Their contribution is acknowledged at the beginning of the relevant chapters.

Dublin, Ireland, April 2025

PUBLICATIONS

This thesis is based on the following publications:

[1] - Marco Radaelli, Gabriel T. Landi, Kavan Modi, Felix C. Binder. "Fisher information of correlated stochastic processes". New J. Phys. **25**, 053037 (2023)

[2] - Joseph A. Smiga, Marco Radaelli, Felix C. Binder, Gabriel T. Landi "Stochastic metrology and the empirical distribution". Phys. Rev. Res. **5**, 033150 (2023)

[3] - Marco Radaelli, Gabriel T. Landi, Felix C. Binder. "Gillespie algorithm for quantum jump trajectories". Phys. Rev. A **110**, 062212 (2024)

as well as the following preprints:

[4] - Marco Radaelli, Joseph A. Smiga, Gabriel T. Landi, Felix C. Binder. "Parameter estimation for quantum jump unraveling". arXiv 2402.06556 (2024)

[5] - Saulo V. Moreira, Marco Radaelli, Alessandro Candeloro, Felix C. Binder, Mark T. Mitchison "Precision bounds for multiple currents in open quantum systems". arXiv 2411.09088 (2024)

and the following works in preparation:

[6] - Marco Radaelli, Joseph A. Smiga, Felix C. Binder, Gabriel T. Landi. "Hypothesis testing for quantum-generated strings", *in preparation*

[7] - Marco Radaelli, Simon C. Milz, Felix C. Binder. "Collisional tomography of a quantum channel", *in preparation*

During the course of the PhD, I have also worked on the following, that is, though, not included in the thesis:

Marco Radaelli, Claudia Benedetti, Stefano Olivares. "On the sampling of an almost-uniform distribution via a quantum walk on a cycle", *in preparation*

CODE AVAILABILITY

The code employed in the simulations is freely accessible on the candidate's GitHub profile.

FUNDING ACKNOWLEDGEMENTS

The research undertaken in this PhD has been funded by a Research Ireland Government of Ireland Postgraduate scheme scholarship, under grant number GOIPG/2022/2321.

I am grateful for hospitality and support to the University of Rochester, and to the group of Prof. Gabriel T. Landi, in Rochester, NY, United States of America.

Part of the calculations and simulations have been performed on external resources. I acknowledge support by Research IT at TCD, as well as the use of the Kesha calculator, belonging to the ToCQS group in TCD. I also acknowledge computational resources and support from the Irish Centre for High Performance Computing (ICHEC), under class C project "Learning of completely positive quantum maps via partial sequential access".

ACKNOWLEDGEMENTS

Saying that completing a PhD is a long undertaking may come across as something completely obvious, but it is not. When I started, right after the end of the COVID pandemic, I had not realised that the PhD would have been the next four years of my life. Not an aspect of those years, something accessory to a life that otherwise proceeds untouched, but the entirety of those years. Everything in these four years has been shaped by this PhD – starting from the choice of moving to Dublin.

My first thank you, then, goes to the city of Dublin, that I have hated at the beginning with all my heart, and then slowly learned to feel and love. To this strange city, in the middle of an island in the Atlantic, to its rainy weather and cold summers, to the short days of winter when light is nowhere to be found, to the unreasonably noisy pubs and the never-ending flow of beer. Thanks for having been the background of these four years, the scenery on which so much has played out.

Thanks to the person who brought me here, my supervisor Prof. Felix C. Binder. His support has truly been invaluable. While never losing sight of my work, he has always left me ample space to explore and enjoy what I was doing. I thank Felix for his way of being, first and foremost, a human being, and for his attitude of care towards the group, and me in particular.

Thanks also to Prof. Gabriel T. Landi, with whom I have shared a number of research projects, and who has also been an amazing host at the University of Rochester, in the US, during a couple of months I spent there.

Theoretical research is something that, in the end, one does alone. However, I would like to thank all the members of my group, that made coming to the office everyday a joy and a pleasure. Their support in the endless frustrations of research, and their ability to sincerely rejoice for the successes one of each other has been one of the best things of this PhD.

Thanks in particular to Magdalini, my fellow PhD student in the group for most of the time. For making me discover the nuances of coffee, but especially for always offering a patient and listening ear.

Thanks to Sindre, with whom I shared the habit of a second breakfast every single morning, for being the kind of friend with whom I almost never agree on anything, without becoming less friends.

Thanks to Laetitia for the exploration of a number of incredible pastry shops in Dublin, and for having shared two trips to two different continents, and never having complained about my obsession with being at the airport four hours before a flight.

Thanks to Simon, post-doc in the group, for his incredible patience with me, and for never shutting

me down despite my countless requests for help and opinions.

Thanks to my old coursemates in Milan: their (not only virtual) closeness in these years has truly been my lifeline in many occasions during these years: Martina, Cesare, Giorgia, Francesco, Eleonora, Elia, Tommaso, Marta, Mattia.

Thanks to my historical friends from Como, for their strong support, many shared discussions, and being something to go back to: Paola, Marco, Miriam, Mattia, Lorenzo.

Thanks to my family, and in particular my siblings Margherita and Giulio, for being the people I grew up with, and with whom I still share the most in our way of understanding the world.

ABSTRACT

Many applications in quantum technologies require precise characterisation and calibration of quantum devices. Oftentimes, one has to estimate dynamical parameters from the outcomes of the quantum systems. In this thesis, we address this problem, focussing specifically on the role of time correlations in the output data. We give novel precision bounds for the estimated parameters, and introduce techniques to saturate such bounds.

The achievable precision when estimating a parameter is bounded by the Cramér-Rao bound, in terms of the Fisher information of the outcomes' probability distribution. While estimation is well understood for independent variables, the role of correlations is still largely unexplored. How do the precision bounds for independent random variables generalise to the correlated case? And how can estimation be actually performed, in a computationally feasible manner?

We first derive precision bounds for classical stationary processes with a finite memory. We prove an asymptotic expression for the Fisher information in the limit of a large amount of data. Quite counterintuively, correlations can both hamper or improve estimation performances: we give general rules of thumb to determine when they are helpful and when they aren't. We apply our results to temperature estimation on an Ising spin chain, and we find crucially different behaviour between the ferromagnetic and anti-ferromagnetic phases, in line with our heuristics.

We then focus on the quantum jump unravelling (QJU) of open quantum system dynamics. These processes describe continuously monitored quantum systems, where time and type of each emission are recorded. We derive analytical expressions for the Fisher information for processes in which all the memory is lost upon every emission, known as renewal processes. Beyond the renewal hypothesis, the problem cannot be treated analytically. We generalise the previously introduced monitoring operator formalism to compute the Fisher information on quantum trajectories. We develop the quantum Gillespie algorithm, for a more efficient simulation on trajectories. We illustrate these numerical methods with case studies from atomic physics and quantum optics.

In quantum thermodynamics, one can express the kinetic and thermodynamic uncertainty relations in terms of the Fisher information on trajectories. Exploiting our results on metrology for the QJU, we introduce a tighter version of the kinetic uncertainty relation for quantum systems undergoing continuous monitoring.

We consider the problem of learning a quantum channel acting on a system, by only having access to a

subsystem, while quantum memory keeps flowing in the unmonitored portion. We term this framework *collisional learning*. As can be expected, not all the features of the channel can be learnt in this scenario, and we characterise the emerging gauge freedoms. We define a new notion of similarity between the estimated channel and the original one, in terms of their action on the monitored subsystem. We develop numerically efficient methods to perform collisional learning, and show that they can effectively learn two-qubits Hamiltonians.

Frequently, the output data are not available in raw form, but rather subject to some post-processing. In general, this reduces the Fisher information in the data, and consequently the estimation precision. We derive expressions for the Fisher information contained in the outcomes of two of the most common post-processing strategies, the empirical distribution and the sample average. For quantum jump unravelling outputs, we show what happens if some of the jump channels cannot be monitored, or cannot be distinguished one from each other. This amounts to a reduction of the Fisher information. Quite surprisingly, the renewal nature of the process can change, depending on the post-processing.

Throughout the thesis, we discuss hypothesis testing problems on the side of estimation. This is a framework in which, instead of estimating continuous parameters, one chooses the most likely among a discrete sets of models that can have generated the process. We give both analytical and numerical results on the error probabilities.

Finally, in a more speculative Outlook chapter, we give some hints at possible directions in which our research could be carried forward.

CONTENTS

Introduction 12

I Basics: Statistical concepts 16

I.1. Stochastic processes 16

I.2. Estimation theory 21

I.3. Hypothesis testing 24

II Basics: quantum dynamics 28

II.1. CPTP maps and their representations 28

II.2. The GKSL master equation and its unravelings 32

III Estimation theory for stochastic processes 34

III.1. The Fisher information is not subadditive 35

III.2. Fisher information for finite Markov order 37

III.3. Estimators based on the sample mean 38

III.4. Application: spin chain thermometry 40

III.5. Learning an ϵ-machine 43

IV Estimation and hypothesis testing for the quantum jump unraveling 44

IV.1. The measurement record of a jump process 45

IV.2. Renewal processes 47

IV.3. Processes with closed patterns 52

IV.4. Non-renewal processes and the monitoring operator formalism 57

V Numerical simulations: the quantum Gillespie algorithm 64

V.1. Simulating quantum jump trajectories 65

V.2. The Gillespie algorithm 67

V.3. Case study: single atom maser model 71

V.4. Case study: coupled quantum dots in an optical cavity 76

V.5. Fisher information and the kinetic uncertainty relation 79

VI Collisional learning of a quantum channel 85

VI.1. Collisional trajectories 87

VI.2. Gauge degrees of freedom and process distance 88

VI.3. A tradeoff between precision and expressivity 91

VI.4. Learning a Hamiltonian 94

VII Estimation with classical postprocessing 98

VII.1. Data-processing inequality 98

VII.2. Empirical distribution and sample average 99

VII.3. Partial monitoring and channel merging 104

Discussion and outlook 107

Bibliography 110

Vi siete mai chiesti veramente che cosa sia il cielo?

Did you ever ask, for real, what is the sky?

P. Florenskij, *Lectures on the origins of Western Philosophy*

Et si habuero prophetiam, et noverim mysteria omnia, et omnem scientiam: et si habuero omnem fidem ita ut montes transferam, caritatem autem non habuero, nihil sum.

And though I have the gift of prophecy, and understand all mysteries, and all knowledge; and though I have all faith, so that I could remove mountains, and have not love, I am nothing

St. Paul, *First letter to the Corinthians*

May the road rise to meet you,
May the wind be always at your back.
May the sun shine warm upon your face,
The rains fall soft upon your fields.
And until we meet again,
May God hold you in the palm of his hand.

Ancient Irish blessing

Introduction

Quantum science and technology are among the most active research areas in theoretical and applied physics. Quantum devices such as fault-tolerant quantum computers and quantum sensors require precise monitoring and control of the dynamical parameters of both wanted (signal) and unwanted (noise) origin. Information about quantum systems can be obtained from their outputs. The theorist can ask what are the ultimate bounds to information one can gain about a physical process. For example, how precisely can we estimate the noise strength, or determine what physical process is the cause of a specific effect? These questions are the object of the research field known as *quantum metrology*.

In most of the theory of quantum metrology, one assumes that the experimenter has access to a number of identical and independent (i.i.d.) copies of the system of interest. For instance, if one is to determine a parameter in a quantum channel, multiple copies of the channel will be available. Research is then mostly concerned with the optimal choice of input states and measurement schemes, to extract as much information as possible about the parameter [8, 9].

The i.i.d. hypothesis immensely simplifies the treatment of metrological problems. However, it is frequently broken by real systems, which exhibit different forms of correlations in time. Furthermore, in many situations one cannot optimise over measurement schemes, as they are fixed by the specific system at hand. This thesis deals with this situation: correlated estimation, with a fixed measurement scheme. Since the measurement scheme is fixed, we use classical statistics techniques, acting on correlated data generated by quantum processes.

The problem of metrology, be it classical or quantum, has historically seen two separate lines of development. On one hand, one can consider the task of determining the value of an unknown parameter encoded in the system. The expected answer to this problem is a continuous parameter, or a collection of continuous parameters. Besides, one may want to determine which one of a finite collection of causes generates a given result: the expected answer is a choice in a finite set. The former problem is conventionally termed *estimation*, the latter is called *hypothesis testing*. Both are widespread tasks, well beyond the realm of physics or even natural sciences. For instance, an example of an estimation problem may be calculating the volatility of some asset on the stock market [10–12], and an epidemiologist trying to determine from population data whether a pathogen is airborne or not [13] is solving a hypothesis testing problem. In the words of Carl W. Helstrom, one of the pioneers of quantum information theory:

> Modern statistical theory also has a normative and methodological aspect, which appears in

> its treatment of hypothesis testing and estimation. It seeks the best procedures for making statements about the condition of a system under observation, statements that are framed as decisions among hypotheses about the system, or as estimates of numerical parameters characterizing it. The statements are based on observational data subject to unavoidable random error. The best methods are those that minimize the influence of error, and by evaluating their quality, it is possible to determine the ultimate limits imposed by statistical uncertainty on the accuracy of decisions and measurements. [14]

The classical theory of estimation [15–17] first appeared in the early 20th century, with the introduction of the Fisher information and the fundamental Cramér-Rao bound [18, 19]. Given a parameter-dependent probability distribution, the theory is concerned with techniques to estimate the parameter from a string of samples. The Cramér-Rao bound limits the maximum precision in the estimation of the parameter in terms of the Fisher information content of the probability distribution.

The first attempts at a quantum theory of estimation started in the 1960s [14, 20–22], with the introduction of the *Helstrom information number*, later the quantum Fisher information for one [23] and multiple [24] parameters. Instead of a classical sample, in quantum estimation one is given a quantum state, and can optimise over the measurement scheme. A significant effort has been devoted to identifying cases in which the bounds afforded by this quantum freedom improve on the classical counterparts [25, 26], and how to saturate these bounds [27]. Later, a deep connection between estimation theory and the geometric structure of the space of quantum states was discovered [28–30]. More recently, researchers have tried to identify the quantum resources required for better-than-classical scalings, such as entanglement [8, 9, 31–33] and squeezing [34].

As mentioned before, in most cases quantum estimation theory assumes access to a number of i.i.d. copies of the system. Beyond the i.i.d. hypothesis, two main avenues of research have emerged. On one hand, one can consider the classical outcomes of a correlated quantum process in time, fixing the measurement scheme. From this premise follow a number of results of metrology for different unravelings of Markovian master equations [35–41]. The present thesis assumes this first line of research. An alternative approach preserves the quantum freedom of choosing the measurement to be applied, generalised to any possible feedback and control scheme, via the use of the process tensors or quantum combs formalism [42–45].

Hypothesis testing [15, 46, 47] is the statistical technique of choosing, among a finite set of causes, the optimal one to explain a given outcome. The theory was first developed in the 1930s [48–50]. The quantum

version has mainly taken the form of *quantum state discrimination*: how to optimally determine which state one is given, between two candidate states? [14, 51–53]. The theory is mostly concerned with the determination of the error exponents, i.e. the behaviour of error probability as a function of the number of available copies of the state. A key result has been the quantum generalisation of the Chernoff-Stein bound [54–59]. All these results assume the possibility to optimise over the measurement strategies. On the same line, researchers have tackled the problem of channel discrimination [60]: how can one choose which quantum map (out of a finite set) one is given?

Also for hypothesis testing, a sideline of research has been dedicated to the case of correlated outcomes for fixed measurement scheme. In analogy with the work in estimation theory, error exponents have been proven for unravelings of Markovian master equations [61–66]. When the optimisation over measurement schemes is allowed, error bounds can be found by using quantum combs [67–70].

This thesis is dedicated to the study of the impact of correlations on estimation and hypothesis testing tasks, when only a classical measurement record is available. This is the case, for instance, of the emission record of a system undergoing continuous monitoring. The quantum system itself is, in this sense, inaccessible. Though, can we learn anything about it by looking at its (correlated) outputs? From the perspective of estimation and hypothesis testing, this is a *classical* problem, in the sense that the measurement record available for the estimation or testing task is a sequence of classical random variables. Nonetheless, the quantum nature of the system leaves signatures in the measurement record, that have to be taken into account.

In the thesis, we employ the language of stochastic processes, ordered sequences of random variables. In Chapter I, we first review their mathematical description. Stochastic processes can have structure, specific ways in which memory about the past influences the future. This structure can be described by Hidden Markov models and ϵ-machines. We then introduce the standard framework of estimation theory and hypothesis testing for i.i.d. random variables.

In Chapter II, we focus on the description of quantum dynamics, introducing quantum maps and their representations. We then focus on Markovian quantum evolutions, reviewing the GKSL equation and its unravelings. This concludes the introductory section of the thesis.

As a first step towards metrology on classical outcomes of quantum processes, in Chapter III, we generalise the bounds of i.i.d. estimation theory to processes with finite memory. We give heuristical insight into when correlations allow to learn more about the system, and when, instead, they allow the system to hide more from learning.

In Chapter IV, we apply the results of the previous chapter to processes generated by quantum systems undergoing continuous monitoring, in the form of the quantum jump unravelling (QJU). We start by focussing on renewal processes, where quantum memory is completely lost upon each jump. We give analytical error bounds, both for estimation and hypothesis testing, for multi-channel renewal processes. If, on the other side, memory flows across the jumps, analytical results are much harder to obtain. We analyse in depth the previously introduced formalism of the monitoring operator [35, 38], that allows the computation of the Fisher information on quantum trajectories. We exploit it to introduce a stochastic expression for the Fisher information rate in terms of the stochastic currents, and show its direct connection to maximum likelihood estimation.

Processes with complex memory patterns require adequate numerical strategies for their simulation. In Chapter V, we introduce the quantum Gillespie algorithm for quantum jump trajectories. We apply it to parameter estimation and hypothesis testing problems in quantum optics and atomic physics. We exploit the connection between the Cramér-Rao bound and the kinetic uncertainty relations (KUR) in quantum thermodynamics [71] to refine the quantum KURs by considering correlations among different currents.

In Chapter VI, we take a more information-theoretic approach, introducing the protocol of *collisional learning*. In this framework, only one of the branches of a bipartite quantum map is accessible to an experimenter who, at each timestep, feeds in a new input state and performs a measurement on that branch. The inaccessible part of the map acts as a quantum memory, and influences the outcomes probability distribution. Through this influence, information can be gathered about the structure of the map, including the inaccessible part. We characterise what features of the map can be learnt in this framework, and the emerging gauge degrees of freedom. We also introduce numerically efficient algorithms to reconstruct as much as possible about the map from this collisional measurement record.

In many practical situations, raw data from a system cannot be directly stored and accessed. Rather, the data are post-processed on-the-fly. In Chapter VII, we give precision bounds for common post-processing strategies, the sample mean and the empirical distribution, on correlated data. For QJU measurement records, we show the reduction in Fisher information if some channels cannot be monitored, or distinguished one from each other. Surprisingly, the latter case can result in the emergence of non-renewal behaviour in otherwise renewal systems.

In all chapters, the theoretical treatment is accompanied by a number of physical applications to systems of interest, in quantum optics, and mesoscopic and atomic physics.

Chapter I

Basics: Statistical concepts

In this first Chapter, we review the statistical concepts that will be employed in the thesis. These are fully classical, and no quantum mechanics will appear. The choice to consider *classical* data represents, indeed, a core idea of this thesis. In conventional quantum metrology [8, 9], the input is a quantum state, and the experimenter can optimise over different measurement strategies. On the contrary, we assume here that the measurement strategy is fixed by the structure of the system. Consequently, the input of our estimation tasks is the classical result of the measurement.

This Chapter assumes a basic knowledge of the language of random variables, and generalises it to stochastic processes in Sect. I.1. The main questions are: (a) how can we describe a sequence of non-independent random variables? (b) how can we represent its memory structure in an optimal way?

We then leave the realm of correlated processes for a while in Sect. I.2, focussing on the problem of estimation: given a parameter encoded in a probability distribution, (c) how and (d) with what precision can we recover its value from a sample?

Finally, we present the formalism of hypothesis testing in Sect. I.3. Assume that we have a sample that is guaranteed to have been extracted from one of two probability distributions: (e) how can we optimally choose the correct one, in such a way as to minimise the errors?

I.1. STOCHASTIC PROCESSES

Let us consider a set of N random variables $X_1, \ldots, X_N$, all drawn from the same finite alphabet $\mathcal{A}$, and distributed according to the multivariate probability distribution $\Pr(X_1, \ldots, X_N)$. The probability distribution is normalised according to the law

$$\sum_{x_1,\ldots,x_N} \Pr(x_1, \ldots, x_N) = 1. \tag{I.1}$$

We adopt the convention of uppercase letters for the random variables, and lowercase letters for their realisations. It is convenient to employ a lighter notation for strings of random variables; we write

$$X_{1:N} = X_1, \ldots, X_N. \tag{I.2}$$

Our notation is, therefore, left- and right-inclusive.

A bi-infinite string of random variables $X_{-\infty:+\infty}$ is a *stochastic process*. It is associated with the joint probability distribution of all the random variables, $\Pr(X_{-\infty:+\infty})$. The joint probability of a process contains all the existing information about it. In particular, it encodes all the correlations among the random variables.

A contiguous substring of the random variables of a process is called a *slice*. The probability distribution of a slice of a process can be obtained by *marginalising* the process probability distribution over all the other random variables:

$$\Pr(x_{a:b}) = \sum_{\substack{x_{-\infty:a-1} \\ x_{b+1,+\infty}}} \Pr(x_{-\infty:+\infty}). \tag{I.3}$$

As mentioned above, the joint probability of a stochatic process $\Pr(X_{-\infty:+\infty})$ contains all information on correlations among the random variables. To make the process tractable, we assume it is *stationary* [17]:

$$\Pr(X_{a:b} = x_{a:b}) = \Pr(X_{a+l:b+l} = x_{a:b}) \quad \forall a, b, l \in \mathbb{N}, b > a. \tag{I.4}$$

The intuition behind this equation is that the process is invariant under time translation. Physically, this means that the underlying mechanism generating the process does not change in time.

Given two random variables A and B, the chain rule $\Pr(A, B) = \Pr(B|A)\Pr(A)$ relates the joint, conditional and marginal probabilities. For a slice of a stochastic process, this translates into

$$\Pr(X_{1:N}) = \Pr(X_N|X_{1:N-1})\Pr(X_{N-1}|X_{1:N-2}) \ldots \Pr(X_2|X_1)\Pr(X_1). \tag{I.5}$$

For a fully general process, each outcome is influenced by the entire string of the past outcomes.

While Eq. (I.5) holds in full generality, in many processes of physical interest future outcomes are

only influenced by the most recent outcomes. The simplest example of this kind are independent and identically distributed (i.i.d.) stochastic processes. They are completely memoryless, in the sense that future outcomes do not depend at all on past outcomes: $\Pr(X_N|X_{-\infty:N-1}) = \Pr(X_N)$. Plugging this into the chain rule of Eq. (I.5), the joint probability factorises:

$$\Pr(X_{1:N}) = \Pr(X_1)\Pr(X_2)\dots\Pr(X_N). \tag{I.6}$$

Physically, i.i.d. processes correspond to independently executed identical experiments.

A stationary process is defined to have Markov order $\mathcal{M}$ [72] if

$$\mathcal{M} = \min_{\ell} \{\ell \text{ s.t. } \Pr(X_1|X_{-\ell:0}) = \Pr(X_1|X_{-\infty:0})\}. \tag{I.7}$$

In other words, a stochastic process has Markov order $\mathcal{M}$ if all conditional probabilities can be truncated to at most $\mathcal{M}$ steps in the past, without increasing the uncertainty about the future. I.i.d. processes have $\mathcal{M} = 0$ (this is sometimes called *super-Markovian* behaviour). *Markovian* processes have Markov order $\mathcal{M} = 1$.

If a process has a finite Markov order $\mathcal{M}$, then a random variable X_N is *conditionally* independent of $X_{-\infty:N-\mathcal{M}-1}$, conditioned on the knowledge of $X_{N-\mathcal{M}:N-1}$. It is worth stressing that this does not imply *unconditional independence*. Indeed, unless $\mathcal{M} = 0$, all the random variables in the process are unconditionally correlated. For the simple case of a Markovian process ($\mathcal{M} = 1$) with transition matrix $Q_{ij} = \Pr(X_1 = i|X_0 = j)$, one has

$$\Pr(X_b|X_a) = \left(Q^{b-a}\right)_{X_b X_a} \quad b > a. \tag{I.8}$$

Since Q is a stochastic matrix, all correlations decay exponentially with the distance between the variables [1]. A similar result holds for processes of finite Markov order, as soon as the two considered random variables are farther apart than the Markov order.

In a process of finite Markov order $\mathcal{M}$, only the last $\mathcal{M}$ steps of the past are relevant to predict the future with full accuracy. More in general, let us consider processes in which the future only depends on some *summary statistics* of the past. In other words, there exists a function Φ of the entire semi-infinite

substring of past outcomes $X_{-\infty:0}$ such that

$$\Pr(X_1|X_{-\infty:0}) = \Pr(X_1|\Phi(X_{-\infty:0})). \tag{I.9}$$

In the case of finite Markov order, the function Φ throws away all the information about the past that is less recent than the Markov order, $\Phi(X_{-\infty:0}) = X_{-\mathcal{M}+1:0}$. For processes of infinite Markov order, this is of course not possible! In some cases, though, one can define a function Φ that distills only the *relevant* information about the past. If this distillation is non-trivial, the function Φ is not injective. It induces equivalence classes of pasts that give the same information about the future [73, 74]. We define a *maximal* Φ function ϵ, that groups together all the pasts yielding the same information about the future. The function ϵ induces an equivalence relation between pasts

$$x_{-\infty:0} \sim_\epsilon x'_{-\infty:0} \iff \Pr(X_{1:+\infty}|x_{-\infty:0}) = \Pr(X_{1:+\infty}|x'_{-\infty:0}). \tag{I.10}$$

We refer to these maximal equivalence classes of pasts as *causal states* $\{\sigma\}$. From the equivalence relation of Eq. (I.10), the causal state to which a past belongs has the same informational content about the future as the past itself.

When observing a new outcome, one must know how to update the causal state. We define a transition function $\lambda(\sigma, x) = \sigma'$, for all causal states σ and all outcomes $x \in \mathcal{A}$. It takes as input the current causal state and the observed outcome, and gives the new causal state. Crucially, given the outcome x, the transition between the causal states is deterministic. This property is commonly known as *unifilarity*.

The interplay of causal states and the λ transition function is represented by diagrams known as ϵ-machines, that are the main tool of the discipline known as *computational mechanics* [73–76].

A. Example: the Even Process

The ϵ-machine for the *Even Process* [77, 78] is represented in Fig. I.1. The only free parameter in the machine is $p \in (0, 1)$. It represents the conditional probability of emitting symbol 1 from causal state A. The even process owes its name to the fact that the symbols 0 always have to be emitted in pairs (because no 1 symbol can be emitted from causal state B).

The model is unifilar: for each causal state, there is at most one outgoing arrow per emission symbol. If one fixes the initial causal state, given a sequence of emissions, the evolution on the causal states

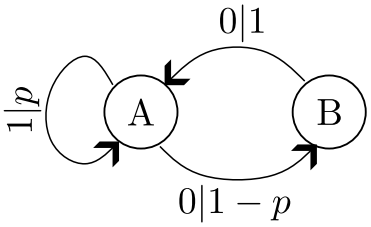

FIG. I.1. Scheme of the ϵ-machine for the Even Process, with the parameter $p \in (0, 1)$. The notation $x|p$ indicates that the transition happens with probability p and is accompanied by the emission of symbol x.

is deterministic. E.g., assume that the system starts in causal state A, and the sequence emitted is 0010000110. Then, the sequence of the causal states is $ABAABABAAAB$.

If the initial causal state is unknown, it can take a number of emissions before one can localise the correct causal state. This is known as the *synchronisation* mechanism [79, 80]. For instance, assume that the same sequence as before is observed, but no information is known about the initial state. Then, there is no way of localising the system until a 1 is observed, because both A and B may emit 0. However, as soon as one sees the emission of a 1, the system synchronises, because the state is A with certainty (a 1 cannot be emitted from causal state B). An arbitrarely long string of symbols can be observed before the synchronisation takes place. However, the probability of the system having synchronised converges asymptotically to 1 for a large number of observed symbols [81].

The very simple structure of the even process hides an infinite Markov order. Indeed, the following truncation rule applies:

$$\Pr(X|000\ldots0001\overleftarrow{x}) = \Pr(X|000\ldots0001), \tag{I.11}$$

where $\overleftarrow{x}$ is any semi-infinite sequence. In fact, the state is completely reset when observing a 1. However, there can be an arbitrarely long string of 0s in between two 1s: one cannot truncate conditionals at a fixed length. It is, however, evident that not all features of the past have an impact on the future probabilities. Indeed, it is sufficient to keep track of the *parity* of the number of zeros that have been observed. In this way, the ϵ-machine distillates only the relevant information about the past.

I.2. ESTIMATION THEORY

In many contexts, one wants to determine the value of a parameter that describes the dynamics of a system. The parameter may not, however, be directly measurable. For instance, in epidemiology, one may want to determine the reproduction rate of a disease [82, 83], but there is no way to obtain such parameter by direct measurement on the pathogenic agent. In finance, one may want to establish the so-called "Greeks" [12] of an asset using the time series of its valuation.

These problems are commonly referred to as *estimation tasks*. In this section, we review the framework of estimation theory [15–17]. We discuss in particular the precision bounds and methods for saturating them.

Let us consider a random variable X, distributed according to a parameter-dependent probability distribution $\Pr_\theta(X)$. The functional dependence of such distribution on θ is precisely known: the exact value of the parameter is the only degree of ignorance. An *estimator* is a function T that takes a sample x from the distribution as input, and outputs an estimate $\hat{\theta} = T(x)$ for the parameter. Of course, one would like this estimate to be close to the *true* value of the parameter (here referred to simply as θ).

We define the *bias* of an estimator T as

$$b = \mathbb{E}_X \left[T(X)\right] - \theta, \tag{I.12}$$

where the expectation value is taken with respect to the distribution of the random variable X. An estimator is *unbiased* if $b = 0$; this is the common framework of estimation theory. The definition of the bias puts this perspective on estimation theory in the frequentist field. Indeed, we assume that the true value θ, though unknown, exists. For an introduction to the alternative approach of Bayesian estimation, see Refs. [84, 85].

Given an unbiased estimator, its precision is measured by the *mean squared error* (MSE)

$$\mathrm{MSE}(T) = \mathbb{E}_X \left[\left(T(X) - \theta\right)^2\right] = \mathrm{Var}_X \left[T(X)\right]. \tag{I.13}$$

Remarkably, for a given parameter dependent probability distribution, it is possible to bound the MSE

of *any* estimator from below:

$$\mathrm{MSE}(T) \geq \frac{1}{F_\theta(X)}. \tag{I.14}$$

This inequality is known as the Cramér-Rao bound [18, 19]. Note that the MSE is only bounded from below: no estimator can do better than the bound, but it is possible to design arbitrarely bad estimators. The quantity in the denominator is the *Fisher information* (contained in the random variable X about the parameter θ), defined as

$$F_\theta(X) = \mathbb{E}_X \left[\left(\frac{\partial}{\partial \theta} \log \Pr_\theta(X) \right)^2 \right]. \tag{I.15}$$

The Fisher information expresses how strongly the probability distribution changes upon a change in the value of θ. This observation allows for an intuitive interpretation of the Cramér-Rao bound. If the probability distribution $\Pr_\theta(X)$ depends strongly on θ, estimating θ is easy. On the contrary, if the dependence is very weak, the estimation is harder, resulting in a higher MSE. The Fisher information can be alternatively defined as

$$F_\theta(X) = -\mathbb{E}_X \left[\frac{\partial^2}{\partial \theta^2} \log \Pr_\theta(X) \right]. \tag{I.16}$$

A proof of the equivalence of the two definitions is given in Ref. [86].

Usually, instead of a single sample x from $\Pr_\theta(X)$, one has a collection of samples $x_{1:N}$, distributed according to $\Pr_\theta(X)$. The estimator T acts therefore on the entire collection $X_{1:N}$. Following from Eq. (I.5), the Fisher information follows the chain rule [86, 87]

$$F_\theta(X_{1:N}) = F_1(\theta) + \sum_{k=2}^{N} F_\theta(X_k | X_{1:k-1}). \tag{I.17}$$

In the i.i.d. case, the joint probability factorises as in Eq. (I.6). The chain rule of Eq. (I.17) reduces to

$$F_\theta(X_{1:N}) = N F_\theta(X_1). \tag{I.18}$$

Correspondingly, the Cramér-Rao bound for i.i.d. random variables reads

$$\mathrm{MSE}(T) \geq \frac{1}{NF_\theta(X_1)}. \tag{I.19}$$

This inverse-linear scaling of the MSE with respect to the number of random variables is referred to as *shot-noise scaling* or *standard quantum limit* [88].

One of the key results of this thesis is the scaling of the Fisher information in the case of finite Markov order processes, see Chapter III [1].

An estimator T is *efficient* if it saturates the Cramér-Rao bound. Usually, efficiency of an estimator is very hard to prove for finite N. If an estimator saturates the Cramér-Rao bound in the limit $N \to \infty$, then it is called *asymptotically efficient*.

A specific estimator, known as *maximum likelihood estimator* (MLE), is always asymptotically efficient, if the probability distribution is regular enough [17, 89]. Let us define the loglikelihood function

$$\ell_x(\tilde{\theta}) = \log \Pr_{\tilde{\theta}}(x). \tag{I.20}$$

For a specific sample x, the loglikelihood function yields the log-probability of such sample, as a function of the parameter $\tilde{\theta}$. The maximum likelihood estimator prescribes $\hat{\theta}$ such that

$$\hat{\theta}_{\mathrm{MLE}} = \mathrm{argmax}_{\tilde{\theta}}\, \ell_x(\tilde{\theta}). \tag{I.21}$$

That is, one chooses the parameter that makes the observed value the most likely. As an example of the application of MLE, we consider a biased coin in the following example.

A. Example: biased coin

Consider a random variable on the alphabet $\mathcal{A} = \{0, 1\}$, with probabilities $\Pr(0) = p$, $\Pr(1) = 1 - p$, $p \in [0, 1]$. We extract N symbols $x_{1:N}$. The probability of the string is

$$\Pr(x_{1:N}) = p^M (1 - p)^{N-M}, \tag{I.22}$$

where M is the number of times the symbol 0 appears in the string $x_{1:N}$. We want to estimate p. The log-likelihood function is

$$\log \ell_{x_{1:N}}(\tilde{p}) = M \log \tilde{p} + (N - M) \log(1 - \tilde{p}). \tag{I.23}$$

We apply the MLE prescription of Eq. (I.21). We maximise the log-likelihood with respect to $\tilde{p}$, by imposing $\partial_{\tilde{p}} \log \ell_{x_{1:N}}(\tilde{p}) \stackrel{!}{=} 0$, to obtain the constraint:

$$0 = \frac{M}{\tilde{p}} - \frac{N - M}{1 - \tilde{p}} \implies \tilde{p} = \frac{M}{N}. \tag{I.24}$$

This corresponds to what we could have expected: we should choose p as the frequency of the outcome 0! The Fisher information for a single outcome is

$$F_p(X) = \frac{1}{p(1-p)}. \tag{I.25}$$

As the tosses are i.i.d., $F_p(X_{1:N}) = NF_p(X)$. The MSE of MLE is just the variance of the binomial distribution

$$\mathbb{E}\left[\left(\frac{M}{N} - p\right)^2\right] = \frac{1}{N^2}\mathrm{Var}\left[M\right] = \frac{p(1-p)}{N}. \tag{I.26}$$

As expected, MLE saturates the Cramér-Rao bound.

This concludes our review of the basics of classical estimation theory. In Chapter III, we will generalise these results for processes with a finite Markov order.

I.3. HYPOTHESIS TESTING

A related but not equivalent problem to estimation theory is hypothesis testing. In this section, we review the problem, and the standard methods to approach it.

In a *simple hypothesis testing* problem [17], one has to choose between two hypotheses, to determine which one better explains the observed data. Like estimation, hypothesis testing is ubiquitous far beyond physics, across the social and natural sciences.

Consider a random variable X, that is distributed according to one of two probability distributions

(*hypotheses*) $\mathcal{H}_0$ and $\mathcal{H}_1$. One wants to explain from which one of the hypotheses a sample x has been extracted. A *tester* is a function $\mathcal{I}(x)$ that, taking as input the sample, retrieves the chosen hypothesis. In the following, we refer to $\mathcal{H}_0$ as the *null hypothesis*, while $\mathcal{H}_1$ is the *alternative hypothesis* [15]. The meaning of the names will become clear shortly.

When applying the tester function on a sample x, four possible situations can happen, represented in Table I.1. One would like to minimise type-1 and type-2 errors at the same time. This is, however, not

		$\mathcal{I}(x)$	
		$\mathcal{H}_0$	$\mathcal{H}_1$
True	$\mathcal{H}_0$	correct	type 1 error (false alarm)
	$\mathcal{H}_1$	type 2 error (miss)	correct

TABLE I.1. The four situations that can verify in hypothis testing tasks.

possible, and one has to choose a trade-off balance between them. Some formalisms are *symmetric*, in the sense that type 1 and type 2 errors are considered on an even basis. Here, we consider the well-known *asymmetric* formalism of Neyman and Pearson [15, 46–48]. This formalism considers a false alarm error to be much more dangerous than a miss. Therefore, one fixes the probability of false alarm $P_{\text{FA}} = \Pr(\mathcal{I}(x) = \mathcal{H}_1|\mathcal{H}_0)$ to some low value, and then minimises the probability of a miss $P_{\text{M}} = \Pr(\mathcal{I}(x) = \mathcal{H}_1|\mathcal{H}_0)$ for fixed P_{FA}. The Neyman-Pearson formalism gives an optimal test $\mathcal{I}(x)$, and allows to bound P_{M} as a function of P_{FA}.

To introduce the working principle of the Neyman-Pearson test, we start from the simple case of Gaussian hypothesis testing [15, 66]. Let us assume a real random variable X, sampled either from a normal distribution $\mathcal{H}_0 = \mathcal{N}(\mu_0, \sigma_0^2)$, or from another normal distribution $\mathcal{H}_1 = \mathcal{N}(\mu_1, \sigma_1^2)$. We assume $\mu_0 < \mu_1$. Given a sample x, we define the test function:

$$\mathcal{I}(x) = \begin{cases} \mathcal{H}_0 & \text{if } x \leq \gamma \\ \\ \mathcal{H}_1 & \text{if } x > \gamma \end{cases}, \tag{I.27}$$

where γ is a threshold value. The situation is represented in Fig. I.2. In the Neyman-Pearson formalism, we first fix the value of the false alarm probability P_{FA}, corresponding to the area under the green curve in Fig. I.2. We have:

$$P_{\text{FA}}(\gamma) = Q\left(\frac{\gamma - \mu_0}{\sigma_0}\right), \tag{I.28}$$

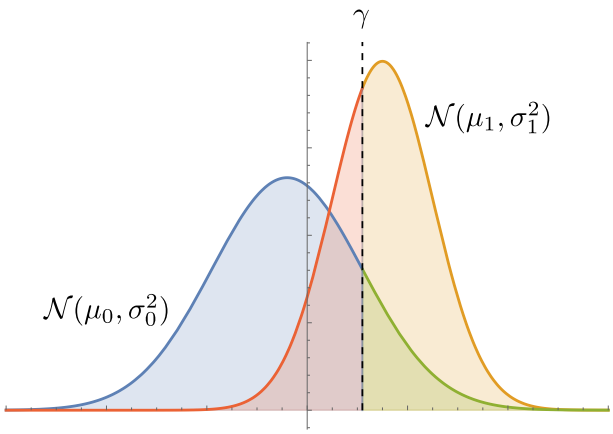

FIG. I.2. Schematic of the Gaussian hypothesis testing scenario. The regions are coloured as follows. (blue) correctly identified as $\mathcal{H}_0$; (red) miss error; (green) false alarm error; (orange) correctly identified as $\mathcal{H}_1$.

where $Q(\bullet) = \frac{1}{2}(1 - \mathrm{erf}(\bullet))$. We fix

$$\gamma = \mu_0 + \sigma_0 Q^{-1}(P_{\mathrm{FA}}). \tag{I.29}$$

At this point, the missed detection probability P_{M} is given by the area under the red curve, for the value of γ fixed by the choice of P_{FA}. We get

$$P_{\mathrm{M}} = 1 - Q\left(\frac{\mu_0 - \mu_1}{\sigma_1} + \frac{\sigma_0}{\sigma_1} Q^{-1}(P_{\mathrm{FA}})\right) \tag{I.30}$$

for the missed detection error. Note that a smaller P_{FA} corresponds to moving the threshold γ to higher values. Correspondingly, the missed detection probability P_{M} increases.

Let us now consider hypothesis testing for general i.i.d. random variables. All the variables of the sample $x_{1:N}$ are extracted from the same probability distribution (either $\mathcal{H}_0$ or $\mathcal{H}_1$).

Let us define the log-likelihood ratio $L(x_{1:N})$ as

$$L(x_{1:N}) = \log \frac{\mathrm{Pr}_1(x_{1:N})}{\mathrm{Pr}_0(x_{1:N})}, \tag{I.31}$$

where $\mathrm{Pr}_j(x)$ is shorthand for the probability of outcome x, according to the hypothesis $\mathcal{H}_j$. The Neyman-Pearson theorem [15, 48] states that the optimal test function $\mathcal{I}(x_{1:N})$, i.e. the one that minimises P_{M}

for fixed P_{FA}, is given by

$$\mathcal{I}(x_{1:N}) = \begin{cases} \mathcal{H}_1 & \text{if } L(x) > \gamma \\ \mathcal{H}_0 & \text{otherwise} \end{cases}, \tag{I.32}$$

for a threshold γ satisfying the equation

$$P_{\mathrm{FA}} = \int_{x_{1:N} \text{ s.t. } L(x)>\gamma} dx_{1:N} \Pr_0(x_{1:N}). \tag{I.33}$$

The longer the string of samples $x_{1:N}$, the more precise the choice of the appropriate hypothesis will be. The Chernoff-Stein lemma [17] yields the asymptotic behaviour of P_{M}:

$$\lim_{N\to\infty} \frac{1}{N} \log P_{\mathrm{M}} = -D_{\mathrm{KL}} \left[\Pr_1(X_{1:N}) || \Pr_0(X_{1:N})\right], \tag{I.34}$$

where D_{KL} is the Kullback-Leibler divergence, or relative entropy, between the two probability distributions, defined as

$$D_{\mathrm{KL}}(\Pr_1(X)||\Pr_2(X)) = \sum_x \Pr_1(x) \log \frac{\Pr_1(x)}{\Pr_2(x)}. \tag{I.35}$$

The Chernoff-Stein lemma has a simple interpretation: the more similar the two candidate distributions are, the harder the hypothesis testing task is, and the slower P_{M} decays.

Chapter II

Basics: quantum dynamics

In this Chapter, we revise the concepts needed to describe the evolution of open quantum systems. We start by reviewing completely positive, trace preserving (CPTP) quantum maps [90–95] as the most general evolution of a quantum system. We also present three useful representations of quantum maps, that will be extensively used in the thesis.

GKSL master equations [96–99] yield a single parameter family of CPTP maps, fulfilling the semigroup requirement. They represent the Markovian evolution of open quantum systems, weakly coupled to an environment. They naturally lead to quantum unravellings, which allow to describe continuously monitored quantum systems.

II.1. CPTP MAPS AND THEIR REPRESENTATIONS

The most general way of representing the state of a quantum system is via a *density matrix* $\rho \in \mathcal{S}(\mathcal{H})$, a positive semidefinite operator on the Hilbert space $\mathcal{H}$ of the system, with trace $\mathrm{Tr}[\rho] = 1$. In the following, we usually omit the specification of the Hilbert space on which $\mathcal{S}$ is defined.

A CPTP map is a function between two spaces of states $\mathcal{E} : \mathcal{S}_A \to \mathcal{S}_B$, with the following properties:

- It is linear: $\mathcal{E}(\alpha\rho + \beta\sigma) = \alpha\mathcal{E}(\rho) + \beta\mathcal{E}(\sigma)$, for all $\alpha, \beta \in \mathbb{C}$, $\rho, \sigma \in \mathcal{S}$;
- It preserves the trace: $\mathrm{Tr}[\mathcal{E}(\rho)] = \mathrm{Tr}[\rho]$;
- It is positive: $\rho \geq 0 \implies \mathcal{E}(\rho) \geq 0$ for all $\rho \in \mathcal{S}_A$;
- It is completely positive: for any possible ancillary space $\mathcal{S}_E$, the map $\mathcal{E} \otimes \mathbb{I}_E : \mathcal{S}_A \otimes \mathcal{S}_E \to \mathcal{S}_B \otimes \mathcal{S}_E$ is positive, where $\mathbb{I}_E$ is the identity map on the space $\mathcal{S}_E$.

CPTP maps are also known as *quantum channels*. In this thesis, we make use of different equivalent representations of CPTP maps.

A. The Kraus representation

Let $\mathcal{E} : \mathcal{S}_A \to \mathcal{S}_B$ be a CPTP map. Then there exists a collection of operators $\{K_j\}_{j=1}^{L}$, $K_j : \mathcal{H}_A \to \mathcal{H}_B$, such that the action of the map can be equivalently written as

$$\mathcal{E}(\rho) = \sum_{j=1}^{L} K_j \rho K_j^\dagger. \tag{II.1}$$

This is known as *operator-sum* or Kraus representation [92, 95]. The trace preservation requirement corresponds to the normalisation

$$\sum_{j=1}^{L} K_j^\dagger K_j = \mathbb{I}. \tag{II.2}$$

The Kraus decomposition of a map is not unique. The minimal number of Kraus operators required to describe a map is the *rank* of the map. We remark that [100]:

- A CPTP map can have rank 1 if and only if it is unitary. For rank-1 maps, the condition of Eq. (II.2) reduces to the definition of unitary operators.

- The maximum rank of a CPTP map is equal to the product of the dimensions of the origin and destination Hilbert spaces.

Oftentimes, the Kraus operators can be traced to individual physical effects. For example, in the modelisation of noise, each of them may correspond to a different noise mechanism [95].

The Kraus representation also connects naturally to quantum trajectories [99, 101], that will be introduced later in this chapter. The application of a channel $\mathcal{E}$, with Kraus operators $\{K_j\}_{j=1}^{L}$, can be equivalently written as

$$\mathcal{E}(\rho) = \sum_{j=1}^{L} K_j \rho K_j^\dagger = \sum_{j=1}^{L} \mathrm{Tr}\left[K_j \rho K_j^\dagger\right] \frac{K_j \rho K_j^\dagger}{\mathrm{Tr}\left[K_j \rho K_j^\dagger\right]} = \mathbb{E}_j \frac{K_j \rho K_j^\dagger}{\mathrm{Tr}\left[K_j \rho K_j^\dagger\right]}, \tag{II.3}$$

where the expectation value is taken with respect to the probability distribution $\Pr(j) = \mathrm{Tr}\left[K_j \rho K_j^\dagger\right]$. The application of the map corresponds, therefore, to the average over all the possible effects (individual Kraus operators), with the appropriate probability.

B. The Choi representation

A useful representation of quantum channels is given by the Choi-Jamiolkowski isomorphism [102–104]. Let us consider the action of a map $\mathcal{E} : \mathcal{S}_A \to \mathcal{S}_B$ on half of the unnormalised maximally entangled state on $\mathcal{S}_A \otimes \mathcal{S}_A$ [105]:

$$\Upsilon_{\mathcal{E}} = (\mathcal{E} \otimes \mathbb{I}) \left|\Phi^+\right\rangle \left\langle\Phi^+\right| = \sum_{ij} \mathcal{E}(|i\rangle \langle j|) \otimes |i\rangle \langle j| , \tag{II.4}$$

where $|\Phi^+\rangle = \sum_j |j\rangle \otimes |j\rangle$ for a basis $\{|j\rangle\}$ of the input space. The matrix $\Upsilon_{\mathcal{E}}$ lives in the space $\mathcal{S}_B \otimes \mathcal{S}_A$, as $\mathcal{E}$ only acts on the first arm of the maximally entangled state. $\Upsilon_{\mathcal{E}}$ is the *Choi state* of the map. The action of the map $\mathcal{E}$ on the state ρ is given by

$$\mathcal{E}(\rho) = \mathrm{Tr}_A \left[\left(\mathbb{I} \otimes \rho^T \right) \Upsilon_{\mathcal{E}} \right] . \tag{II.5}$$

While the rule for the action of the Choi state on ρ of Eq. (II.5) may seem cumbersome, the Choi representation has extremely neat analytical properties. Indeed, as proven in Ref. [105]:

- The map $\mathcal{E}$ is completely positive if and only if $\Upsilon_{\mathcal{E}}$ is positive;
- The map $\mathcal{E}$ is trace preserving if and only if $\mathrm{Tr}_B [\Upsilon_{\mathcal{E}}] = \mathbb{I}$.

The concatenation of two channels $\mathcal{E} : \mathcal{S}_A \to \mathcal{S}_B$, and $\mathcal{F} : \mathcal{S}_B \to \mathcal{S}_C$ is a new map $\mathcal{F} \circ \mathcal{E} : \mathcal{S}_A \to \mathcal{S}_C$. Let $\Upsilon_{\mathcal{E}}$ and $\Upsilon_{\mathcal{F}}$ be the corresponding Choi states. The concatenation in terms of the Choi states is given by the *link product* rule [106]

$$\Upsilon_{\mathcal{F}\circ\mathcal{E}} = \Upsilon_{\mathcal{F}} \star \Upsilon_{\mathcal{E}} = \mathrm{Tr}_B \left[\left(\mathbb{I}_C \otimes \Upsilon_{\mathcal{E}}^{T_B} \right) \left(\Upsilon_{\mathcal{F}} \otimes \mathbb{I}_A \right) \right] , \tag{II.6}$$

where the superscript T_B denotes partial transposition on subsystem B. As $\Upsilon_{\mathcal{E}}$ lives in the space $\mathcal{S}_B \otimes \mathcal{S}_A$ and $\Upsilon_{\mathcal{F}}$ in $\mathcal{S}_C \otimes \mathcal{S}_B$, the composition Choi state $\Upsilon_{\mathcal{F}\circ\mathcal{E}}$ belongs to $\mathcal{S}_C \otimes \mathcal{S}_A$. This rule can be generalised for a more complex structure of maps with different input and output spaces [106].

Due to the neat characterisation of CPTP maps, the Choi formalism is in widespread use for the analytical study of the structural properties of quantum maps, and is at the base of the language of *quantum combs*. These mathematical objects describe the most general evolution of quantum systems, allowing for (beyond-Markovian) memory, measurements, and feedback [67, 105]. The exposition of

quantum combs lies outside of the scope of the present thesis. However, they constitute the natural language to express correlated quantum processes, whenever the experimenter can optimise over the measurement schemes.

C. The Liouville representation

While the Choi representation gives clear analytical insight into CP- and TP-ness of maps, its complicated composition rule Eq. (II.6) makes it less amenable to numerical use. An alternative approach is given by the Liouville representation [107, 108]. Assume a state $\rho \in \mathcal{S}$ has matrix elements $\rho = \sum_{ij} \rho_{ij} |i\rangle \langle j|$ on a basis $\{|i\rangle\}$ of the underlying Hilbert space. Define

$$|\rho\rangle\rangle = \sum_{ij} \rho_{ij} |i\rangle \otimes |j\rangle . \tag{II.7}$$

As mixed states are mapped to vectors in a larger Hilbert space, quantum maps are mapped to matrices on the same Hilbert space. We define a *vectorisation* function vec, mapping states (ρ) to vectors ($|\rho\rangle\rangle$). Given three matrices A, B, C, of such dimensions that their product is well defined, the vectorisation function has the following important property [107]:

$$\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B). \tag{II.8}$$

This allows us to write the Liouville representation of a CPTP map from its Kraus operators $\{K_j\}$:

$$\mathrm{vec}(\mathcal{E}(\rho)) = \mathrm{vec}\left(\sum_j K_j \rho K_j^\dagger\right) = \left(\sum_j K_j^* \otimes K_j\right) |\rho\rangle\rangle \implies \tilde{\mathcal{E}} = \sum_j K_j^* \otimes K_j. \tag{II.9}$$

The main strength of the Liouville representation of quantum maps is the very straightforward composition rule: the Liouville representation of the concatenation of two maps is just the matrix product of their Liouville representations (on the same basis). This makes this representation particularly useful for numerical calculations. On the other hand, there is no simple way of writing the complete positivity constraint in terms of properties of the Liouville matrix $\tilde{\mathcal{E}}$.

II.2. THE GKSL MASTER EQUATION AND ITS UNRAVELINGS

The description of quantum evolution we presented so far is completely general and all-encompassing. Let us now focus on a continuous-parameter family of quantum channels, that represents the evolution of an open quantum system weakly coupled with a Markovian environment.

We partition the universe into a part that is under experimental control, the *system*, and a part that is not, the *environment*. We assume that the interactions between system and environment are weak, and that the environment does not carry memory about its interaction with the system. This is equivalent to saying that the environment dissipates any change induced by the interaction with the system on a much faster timescale than the typical timescale of the system's evolution. Under these conditions, one can write the evolution rule for the state ρ using the Gorini-Kossakowski-Sudarshan-Lindblad master equation [96–98]:

$$\frac{d}{dt}\rho = \mathcal{L}\rho = -i[H,\rho] + \sum_k \gamma_k \left(L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\} \right). \tag{II.10}$$

The map $\mathcal{L}$ is the Lindbladian superoperator. Superoperators are linear operators on the space of operators. In the following, we indicate superoperators with calligraphic notation. The first term of the sum represents the coherent evolution of the system, under a Hamiltonian H, while the sum over k represents the dissipative, irreversible evolution. The operators L_k are known as *collapse* or *jump* operators, for reasons that will become clear in the following. The parameters γ_k are positive real numbers, known as *jump rates*.

The one-parameter family of quantum channels that solves the GKSL equation for all times t maps the state $\rho(0)$ to its time-evolved version $\rho(t)$. By formal integration of Eq. (II.10), it has the form $\Phi(t) = e^{\mathcal{L}t}$. In reality, solving the GKSL equation analytically is usually very hard, and numerical methods are employed. We will present a new method in Chapter V.

We focus on the infinitesimal time evolution:

$$\begin{aligned} &\Phi(dt) = e^{\mathcal{L}dt} \sim \mathbb{I} + \mathcal{L}dt + \mathcal{O}(dt^2) = \\ &= \mathbb{I} - iH\rho dt + i\rho H dt + \\ &+ \sum_k \gamma_k L_k \rho L_k^\dagger dt - \frac{1}{2}\sum_k \gamma_k L_k^\dagger L_k \rho dt - \frac{1}{2}\sum_k \gamma_k \rho L_k^\dagger L_k dt + \mathcal{O}(dt^2). \end{aligned} \tag{II.11}$$

We can write this equation, up to order dt, in terms of the Kraus operators:

$$M_0 = \mathbb{I} - iHdt - \frac{1}{2}\sum_k \gamma_k L_k^\dagger L_k dt \quad ; \quad M_k = \sqrt{\gamma_k} L_k \sqrt{dt}. \tag{II.12}$$

They satisfy the normalisation condition $\sum_j M_j^\dagger M_j = \mathbb{I} + \mathcal{O}(dt^2)$, up to first order in dt.

Let us now apply the same logic of Eq. (II.3) to $\{M_j\}$. In particular:

$$\Phi(dt)(\rho(0)) = \mathbb{E}_j \frac{M_j \rho(0) M_j^\dagger}{\mathrm{Tr}\left[M_j \rho M_j^\dagger\right]}. \tag{II.13}$$

The evolution for a finite amount of time Δt is given by the repeated application (say, N times) of the infinitesimal evolution[1]. We then get:

$$\Phi(\Delta t)(\rho(0)) = \mathbb{E}_{\omega_{1:N}} \frac{M_{\omega_N} \dots M_{\omega_1} \rho(0) M_{\omega_1}^\dagger \dots M_{\omega_N}^\dagger}{\mathrm{Tr}\left[M_{\omega_N} \dots M_{\omega_1} \rho(0) M_{\omega_1}^\dagger \dots M_{\omega_N}^\dagger\right]}. \tag{II.14}$$

The evolution of a quantum state according to the GKSL equation is, therefore, given by the average over repeated, stochastic applications of Kraus operators. Each individual sequence of Kraus operators in the expectation value forms a *quantum trajectory*.

Up to this point, quantum trajectories may appear a purely mathematical artefact, following from the Kraus representation of CPTP maps. In this thesis, we use quantum trajectories to describe continuously monitored quantum systems. In particular, we associate a *detection event*, a click, to each collapse operator L_k. We then interpret the evolution along a single trajectory as a sequence of no-jump infinitesimal steps (corresponding to the application of the Kraus operator M_0), interspersed by jump steps (corresponding to the application of any other Kraus operator). This is an accurate description of the quantum jumps observed in experiments [109] on single ions [110–112], superconducting qubits [113, 114], single atoms [115], and cavity QED [116, 117].

[1] This property is, in reality, far from trivial, and is known as the *semigroup* structure of the GKSL evolution. See for instance [98].

Chapter III

Estimation theory for stochastic processes

In this Chapter, we generalise parameter estimation to temporally correlated stochastic processes with finite memory. The main results are the expression for the Fisher information for systems with finite Markov order, Eq. (III.7), and the heuristics on the impact of the correlations sign. As an illustration of the latter, we apply the results to temperature estimation on an Ising spin chain.

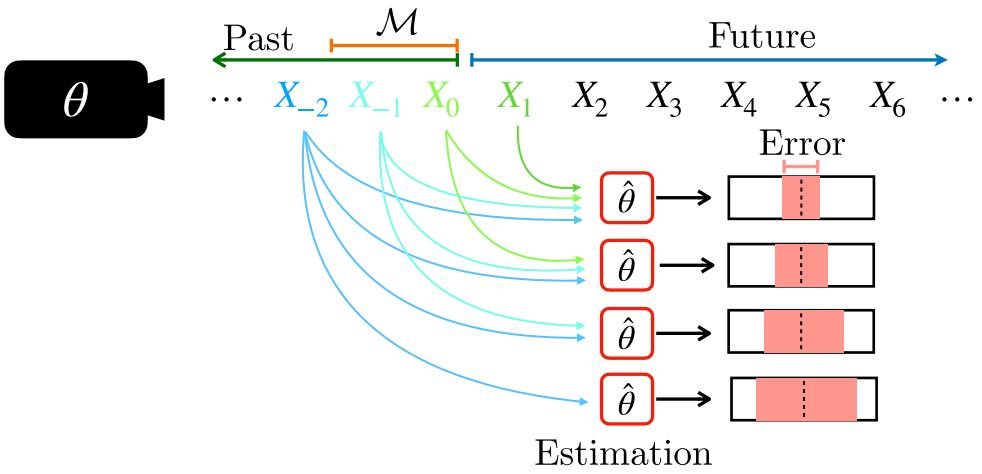

FIG. III.1. Scheme of the correlated estimation setup: a black box outputs outcomes with a probability depending on an unknown parameter θ. The outcomes are not independent: in this case, they have Markov order 2. Reproduced from [1].

In Sect. I.2, we introduced the formalism of estimation theory, and in particular the Cramér-Rao bound (I.14) and the Fisher information (I.15), that bound the precision of an estimation task. Oftentimes, in estimation problems one assumes that the considered random variables are independent and identically distributed (i.i.d.). In this case, the probability factorises in the form (I.6), and the Fisher information is additive. If we want to estimate parameters encoded in a process, however, we need to go beyond the i.i.d. hypothesis. For example, in the following chapters we will want to determine a parameter of the GKSL equation (II.10) of a continuously monitored syste from the sequence of measurements outcomes. These outcomes are correlated, because the state of the system keeps changing all along the evolution, as a result both of its intrinsic dynamics and measurement backaction. We want to answer the following questions: (a) how does the Fisher information behave in a correlated process? (b) when is it advantageous to estimate along a stochastic process, with respect to only considering its single-time

marginals?

References and collaboration information

The main reference for this chapter is the work [1], written in collaboration with G.T. Landi, K. Modi and F.C. Binder. As the first author of the paper, I was responsible for most of the analytical calculations and the simulations, as well as for writing the first version of the manuscript. Sect. III.5 refers to work carried out by P. Riechers in Ref. [118]. The results there go beyond [1]. I had the occasion to discuss the results with the Author before the publication, as acknowledged in the paper itself.

III.1. THE FISHER INFORMATION IS NOT SUBADDITIVE

We start by observing that the Fisher information is not always subadditive. Nothing can be said, in general, about the inequality

$$F_\theta(X_{1:N}) \overset{?}{\leqslant} F_\theta(X_1) + F_\theta(X_2) + \ldots + F_\theta(X_N). \tag{III.1}$$

This fact sets the Fisher information aside from other entropy-like quantities, which are always convex [17]. Consequently, the answer to question (b) above will be process-dependent. The following example illustrates the additivity of the Fisher information for two Gaussian variables.

A. Example: two Gaussian variables

Let us consider two Gaussian random variables X, Y, with the same mean μ and covariance matrix

$$\sigma = \gamma_0 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{III.2}$$

where $\rho \in [-1, 1]$ and $\gamma_0 > 0$. We set to estimate μ. The Fisher information for a generic parameter θ for Gaussian random variables is [15]

$$F_\theta(X, Y) = (\partial_\theta \vec{\mu})^T \, \sigma^{-1} \, (\partial_\theta \vec{\mu}) + \mathrm{Tr}\left[\sigma^{-1} \, (\partial_\theta \sigma) \, \sigma^{-1} \, (\partial_\theta \sigma)\right]/2, \tag{III.3}$$

where $\vec{\mu} = (\mu, \mu)$ in our case. If $\theta = \mu$, the joint Fisher information is

$$F_\mu(X, Y) = \frac{1}{\gamma_0} \frac{2}{1+\rho}. \tag{III.4}$$

The marginal Fisher information is, instead

$$F_\mu(X) = F_\mu(Y) = \frac{1}{\gamma_0}. \tag{III.5}$$

We get a super-additive behaviour $F_\mu(X, Y) > F_\mu(X) + F_\mu(Y)$ for anticorrelated variables ($\rho < 0$), but a sub-additive behaviour if the correlation is positive ($\rho > 0$). This behaviour is not an isolated case for Gaussian variables. On the contrary, we present heuristic arguments in the following connecting the sign of the correlations and the additivity of the Fisher information.

The inequality (or lack thereof) in Eq. (III.1) can be rather surprising: how can the marginals of a process contain *more* information about a parameter in the process, than the full joint probability? To illustrate this point, let us consider an extreme case. Assume that we toss a biased coin (with probability of heads θ) only once, and after the toss the coin remains stuck in its final state; our aim is to determine θ. The stochastic process generated by repeatedly observing the coin is pathological: once an observation is made, the following outcomes are completely fixed. The conditional probability is independent on θ. In terms of the chain rule for the Fisher information of Eq. (I.17), all conditional Fisher information terms are therefore zero. Hence, $F_\theta(X_{1:N}) = F_\theta(X_1)$. This gives the left hand side of Eq. (III.1). Let us now focus on the right hand side. In this case, one has to sample the random variables i.i.d. from the process. This means that one has to throw away the coin after the first toss, and the fact that it remains stuck is immaterial. Of course, in this second case the information collected about θ is larger. It is important, however, to keep in mind that the mechanism with which the random variables are generated in the two cases is different.

III.2. FISHER INFORMATION FOR FINITE MARKOV ORDER

Let us consider a stationary process, according to the definition of Eq. (I.4), with finite Markov order $\mathcal{M}$, see Eq. (I.7). For the joint Fisher information of N random variables, the chain rule (I.17) yields

$$F_\theta(X_{1:N}) = F_\theta(X_1) + \sum_{k=2}^{N} F_\theta(X_k|X_{1:k-1}). \tag{I.17 revisited}$$

By definition of the Markov order, we can truncate all conditional probabilities to at most $\mathcal{M}$ steps in the past. The conditional Fisher information can be truncated in the same way. We have, therefore:

$$F_\theta(X_{1:N}) = F_\theta(X_1) + \sum_{k=2}^{\mathcal{M}+1} F_\theta(X_k|X_{1:k-1}) + \sum_{k=\mathcal{M}+2}^{N} F_\theta(X_k|X_{k-\mathcal{M}:k-1}). \tag{III.6}$$

All the contributions in second sum are exactly the same, by stationarity of the process. The first sum can be reabsorbed as the joint Fisher information of the first $\mathcal{M}$ random variables:

$$F_\theta(X_{1:N}) = F_\theta(X_{1:\mathcal{M}}) + (N-\mathcal{M})F_\theta(X_{\mathcal{M}+1}|X_{1:\mathcal{M}}). \tag{III.7}$$

This is the first main result of Ref. [1]. We now explore its implications in terms of estimation performance. We can rewrite Eq. (III.7) as

$$F_\theta(X_{1:N}) = Nf(\theta) + \epsilon(\theta), \tag{III.8}$$

by defining the Fisher information rate $f(\theta)$ as

$$f(\theta) = \lim_{N\to\infty} \frac{1}{N} F_\theta(X_{1:N}) = F_\theta(X_{\mathcal{M}+1}|X_{1:\mathcal{M}}), \tag{III.9}$$

and the excess Fisher information $\epsilon(\theta)$ as

$$\epsilon(\theta) = F_{1:N}(\theta) - Nf(\theta). \tag{III.10}$$

The Cramér-Rao bound (I.14) yields

$$\mathrm{MSE}(T(X_{1:N})) = \frac{1}{F_\theta(X_{1:N})} = \frac{1}{Nf(\theta) + \epsilon(\theta)} \sim \frac{1}{Nf(\theta)}, \tag{III.11}$$

where the last approximation is taken in the large N limit. It is instructive to compare the large-N expression for the correlated Cramér-Rao bound of Eq. (III.11) with the same bound for i.i.d. random variables of Eq. (I.19). In both cases, the precision increases linearly with N. The rate, however, is different. As discussed above, nothing can be said in general about which case gives a higher precision. Though, we present heuristical arguments in the following, to connect the sign of the correlations with their contribution to estimation.

III.3. ESTIMATORS BASED ON THE SAMPLE MEAN

We now focus on a specific subclass of estimators, the ones based on the sample mean. We show that, for this subclass, there is a direct connection between the sign of the correlations and the addivity of the Fisher information.

Given a process $X_{1:N}$, its sample mean is the random variable

$$Y = \frac{X_1 + \dots X_N}{N}. \tag{III.12}$$

The sample mean is a *summary statistic* for the process. This means that there is always a way to uniquely map a process $X_{1:N}$ into its sample mean Y, but the converse may not hold true. We consider estimators operating on the sample mean Y. Of course, these estimators are, in general, less precise than estimators that consider entire strings. Let us call $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$, respectively, the expectation value and the variance of the individual random variables. The process being stationary, we do not need to specify the indices. For the same reason, the covariance $\mathrm{Cov}(X_i, X_j)$ is only a function of the distance $i - j$. Let us define $C_{i-j} = \mathrm{Cov}[X_i, X_j]$. Two variables X_i and X_j are *positively correlated* if $C_{i-j} > 0$, *negatively correlated* otherwise.

The central limit theorem (whose results can be extended to the correlated case under suitable conditions [119]) guarantees that Y converges in distribution to a normal distribution, with expectation value

μ, and variance

$$\nu = \frac{1}{N}\sigma^2 + \frac{2}{N^2}\sum_{i,j>i}\mathrm{Cov}[X_i, X_j]. \tag{III.13}$$

For the second term of the right-hand side, we note that

$$\sum_{i,j>i}\mathrm{Cov}[X_i, X_j] = \sum_{i=0}^{N-1}\sum_{j=i+1}^{N-1} C_{i-j} = \sum_{i=1}^{N-1}(N-i)C_i, \tag{III.14}$$

hence

$$\nu = \frac{1}{N}\left[\sigma^2 + 2\sum_{i=1}^{N-1} C_i\right]. \tag{III.15}$$

The Fisher information contained in the sample mean Y is the Fisher information of a normal distribution as

$$F_\theta(Y) = N\left[\frac{1}{\nu}\left(\frac{\partial}{\partial\theta}\mu\right)^2 + \frac{1}{2N}\frac{1}{\nu^2}\left(\frac{\partial}{\partial\theta}\nu\right)^2\right]. \tag{III.16}$$

As the second term vanishes for $N \to \infty$, we are left with the asymptotic expression [120]

$$F_\theta(Y) \sim N\left(\frac{\partial\mu}{\partial\theta}\right)^2 \frac{1}{\sigma^2 + 2\sum_{i=1}^{N-1} C_i}. \tag{III.17}$$

This expression makes the role of correlations evident:

- positively correlated variables ($C_i > 0$) increase the denominator, yielding a lower Fisher information;
- negatively correlated variables ($C_i < 0$) the opposite happens. This results in an overall higher Fisher information.

It is interesting to see the same result from an intuitive perspective. Let us decompose the random variables of the process in terms of fluctuations around their mean, $X_i = \mu + \delta X_i$. The sample mean is then

$$Y = \mu + \frac{\delta X_1 + \ldots + \delta X_N}{N}. \tag{III.18}$$

Say now that two variables in the process, X_i and X_j, are negatively correlated. If one of them has a positive oscillation above the mean, $\delta X_i > 0$, then there is a tendence to have $\delta X_j < 0$. Consequently, the errors induced by the two variables tend to cancel out. On the contrary, for positive correlations the errors tend to add up.

These observations hint at a general tendency: negative correlations are better for estimation than positive correlations. However, we underline that, formally speaking, this argument only hold for estimators based on the sample mean. Since these estimators are a subset of the general set of estimators, the Fisher information of the sample mean lower-bounds the Fisher information for the general set:

$$F_\theta(Y) \leq F_\theta(X_{1:N}). \tag{III.19}$$

In the next section, we explore the application of correlated estimation theory to thermometry on classical spin chains. While there are, in that case, small violations of this relation between estimation advantage and sign of the correlations, we show that it is, indeed, a good rule of thumb.

III.4. APPLICATION: SPIN CHAIN THERMOMETRY

As an example of correlated estimation, we consider thermometry on an Ising spin chain. This model, first introduced by Ising [121] is widely studied in statistical mechanics [122].

Let us consider a chain of N spins $\{x_1, \ldots, x_N\}$, each taking values ± 1, with (classical) Hamiltonian

$$H(x_1, \ldots, x_N) = -B \sum_j x_j - \sum_k J_k \sum_j x_j x_{j+k}, \tag{III.20}$$

where B is an external magnetic field), and J_k is the coupling strength between two spins at distance k from each other.

The system is prepared in a thermal state at temperature T, yielding a probability distribution

$$\Pr(X_1, \ldots, X_N) = \frac{1}{Z} \exp\left[-\frac{1}{T} H(X_1, \ldots, X_N)\right]. \tag{III.21}$$

We consider a spin chain as a stochastic process, in which the experimenter measures the spins one at a time, sequentially, from the first to the last. The experimenter is interested in estimating the temperature T. The process is correlated, due to the presence of the J_k coupling.

Intuitively, one may see that the Markov order of the process is exactly the same as the largest k for which $J_k \neq 0$ (the *interaction range* R of the chain). It has been proven in full generality [123] that $R \geq \mathcal{M}$. However, whether equality holds or not is unclear [124] in the general case, but known for $\mathcal{M} = R = 1$ [1]. This is a good spot to highlight again that the Markov order indicates *conditional* independence between past and future, but not *unconditional* independence. Indeed, spins on the chain are all unconditionally dependent. However, when considering the chain as a stochastic process, the previous spin contains as much conditional knowledge about the future as the entirety of the past.

Let us focus on the case $R = 1$, $J_k = 0$ for $k > 1$, that is the typical nearest-neighbors Ising chain. The Fisher information scales according to Eq. (III.7), with $\mathcal{M} = 1$. The Fisher information rate is $f(T) = F_T(X_2|X_1)$. Can we say anything about the inequality in Eq. (III.1) for temperature estimation on a spin chain?

First of all, let us clarify the physical meaning of $F_T(X_{1:N})$, as compared to $NF_T(X_1)$ in the case of spin chain thermometry. If the access to the chain is sequential, then the precision in temperature estimation is bounded by the correlated Fisher information $F_T(X_{1:N})$. Let us now assume that one has access to N copies of the spin chain, all of them in thermal states of the same temperature, but potentially with different microstate configurations. If one were to measure one spin from each of those chains, the precision would be bounded by $NF_T(X_1)$. The latter approach is clearly not very physical. However, since correlations decay fast enough on a spin chain, one can achieve an approximately similar scaling by measuring N spins on the same chain, taking care that they are far away from each other.

To discuss the relation between correlated and uncorrelated Fisher information, we define the quantity

$$\eta = \frac{F_T(X_{1:2})}{2F_T(X_1)}, \tag{III.22}$$

such that super-additive behaviour gives $\eta > 1$, and sub-additive behaviour $\eta < 1$. We represent η for a wide choice of temperatures and couplings in Fig. III.2a. There is a remarkable difference between the ferromagnetic ($J > 0$) and anti-ferromagnetic ($J < 0$) regimes. Indeed, in the ferromagnetic case, the Fisher information is slightly sub-additive almost everywhere, $\eta \lesssim 1$. On the contrary, for $J < 0$ we observe a strongly super-additive regime, with η achieving high values. This connects back to our previous observation regarding sign of the correlations and additivity of the Fisher information.

To better understand the behaviour of Fig. III.2a, it is instructive to look at the one- and two-sites marginals as a function of temperature. They are represented in Figs. III.2b and III.2c for the antiferro-

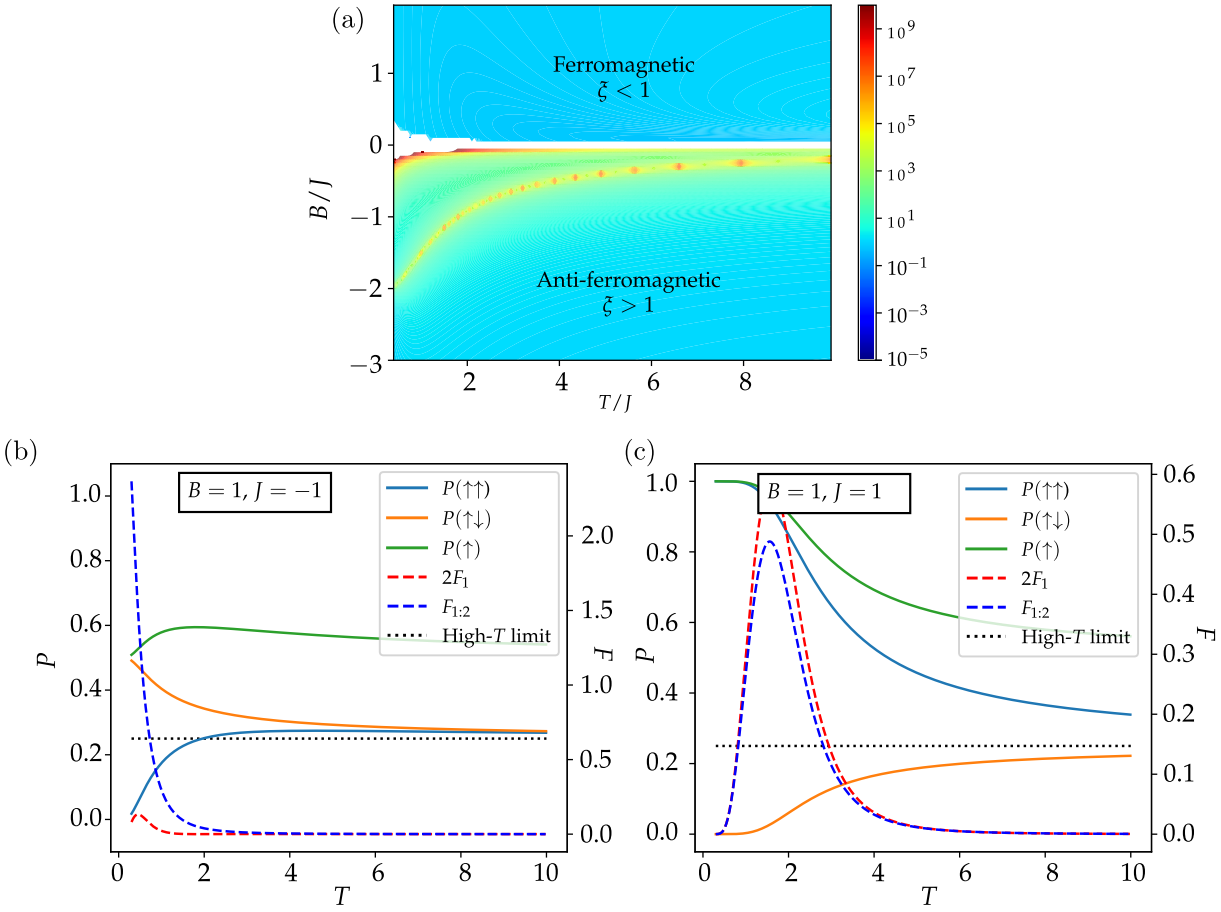

FIG. III.2. (a) Colormap plot of η, representing the Fisher information correlation advantage on an Ising spin chain, as a function of the temperature and the coupling strength. The emerging features are discussed in the main text. All areas with $\eta > 10^{10}$ are shaded in white. (b-c) One- and two-sites marginals for the antiferromagnetic and ferromagnetic cases. Reproduced from [1]

magnetic and the ferromagnetic case, respectively.

We note that the reason for $\eta \gg 1$ in the antiferromagnetic case is mostly that $F(X_1)$ is particularly small. This is a consequence of the degeneracy of the two-sites ground state. In fact, in the antiferromagnetic case, in the $T \to 0$ limit, one has $\Pr(\uparrow\downarrow) = \Pr(\downarrow\uparrow) = \frac{1}{2}$, and $\Pr(\uparrow\uparrow) = \Pr(\downarrow\downarrow) = 0$. In the high-$T$ limit, all the two-sites probabilities converge to the same value of $\frac{1}{4}$. The one-site marginal is obtained as $\Pr(\uparrow) = \Pr(\uparrow\downarrow) + \Pr(\uparrow\uparrow)$. Now: for increasing T, the first term has to decrease by $\frac{1}{4}$ to achieve the high-T value; the second has to increase by the same value. Consequently, the two derivatives almost cancel out, and $\Pr(\uparrow)$ has a very small dependence on T, resulting in a small one-site Fisher information. This effect does not take place in the ferromagnetic regime, where instead the ground state is non-

degenerate, as $\Pr(\uparrow\uparrow) = 1$ and $\Pr(\uparrow\downarrow) = \Pr(\downarrow\uparrow) = \Pr(\downarrow\downarrow) = 0$. In the one-site marginal $\Pr(\uparrow)$, $\Pr(\uparrow\uparrow)$ has to decrease by $\frac{3}{4}$, while $\Pr(\uparrow\downarrow)$ only increases by $\frac{1}{4}$; the two derivatives do not cancel out, and the one-site marginal carries a significant Fisher information.

This line of thinking allows us also to understand the emerging feature of a region of very high η starting at $\frac{B}{J} = -2$, corresponding to null one-site Fisher information. In the antiferromagnetic case, the one-site marginal $\Pr(\uparrow)$ has the same value $\frac{1}{2}$ in both the $T \to 0$ and the $T \to \infty$ limits. In between, there is a maximum, with zero derivative and zero Fisher information. On the contrary, in the ferromagnetic case $\Pr(\uparrow)$ monotonically decreases from 1 to $\frac{1}{2}$ for increasing T.

Finally, a comment is in order regarding the low-B limit. In the absence of an external magnetic field, the spin chain is symmetric under exchange of $\uparrow$ and $\downarrow$; there is consequently no Fisher information carried in the one-site marginals.

III.5. LEARNING AN ϵ-MACHINE

In the previous Sections, we discussed correlated estimation theory for processes with finite Markov order. As a natural generalisation, we consider estimation theory for processes for which an ϵ-machine with a finite number of causal states is known. This theory was recently developed in Ref. [118], and has significant implications in the field of machine learning of stochastic processes.

Consider an ϵ-machine whose transition probabilities depend on the vector of parameters $\vec{\theta}$, that emits symbols in the alphabet $\mathcal{A}$. We assume that the topology of the machine does not change under small modifications of the parameters. Let $\{\sigma\}$ be the set of the causal states, and π be the stationary state of the machine: $\pi(\sigma)$ is the probability of finding the machine in causal state σ, averaged over time and emissions. The Fisher information matrix rate can be proven to be [118]:

$$f_{mn}(\theta) = -\sum_{\sigma} \pi_{\vec{\theta}}(\sigma) \sum_{x\in\mathcal{A}} \Pr(x|\sigma)\partial_{\theta_m}\partial_{\theta_n} \log \Pr_{\vec{\theta}}(x|\sigma). \tag{III.23}$$

We note here the power of the causal shielding effect induced by unifilarity: the complex structure of the joint Fisher information $F_{\vec{\theta}}(X_{1:N})$ is reduced to conditionals depending only on the causal state. We will return to this kind of language and results again when considering processes with closed patterns in Sect. IV.3.

Chapter IV

Estimation and hypothesis testing for the quantum jump unraveling

Continuously monitored quantum systems are characterised by instantaneous emissions, if the monitoring is photodetection-like. The emission record contains information about the dynamics of the emitting system. In this Chapter, we give precision bounds on parameter estimation and hypothesis testing based on emission records.

The main results are the expression of the Fisher information rate and the error exponents for multi-channel renewal processes of Eq. (IV.12) and Eq. (IV.23), the splitting of the Fisher information in terms of channels and times of Eq. (IV.15), error bounds for systems with closed patterns in Sect. IV.3, and the in-depth study of the monitoring operator formalism of Sect. IV.4 A for non-renewal processes.

In quantum metrology [8, 9, 31], one usually assumes that the experimeter has access to a large number of identical (i.i.d.) copies of the physical system at hand, and is able to optimise the measurement to be performed on the system. When dealing with processes, though, one has to deal with correlations. Fundamental bounds on the precision of parameter estimation for correlated quantum processes have been very recently derived using the formalism of quantum combs [42–45, 125]. All these works still assume the capability to optimise over the measurement scheme.

Here, we take a different perspective. We assume no direct access to the physical system: the experimenter can only detect and register emissions. We ask how much one can learn about parameters in the system or decide between two hypotheses by only looking at the emissions record. In this sense, the results of this Chapter generalise the approach of Refs. [35, 36] in parameter estimation, and Refs. [61–66] in hypothesis testing. Work in a similar direction has recently been done for the specific case of thermometry in Ref. [37]. Elegant bounds on the precision of estimation, optimised over all the possible continuous measurement schemes, corresponding to all the possible unravelings of the GKSL equation, have been proposed in Refs. [39–41].

The main questions we set to answer are the following: (a) how precise can estimation and hypothesis testing based on emission records be? (b) how can these tasks be actually performed?

References and collaboration information

The main references for this Chapter are the works in Refs. [4, 6]. Note that the latter, at thesis submission stage, is still in preparation. Both papers were written in collaboration with J.A. Smiga, G.T. Landi and F.C. Binder. I was first author on both works, responsible for most of the analytical and numerical results. The numerical part of the two papers was moved to Chapter V. The discussion of classical post-processing appears in Chapter VII. The discussion of estimation theory for closed patterns in Sect. IV.3 has not appeared before.

IV.1. THE MEASUREMENT RECORD OF A JUMP PROCESS

Let us consider a quantum system evolving according to the GKSL master equation (II.10), introduced in Sect. II.2. We consider specifically jump unraveling. The infinitesimal time evolution of the system is given by

$$\rho_j(t+dt) = \frac{M_j \rho(t) M_j^\dagger}{p_j}, \qquad \text{((II.3) revisited)}$$

where the Kraus operators $\{M_j\}$ depend on the jump operators as

$$M_0 = \mathbb{I} - iHdt - \frac{1}{2}\sum_k \gamma_k A_k^\dagger A_k dt \quad ; \quad M_k = \sqrt{\gamma_k} A_k \sqrt{dt}. \qquad \text{((II.12) revisited)}$$

From a continuos monitoring perspective [99], each jump is associated with a click in a detector. The experimenter can to retrieve a *measurement record*, that is the list of the times and the channels of the jumps. Formally, the measurement record is the list of pairs

$$\omega_{1:N} = \{(\tau_1, k_1), (\tau_2, k_2), \ldots, (\tau_N, k_N)\}, \tag{IV.1}$$

where k_i labels the channel of the jump, and τ_j is the time in between jumps j and $j-1$.

Let us introduce the jump superoperator

$$\mathcal{J}_k \bullet = \gamma_k A_k \bullet A_k^\dagger, \tag{IV.2}$$

representing the (unnormalised) action of a jump on the quantum state. We also define the no-jump Lindblad superoperator

$$\mathcal{L}_0 = \mathcal{L} - \sum_{k\in\mathbb{M}} \mathcal{J}_k, \tag{IV.3}$$

where $\mathcal{L}$ is the ordinary Lindblad operator, defined in Eq. (II.10). The set $\mathbb{M}$ includes *monitored* jumps. In general, in fact, some of the jump channels may remain unmonitored. This means that jumps in the unmonitored channel(s), when they take place, do not appear in the measurement record.

In the jump unraveling, the system undergoes a smooth, continuous evolution governed by the no-jump Lindbladian $\mathcal{L}_0$, interspersed by abrupt transitions, corresponding to monitored jumps. In the language of stochastic calculus [99, 126], one can write the evolution of the state as

$$d\rho = dt\mathcal{L}\rho + \sum_{k\in\mathbb{M}} (dN_k(t) - dt\mathrm{Tr}[\mathcal{J}_k\rho]) \left(\frac{\mathcal{J}_k\rho}{\mathrm{Tr}[\mathcal{J}_k\rho]} - \rho \right). \tag{IV.4}$$

In this expression, each Poisson increment $dN_k(t)$ is a stochastic variable that takes values 0 (no jump in channel k at time t) or 1 (jump in channel k at time t). The probability of a jump at time t in channel k is

$$\Pr[dN_k(t) = 1] = \mathrm{Tr}[\mathcal{J}_k\rho(t)]dt. \tag{IV.5}$$

Since the jump probability is of order dt, no jumps will happen in the vast majority of the timesteps. This has significant consequences in the computational simulation of quantum jump trajectories, see Chapter V.

By integrating the evolution equation (IV.4), we write the probability of a measurement record, given an initial state ρ_0, as

$$\Pr(\omega_{1:N}|\rho_0) = \mathrm{Tr}\left[\mathcal{J}_{k_N} e^{\mathcal{L}_0\tau_N} \dots \mathcal{J}_{k_1} e^{\mathcal{L}_0\tau_1}\rho_0\right]. \tag{IV.6}$$

For brevity, in the following the conditioning on the initial state will be mostly omitted. This probability

is normalised as

$$\int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \ldots \int_0^\infty d\tau_N \sum_{k_1 \in \mathbb{M}} \sum_{k_2 \in \mathbb{M}} \ldots \sum_{k_N \in \mathbb{M}} \Pr(\omega_{1:N}) = 1. \tag{IV.7}$$

The integrals are well defined if the system has no dark subspaces. In other words, from any state, a jump always happens, possibly after a long time [3].

The probability density $\Pr(\omega_{1:N})$ is the starting point for the estimation theory and hypothesis testing tasks of this chapter. It plays the role of the overall joint probability $\Pr(X_{1:N})$ of a stochastic process. In particular, it encodes all the correlations of the *classical* stochastic process emerging from the quantum process through the measurements.

IV.2. RENEWAL PROCESSES

The probability density of Eq. (IV.6) is a difficult object to work with, as it encodes complex long-range correlations. Analytical results can, however, be obtained when all the monitored jumps satisfy the *renewal* property

$$\mathcal{J}_k \rho = \mathrm{Tr}\left[\mathcal{J}_k \sigma_k\right]. \tag{IV.8}$$

Here, σ_k is a state that may depend on the jump channel k, but not on the state before the jump ρ. All quantum memory in the system is lost upon a jump. Processes for which all the jump operators satisfy the renewal property are called *renewal processes* [127]. If a process is renewal, then the measurement record probability of Eq. (IV.6) factorises as

$$\Pr(\omega_{1:N}) = W(\tau_N, k_N | k_{N-1}) \ldots W(\tau_1, k_1 | k_0). \tag{IV.9}$$

The function $W(k, \tau | q)$ is the *waiting time distribution* (WTD). It expresses the probability that a jump will occur in channel k, τ time after a jump in channel q has taken place. We get an explicit expression for the WTD by plugging Eq. (IV.8) into Eq. (IV.6), getting

$$W(k, \tau | q) = \mathrm{Tr}\left[\mathcal{J}_k e^{\mathcal{L}_0 \tau} \sigma_q\right]. \tag{IV.10}$$

The waiting time distribution is normalised as

$$\sum_{k\in\mathbb{M}}\int_0^\infty W(k,\tau|q)=1, \tag{IV.11}$$

for all $q\in\mathbb{M}$, assuming that there are no dark subspaces. If the system has dark subspaces, then the WTD is not normalisable [3].

A. Analytical results for estimation theory

A renewal quantum jump process generates a classical Markov process of order 1 for its output measurement record. Indeed, the probability of the pair $(\tau_j,k_j)\in\omega_{1:N}$ depends only on the previous pair (τ_{j-1},k_{j-1}). We apply the results of Chapter III. Let us now assume that a parameter θ is encoded in the dynamics of the system (i.e., in the Lindblad operator $\mathcal{L}$). As per Eq. (III.7) with $\mathcal{M}=1$, then, the Fisher information scales linearly with the number of jumps. The Fisher information rate is given by the condtional Fisher information of two consecutive jumps, and is computed in detail in the appendices of Ref. [4], yielding

$$F_\theta(\omega_{1:N})=N\sum_{k,q\in\mathbb{M}}p_q\int_0^\infty d\tau\frac{\left[\partial_\theta W(k,\tau|q)\right]^2}{W(k,\tau|q)}. \tag{IV.12}$$

Here, p_q represents the probability that a jump happens in channel q from the steady state of the dynamics, conditioned on the fact that a jump occurs at all:

$$p_q:=\frac{\mathrm{Tr}\left[\mathcal{J}_q\rho_{\mathrm{SS}}\right]}{\sum_{j\in\mathbb{M}}\mathrm{Tr}\left[\mathcal{J}_j\rho_{\mathrm{SS}}\right]}. \tag{IV.13}$$

This Fisher rate information is coherent with the upper-bound obtained by optimising over all the unravelings [40].

It is possible to separate the contributions to the Fisher information given by the jump channels and the jump times. In general, the Fisher information chain rule of Eq. (I.17) yields

$$F_\theta(\omega_{1:N})=F_\theta(k_1,\ldots,k_N)+F_\theta(\tau_1,\ldots,\tau_N|k_1,\ldots,k_N), \tag{IV.14}$$

where the first term refers to the Fisher information in the jump channels alone. The second term gives

the residual Fisher information in the jump times, assuming that the jump channels are already known. The jump channels Fisher information is easy to compute. Let us start with the marginalisation over times of the WTD:

$$p(k|q) = \int_0^\infty W(k,\tau|q) = -\mathrm{Tr}\left[\mathcal{J}_k \mathcal{L}_0^{-1} \sigma_q\right]. \tag{IV.15}$$

This quantity represents the conditional probability that a jump k is observed directly (i.e. without other monitored jumps in between) after a jump q, regardless of the time in between. The stochastic process on the channels only, $k_{1:N}$, forms a Markov order 1 chain as well. The joint Fisher information is then

$$F_\theta(k_1,\ldots,k_N) = N \sum_{k,q\in\mathbb{M}} p(k|q)p_q \left(\frac{\partial}{\partial\theta}\log p(k|q)\right)^2. \tag{IV.16}$$

The residual Fisher information of the jump times can be computed by difference:

$$F_\theta(\tau_1,\ldots,\tau_N|k_1,\ldots,k_N) = N \sum_{k,q\in\mathbb{M}} p(k|q)p_q \int_0^\infty d\tau \frac{[\partial_\theta W(\tau|k,q)]^2}{W(\tau|k,q)}, \tag{IV.17}$$

where

$$W(\tau|k,q) = \frac{W(k,\tau|q)}{p(k|q)}. \tag{IV.18}$$

The decomposition of Eq. (IV.14) can also be written by inverting the roles of jump times and channels, as

$$F_\theta(\omega_{1:N}) = F_\theta(\tau_1,\ldots,\tau_N) + F_\theta(k_1,\ldots,k_N|\tau_1,\ldots,\tau_N). \tag{IV.19}$$

While the equation is valid, the Fisher information $F_\theta(\tau_{1:N})$ cannot be analytically expressed. In fact, when jump channels are ignored, the process may lose its renewal structure[2]. Therefore, $\tau_{1:N}$ is not, in general, a finite Markov order process, and does not satisfy the hypotheses of Eq. (III.7). Though, the Fisher information for the times alone is upper-bounded by the overall Fisher information of the renewal process of Eq. (IV.12). Its behavior is then at most linear with the number of jumps.

[2] Indeed, the process *always* loses its renewal structure, unless all the jump channels lead to the same outgoing state

B. Example: qubit thermometry

A simple example of a renewal system is a qubit evolving according to the Hamiltonian

$$H = \frac{\omega}{2}\sigma_z + \frac{\Omega}{2}\sigma_x, \tag{IV.20}$$

where ω and Ω are parameters. Physically, ω represents a detuning and Ω the Rabi frequency of a driving e.m. field. The qubit is weakly coupled to a thermal bath.The coupling is described by the jump operators

$$L_+ = \sqrt{\gamma\bar{n}}\sigma_+ \qquad ; \qquad L_- = \sqrt{\gamma(\bar{n}+1)}\sigma_-, \tag{IV.21}$$

where γ represents the strength of the coupling, and $\bar{n}$ is the bosonic average occupation number, connected to the temperature of the bath via the Bose-Einstein distribution. We assume that both the jump channels are observable, $\mathbb{M} = \{L_+, L_-\}$, and set out to estimate $\bar{n}$. The Fisher information for $\bar{n}$ can be computed analytically via Eq. (IV.12) (the expression is cumbersome, but given in Ref. [4]). The splitting between jump channels and jump times can be computed as well, as represented in Fig. IV.1.

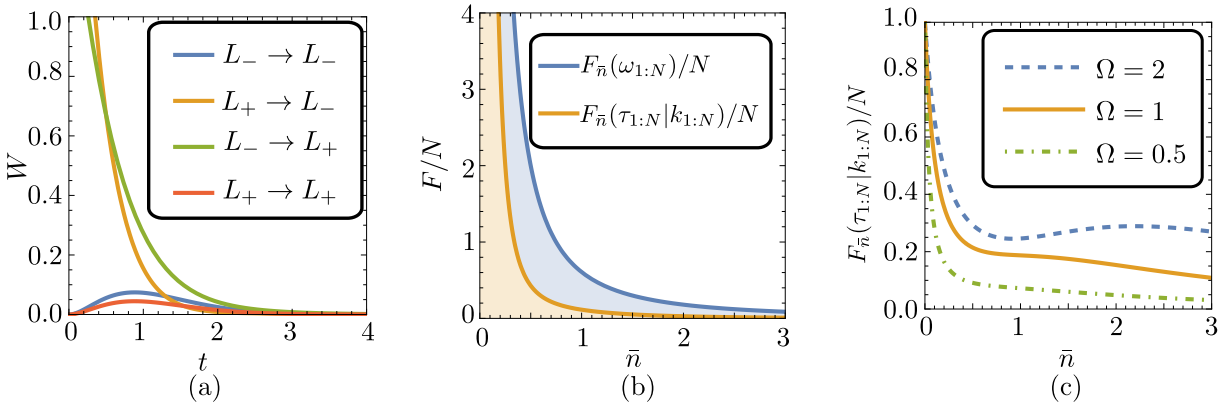

FIG. IV.1. (a) the four waiting time distributions between pairs of jumps for qubit thermometry; (b) splitting of the Fisher information in the measurement record between jump times and jump channels; (c) the Fisher information in the channels alone, for different values of the Rabi frequency. Reproduced from [4].

It is interesting to note the behaviour in the $\Omega \to 0$ limit. In this case, the only transition mechanism is incoherent, and the model becomes fully classical. That is, it cannot generate coherence. In the absence of a Hamiltonian driving, the jumps have to alternate: L_+ is always followed by L_-, and vice versa.

Since the distribution of jump channels is deterministic, there is no Fisher information contained in the jump channels. All the information about the parameter $\bar{n}$ is then located in the jump times.

C. Analytical results for hypothesis testing

Let us now consider renewal processes from the perspective of hypothesis testing. We consider two candidate systems as possible explanations of a given mesurement record, as null and alternative hypothesis. Both the systems are renewal. For a fixed probability of false alarm, we would like to determine the Chernoff-Stein scaling [17] of the missed detection error.

Since both the hypotheses are renewal systems, one can define waiting time distributions according to both. We label them as W_0 for the null hypothesis, and W_1 for the alternative one.

We introduce the matrix

$$M(k,q) = \int_0^\infty d\tau W_1(k,\tau|q) \log \frac{W_1(k,\tau|q)}{W_0(k,\tau|q)}, \tag{IV.22}$$

The expectation value of the log-likelihood ratio according to the alternative hypothesis is the Kullback-Leibler divergence between the processes

$$\mathbb{E}_1[L(\omega_{1:N})] \sim N \sum_{k,q\in\mathbb{M}} p_{1,q} M(q,k). \tag{IV.23}$$

The probability $p_{1,q}$ is defined as in Eq. (IV.13), in terms of the alternative hypothesis system. This result can be read in parallel with the one for the Fisher information of renewal processes in Eq. (IV.12). The Chernoff-Stein decay rate of the missed detection error is then

$$\lim_{N\to\infty} \frac{1}{N} P_{\mathrm{M}}(\omega_{1:N}) = \sum_{k,q\in\mathbb{M}} p_{1,q} M(q,k). \tag{IV.24}$$

D. Example: three level maser

We consider the problem of distinguishing between the "coherent" and the "incoherent" versions of the three level maser, a well-known model in quantum thermodynamics [5, 128–130]. The system is composed by three levels $|1\rangle$, $|2\rangle$, and $|3\rangle$. They are coupled with two bosonic thermal baths with average occupations

$\bar{n}_H$ and $\bar{n}_C$, with $\bar{n}_H > \bar{n}_C$. The coupling parameters for the baths are γ_H and γ_C, respectively. The system Hamiltonian, in an appropriate rotating frame, is

$$H = \epsilon(\sigma_{12} + \sigma_{21}) + \Delta\sigma_{22}, \tag{IV.25}$$

where $\sigma_{ij} = |i\rangle\langle j|$, and ϵ is the driving strength. Here, we set $\Delta = 0$. The coupling with the baths generates four jump channels, with operators

$$\begin{aligned} L_{H,\text{out}} &= \sqrt{\gamma_H(\bar{n}_H + 1)}\sigma_{13} \quad ; \quad L_{H,\text{in}} = \sqrt{\gamma_H \bar{n}_H}\sigma_{31} \\ L_{C,\text{out}} &= \sqrt{\gamma_C(\bar{n}_C + 1)}\sigma_{23} \quad ; \quad L_{C,\text{in}} = \sqrt{\gamma_C \bar{n}_C}\sigma_{32}. \end{aligned} \tag{IV.26}$$

In the literature, the three level maser has been compared with an incoherent model [5, 130]. It is obtained via Fermi's golden rule [130] to yield the same steady state currents. In the incoherent model, Hamiltonian driving is replaced by two incoherent, unmonitored jumps with operators

$$L_{N,\text{in}} = \sqrt{\gamma_N}\sigma_{21} \quad ; \quad L_{N,\text{out}} = \sqrt{\gamma_N}\sigma_{12}, \tag{IV.27}$$

where $\gamma_N = \frac{2\epsilon^2}{(\gamma_H n_H + \gamma_C n_C)/2}$. Both the coherent and the incoherent models are renewal.

Let us now consider the following hypothesis testing task. We are given a black box that contains either the classical (null hypothesis) or the quantum (alternative hypothesis) model of the three level maser. Can we distinguish between them? In Figs. IV.2b-c, we showcase our results, illustrating the behaviour of the log-likelihood ratio with N. Interestingly, the distribution of the log-likelihood under the null hypothesis (classical system) is much less peaked than the one under the alternative hypothesis. This can be traced back to the finer structure of the quantum waiting time distributions, which have a complicated pattern of oscillations, with respect to the simpler sum of exponentials of the classical case.

IV.3. PROCESSES WITH CLOSED PATTERNS

In this section, we present estimation theory and hypothesis testing results on quantum jump processes with closed patterns. These processes, first introduced in Ref. [131], slightly generalise the renewal condition of Eq. (IV.8). Interestingly, the behaviour of these processes closely recalls the one of ϵ-machines, introduced in Sect. I.1.

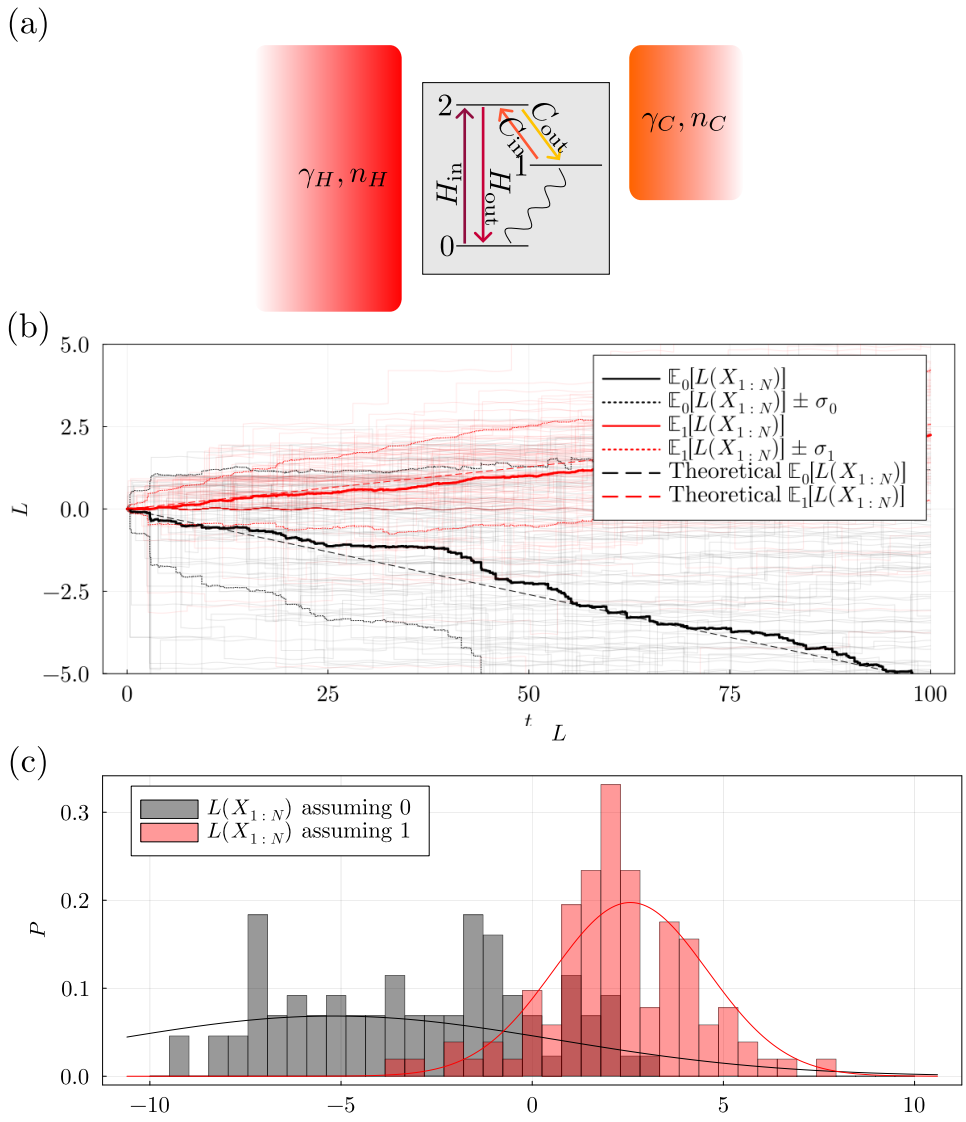

FIG. IV.2. (a) Schematic of the (quantum) three level maser; (b) simulation of the hypothesis testing task over 100 trajectories, showing the loglikelihood as a function of time; (c) histogram of the distribution of the loglikelihood ratio at the end of the evolution. Parameters: $\epsilon = 0.5$, $\gamma_H = 0.1$, $\gamma_C = 0.2$, $\bar{n}_H = 0.8$, $\bar{n}_C = 0.2$.

For renewal systems, we gave an analytical treatment of the separate metrological contribution of jump channels and jump times. Let us now focus on the channels alone, throwing away all the information about the jump times. We define the jump superoperators integrated over time

$$\mathcal{M}_k = \int_0^\infty d\tau \mathcal{J}_k e^{\mathcal{L}_0 \tau} = -\mathcal{J}_k \mathcal{L}_0^{-1}. \tag{IV.28}$$

The action of the superoperator $\mathcal{M}_k$ on the state represents a jump in channel k *at some point* in time. We can think of the dynamics on $\{k_{1:N}\}$ alone as a discrete stochastic process, with evolution rule

$$\rho_{j+1} = \frac{\mathcal{M}_{j+1}\rho_j}{\mathrm{Tr}[\mathcal{M}_{j+1}\rho_j]}. \tag{IV.29}$$

A system supports a closed pattern if there exists a finite set $\mathbb{B} = \{\sigma_1, \sigma_2, \ldots\}$ of states such that the dynamics of Eq. (IV.29) is closed in that set, i.e.

$$\frac{\mathcal{M}_k\sigma_j}{\mathrm{Tr}[\mathcal{M}_k\sigma_j]} = \sigma_{f(j,k)} \in \mathbb{B}, \text{ for all } \sigma_j \in \mathbb{B}, \tag{IV.30}$$

where f is the function that expresses the evolution rule. All renewal processes support closed patterns. In the renewal case, in particular, the function f depends on the jump index k, but not on the previous state index j. However, there are examples of non-renewal systems that support closed patterns [131].

A very clear connection can be established between them and ϵ-machines [73, 74]. Eq. (IV.29) defines an unifilar process [74]. Given the current state $\sigma \in \mathbb{B}$ and the emitted symbol k, in fact, the evolution given by f is fully deterministic. Only the last state σ is relevant for the future behaviour, acting therefore as a causal shield between past and future.

Remarkably, the state that plays the role of the stationary state of the stochastic model is not the steady state ρ_{SS} of the OQS evolution, but rather the jump-steady state [131]

$$\rho_{\mathrm{JSS}} = \frac{\mathcal{J}\rho_{\mathrm{SS}}}{\mathrm{Tr}[\mathcal{J}\rho_{\mathrm{SS}}]}, \tag{IV.31}$$

where $\mathcal{J} = \sum_{k\in\mathbb{M}} \mathcal{J}_k$.

Given the connection with the ϵ-machine formalism, the results of Ref. [118] apply directly. In particular, assume that a parameter θ is encoded in the dynamics, and that the process supports closed patterns. The Fisher information rate is given by

$$f(\theta) = -\sum_{j\in\mathbb{M}}\sum_{\sigma\in\mathbb{B}} \mathrm{Tr}[\rho_{\mathrm{JSS}}\sigma]\mathrm{Tr}[\mathcal{M}_j\sigma]\,\frac{\partial^2}{\partial\theta^2}\mathrm{Tr}[\mathcal{M}_j\sigma]. \tag{IV.32}$$

For renewal processes, this exactly corresponds to the Fisher information of jump channels alone in Eq. (IV.16).

From the perspective of hypothesis testing, assume that both the null and the alternative hypotheses

are processes that support close patterns. The Chernoff-Stein exponent for the missed detection error is given by

$$\mathbb{E}_1[L(\omega_{1:N})] = \sum_{j\in\mathbb{M}}\sum_{\sigma\in\mathbb{B}} \mathrm{Tr}[\mathcal{M}_{1,j}\sigma]\mathrm{Tr}[\rho_{0,\mathrm{JSS}}\sigma]\log\frac{\mathrm{Tr}[\mathcal{M}_{1,j}\sigma]}{\mathrm{Tr}[\mathcal{M}_{0,j}\sigma]}. \tag{IV.33}$$

More details on this result will be given in the publication [6].

A. Example: two coupled qubits

We consider a system composed of two coupled qubits, where injections always happen in the left-most qubit, and emissions from the right qubit. There is an Hamiltonian exchange of excitations between the qubits, as

$$H = g(\sigma_+^L\sigma_-^R + \sigma_-^L\sigma_+^R), \tag{IV.34}$$

where g is the coupling strength, $\sigma_+^j = |1\rangle\langle 0|$ on qubit j, $\sigma_-^j = (\sigma_+^j)^\dagger$. The system has two jump channels, with operators

$$L_I = \sqrt{\gamma_I}\sigma_+^L \qquad ; \qquad L_E = \sqrt{\gamma_E}\sigma_-^R. \tag{IV.35}$$

The system is not renewal: an injection resets the state of the left qubit, but not the right one; an emission resets the right qubit, but not the left one. However, if one integrates over time, the system exhibits closed patterns [131] with $\mathbb{B} = \{|00\rangle, |\psi\rangle, |11\rangle\}$, where the state $|\psi\rangle$ is any state in the span of $|10\rangle$ and $|01\rangle$. Due to the structure of the evolution, represented in Fig. IV.3b, there can never be three consecutive jumps of the same kind. In fact, from $|00\rangle$ one can only observe an injection, leading to $|\psi\rangle$; from there, both an emission (going back to $|00\rangle$), and a further injection (to $|11\rangle$) are possible. From $|11\rangle$, only emissions can take place. The dynamics is controlled by the lone free parameter q, that is the conditional probability of an injection from state $|\psi\rangle$, given by

$$q = \frac{\gamma_I}{\gamma_I + \gamma_E}. \tag{IV.36}$$

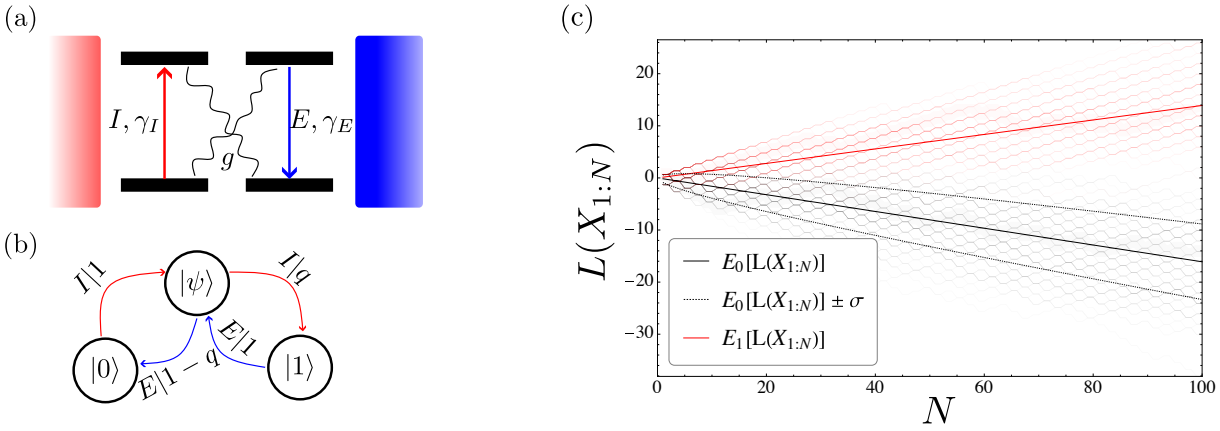

FIG. IV.3. (a) diagrammatic representation of the two coupled qubits system with closed patterns. (b) ϵ-machine-like representation of the dynamics among the recurrent states. (c) hypothesis testing diagram. Parameters for (c): $q = 0.5$ for the null hypothesis, $q = 0.2$ for the alternative hypothesis.

Observe that the parameter g does not appear in the time-integrated dynamics. In fact, g controls the internal temporal dynamics of the $|\psi\rangle$ state, but acts neither on $|00\rangle$ nor on $|11\rangle$. Consequently, there is no way of estimating g from the sequence of jump channels alone, if the times are not accessible. In the same way, two systems that only differ by g cannot be distinguished without access to the jump times. In Fig. IV.3c, we show the hypothesis testing diagram when choosing between two processes with different q.

We compute also the Fisher information rate via Eq. (IV.32). The jump-steady state is

$$\Pr(|00\rangle) = \frac{1-q}{2} \qquad \Pr(|\psi\rangle) = \frac{1}{2} \qquad \Pr(|11\rangle) = \frac{q}{2}, \tag{IV.37}$$

and the Fisher information rate for q is, consequently:

$$f(q) = \frac{1}{2(q-q^2)}. \tag{IV.38}$$

The Fisher information is very high for $q \sim 0$ and $q \sim 1$, while it is much lower for intermediate values of q. If $q \sim 0$, the system oscillates practically only between $|00\rangle$ and $|\psi\rangle$, as the transition $|\psi\rangle \to |11\rangle$ is strongly suppressed. As a result, the observation of the (extremely rare) II sequence leads to a very precise estimation of q. This does not happen when $q \sim \frac{1}{2}$, as there is not a strongly favoured path in the dynamics. A similar logic can be applied for $q \sim 1$.

IV.4. NON-RENEWAL PROCESSES AND THE MONITORING OPERATOR FORMALISM

Let us now focus on the general case, in which the probability of the measurement record of Eq. (IV.6) cannot be factorised. In this situation, long-lasting quantum correlations flow through the system and across jumps, making the problem not amenable to an analytical discussion. Averaging any quantity over all the possible measurement records requires the computation of extremely high-dimensional integrals, according to the normalisation rule of Eq. (IV.7). For example, the Fisher information for a parameter θ would be

$$F_\theta(\omega_{1:N}) = \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \dots \int_0^\infty d\tau_N \sum_{k_1\in\mathbb{M}} \sum_{k_2\in\mathbb{M}} \dots \sum_{k_N\in\mathbb{M}} \Pr(\omega_{1:N}) \left(\frac{\partial}{\partial\theta} \log \Pr(\omega_{1:N}) \right)^2 . \tag{IV.39}$$

If one was to solve the integration with a traditional grid technique, one would split the integration domain in small regions, and compute the value of the integrand for each of them. Problematically, the integration routine would spend the vast majority of the time computing values in regions of extremely low probability, and negligible contribution to the Fisher inforrmation. To avoid this problem, the trick is to compute the quantities of interest *on quantum trajectories*. We present the idea in general, and then specialise it to the cases of the Fisher information and the log-likelihood ratio. Similar methods have been developed for the computation of two-points correlation functions [35, 132].

Assume we want to compute a quantity Q, and that it can be written as the expectation value of a scalar q on all the possible measurement records: $Q = \mathbb{E}_{\omega_{1:N}}[q(\omega_{1:N})]$. Then, one can proceed as follows:

1. Write evolution equations on trajectories for the state;
2. Write evolution equations on trajectories for the quantity q;
3. Evolve the state, and the object q in parallel; record the values of q along the trajectory;
4. Repeat for a large number of trajectories;
5. The value of Q is given by the arithmetic average of the outcomes.

Formally, this is equivalent to computing an integral like Eq. (IV.39) with a Monte Carlo approach [133, 134]. The advantage lies in the fact that, instead of giving all regions of the grid the same weight, the evolution of the state drives the integration, and the most relevant regions are explored with priority. In practice, this gives a much more efficient computation. We will show also that this method allows

for some interesting analytical insights. In the following, we specialise this method to treat estimation theory and hypothesis testing on quantum jump trajectories beyond the renewal case.

A. The monitoring operator formalism for the computation of the Fisher information on quantum trajectories

In this section, we review the formalism of the monitoring operator, introduced first in Ref. [35], and further developed in Ref. [38]. This formalism allows us to compute the Fisher information on quantum trajectories. We define the *conditional unnormalised* state on the measurement record $\omega_{1:N}$ as

$$\tilde{\rho}_{\omega_{1:N}} = \mathcal{J}_{k_N} e^{\mathcal{L}_0 \tau_N} \mathcal{J}_{k_{N-1}} \dots \mathcal{J}_{k_1} e^{\mathcal{L}_0 \tau_1} \rho_0. \tag{IV.40}$$

Crucially, the trace of the unnormalised conditional state corresponds to the probability density of the trajectory $\omega_{1:N}$, as per Eq. (IV.6). Let us now introduce the *monitoring operator* $\xi_{\omega_{1:N}}$ for the parameter θ as

$$\xi_{\omega_{1:N}} = \frac{\partial_\theta \tilde{\rho}_{\omega_{1:N}}}{\mathrm{Tr}[\tilde{\rho}_{\omega_{1:N}}]}. \tag{IV.41}$$

The monitoring operator allows to keep track of how the process depends on the parameter θ. It is connected to the Fisher information contained in the measurement record as

$$F_\theta(\omega_{1:N}) = \mathbb{E}_{\omega_{1:N}} \left[\left(\mathrm{Tr}[\xi_{\omega_{1:N}}] \right)^2 \right], \tag{IV.42}$$

where the expectation value is taken on the space of all quantum trajectories, with the respective measure. In terms of the general description above, here Q is the Fisher information, and $q = \mathrm{Tr}\left[\xi_{\omega_{1:N}}\right]^2$. We want to write evolution equations for $\xi_{\omega_{1:N}}$. We assume inductively that we know $\xi_{\omega_{1:N-1}}$, and we set out to determine $\xi_{\omega_{1:N}}$. We start from the evolution equation for the unnormalised state

$$\tilde{\rho}_{\omega_{1:N}} = \mathcal{J}_{k_N} e^{\mathcal{L}_0 \tau_N} \tilde{\rho}_{\omega_{1:N-1}}, \tag{IV.43}$$

where $\mathcal{J}_{k_N}$, $\mathcal{L}_0$ and $\tilde{\rho}_{\omega_{1:N-1}}$ can all depend on θ. Differentiating the unnormalised conditional state with respect to θ yields

$$\partial_\theta \tilde{\rho}_{\omega_{1:N}} = (\partial_\theta \mathcal{J}_{k_N}) e^{\mathcal{L}_0 \tau} \tilde{\rho}_{\omega_{1:N-1}} + \mathcal{J}_{k_N} (\partial_\theta \mathcal{L}_0 \tau_N) e^{\mathcal{L}_0 \tau_N} \tilde{\rho}_{\omega_{1:N-1}} + \mathcal{J}_{k_N} e^{\mathcal{L}_0 \tau_N} (\partial_\theta \tilde{\rho}_{\omega_{1:N-1}}). \tag{IV.44}$$

Using the relation $\tilde{\rho}_{\omega_{1:N}} = \rho_{\omega_{1:N}} \mathrm{Tr}[\tilde{\rho}_{\omega_{1:N}}]$, where $\rho_{\omega_{1:N}}$ denotes the *normalised* conditional state, we can finally write the evolution equation for the monitoring operator as

$$\xi_{\omega_{1:N}} = \frac{(\partial_\theta \mathcal{J}_{k_N}) e^{\mathcal{L}_0 \tau_N} \rho_{\omega_{1:N}} + \mathcal{J}_{k_N} (\partial_\theta \mathcal{L}_0 \tau_N) e^{\mathcal{L}_0 \tau_N} \rho_{\omega_{1:N-1}} + \mathcal{J}_{k_N} e^{\mathcal{L}_0 \tau_N} \xi_{1:N-1}}{\mathrm{Tr}\left[\mathcal{J}_{k_N} e^{\mathcal{L}_0 \tau_N} \rho_{\omega_{1:N}}\right]}. \tag{IV.45}$$

As observed in Ref. [38], it is possible to express the evolution rule for the monitoring operator in terms of the *normalised* conditional state. This is important for numerical implementations, since the trace of the unnormalised conditional state vanishes exponentially.

Let us now shortly consider the problem of the initial value of the monitoring operator. By the definition of Eq. (IV.41), the monitoring operator for a normalised state ρ_0 is traceless. Indeed, there is no information at the beginning about θ, $F_\theta(\{\}) = 0$. However, nothing forces the monitoring operator at the beginning of the evolution to be the zero operator. Indeed, in many situations of practical interest the evolution starts, for instance, from the steady state, $\rho_0 = \rho_{\mathrm{SS}}$ which, in general, depends on the value of θ, and $\xi_0 \neq 0$. Numerically we observe, though, that the dependence on the initial state is quickly lost. The choice of the initial value for the monitoring operator has, therefore, a negligible impact on the estimation performances.

The main reason to introduce the monitoring operator is numerical; see the discussion in Chapter V. However, it is instructive and insightful to give a deeper physical interpretation of its meaning. In particular, we establish a connection between the monitoring operator and maximum likelihood estimation performed on the measurement record $\omega_{1:N}$.

Given a fixed measurement record, the MLE prescription of Eq. (I.21) requires to maximise its probability by tuning the parameter $\hat{\theta}$. For quantum jump trajectories, this is very tricky, as the probability vanishes exponentially with the number of jumps N, and there is, in general, no way of factorising it. A direct way of performing MLE on quantum trajectories is discussed in Ref. [35], using a renormalised conditional state. Alternatively, we show that the monitoring operator can be used.

Let us define $\tilde{\rho}^{\tilde{\theta}}_{\omega_{1:N}}$ as the unnormalised conditional state for the measurement record $\omega_{1:N}$, assuming $\tilde{\theta}$

as the value for the parameter. Let $\xi^{\tilde{\theta}}_{\omega_{1:N}}$ be the corresponding monitoring operator. The loglikelihood function is then

$$\ell_{\omega_{1:N}}(\tilde{\theta}) = \Pr\left[\omega_{1:N}|\,\tilde{\theta}\right] = \mathrm{Tr}\left[\tilde{\rho}^{\tilde{\theta}}_{\omega_{1:N}}\right], \tag{IV.46}$$

and the maximisation condition yields $\partial_{\tilde{\theta}}\ell(\tilde{\theta}) \stackrel{!}{=} 0$, which implies:

$$\mathrm{Tr}\left[\xi^{\tilde{\theta}}_{\omega_{1:N}}\right] \stackrel{!}{=} 0. \tag{IV.47}$$

The problem of performing MLE on the measurement record is then mapped onto finding the value of $\tilde{\theta}$ that makes the monitoring operator traceless. Importantly, the monitoring operator does not vanish for long trajectories, allowing for numerically stable estimation.

The monitoring operator formalism allows for an analytical (though still stochastic) expression for the Fisher information rate over time. Consider an infinitesimal time increment dt, and let us represent the evolution of the system in these infinitesimal timesteps. The Fisher information chain rule of Eq. (I.17) yields:

$$F_\theta(\omega_{0:t+dt}) - F_\theta(\omega_{0:t}) = F_\theta(\omega_{0:t+dt}|\omega_{0:t}) = \sum_{k\in\mathbb{M}} \mathbb{E}_{\omega_{1:t}}\left[\frac{1}{I^k_t}\left(\partial_\theta I^k_t\right)^2\right] dt, \tag{IV.48}$$

where $I^k_t = \mathrm{Tr}\left[\mathcal{J}_k\rho_t\right]/dt$ is the stochastic current at time t in channel k. The expression is definitely suggestive, as it recalls the definition of the Fisher information (as $I^k_t dt$ is a probability distribution). Remarkably, no terms including the interplay of different currents appear. Despite our efforts, we did not manage to analytically compute the expectation value. So far, then, the meaning of the equation remains mostly formal: in practice, only a numerical approach is feasible when the process is non-renewal.

B. Example: two coupled qubits

We consider two coupled qubits labelled respectively as A and B, with Hamiltonian

$$H = \Omega_A\sigma^A_x + \Omega_B\sigma^B_x g\left(\sigma^A_+\sigma^B_- + \sigma^A_-\sigma^B_+\right), \tag{IV.49}$$

where σ_j^K is the j-th Pauli matrix on qubit K, $\sigma_- = (\sigma_x - i\sigma_y)/2$, Ω_K are the Rabi frequencies, and g represents the coupling strength. Only qubit B can emit, with jump operator

$$L = \sqrt{\gamma}\sigma_-^B. \tag{IV.50}$$

A schematic of the system is presented in Fig. IV.4a, and an example of a stochastic trajectory for the expectation value of the number operator on both qubits in Fig. IV.4b. Whenever a jump takes place, the state of qubit B is completely reset, but the state of qubit A is not. Indeed, qubit A acts as a quantum memory for the process, and prevents the renewal factorisation. We are interested in estimating the jump rate γ along the process. We use the monitoring operator method to compute the Fisher information on trajectories; the results are shown in Fig. IV.4c. The evolution has been calculated with the Gillespie algorithm, discussed in Chapter V. We note that the Fisher information scales linearly with time (or number of jumps), as expected for a process with exponentially vanishing temporal correlations. The Fisher information is monotonically increasing with time. However, this is only true at the level of the average over all trajectories. On each individual trajectory, the quantity $\mathrm{Tr}[\xi_{\omega_{1:N}}]^2$ (connected to the Fisher information via Eq. (IV.42)) has no monotonicity constraints. We also performed maximum likelihood estimation on a number of trajectories, by using the monitoring operator method explained above. In Fig. IV.4d, we compare MLE with the Cramér-Rao bound. While perfect saturation is not perfectly achieved, since only a finite amount of data is available, the MSE of the estimator is close to the bound.

C. Loglikelihood ratio on trajectories for hypothesis testing

Let us now consider hypothesis testing tasks for non-renewal quantum jump processes. For the error rates, we need the relative entropy between the string probability densities according to the different hypotheses. As previously observed, a direct computation would require integrating over the extremely large space of all measurement records. Alternatively, we can integrate on quantum trajectories. Note that the relative entropy $D_{\mathrm{KL}}(\mathrm{Pr}_0||\mathrm{Pr}_1)$ is the average on trajectories of the loglikelihood ratio $L(\omega_{1:N})$.

We consider again the unnormalised conditional states, but we label them with a superscript to indicate the reference hypothesis. $\tilde{\rho}^{(0)}_{\omega_{1:N}}$ is the unnormalised conditional state after the measurement record $\omega_{1:N}$, assuming the null hypothesis; $\tilde{\rho}^{(1)}_{\omega_{1:N}}$ is the same for the alternative hypothesis. The log-likelihood ratio

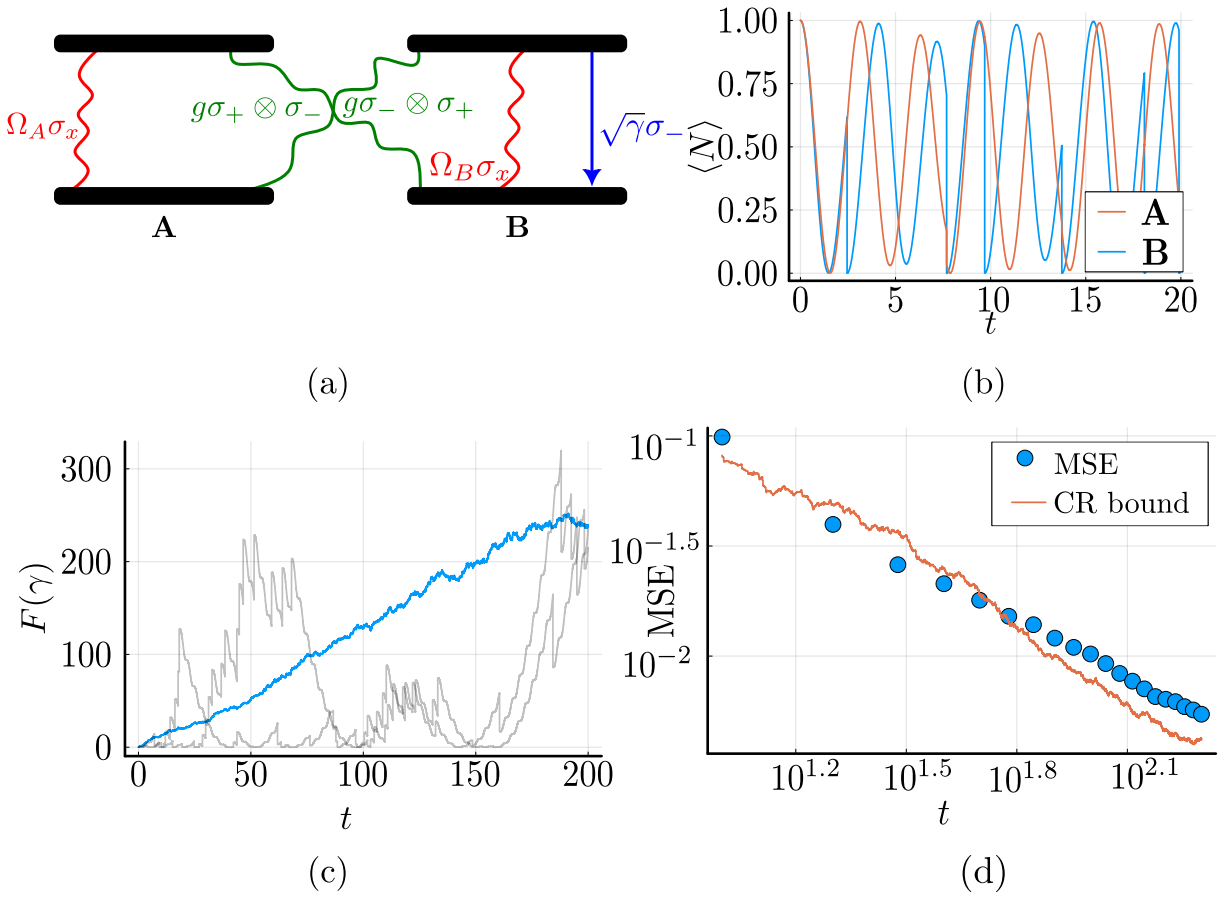

FIG. IV.4. (a) Scheme of the two qubits system. (b) Stochastic quantum trajectory for the expectation value of the number operator on the two qubits (to make the plot clearer, here $g = 0.1$). (c) Gillespie-Fisher simulation of the Fisher information for γ. The average over 500 trajectories is highlighted with the red line, and some trajectories for $\mathrm{Tr}[\xi_{\omega_{1:N}}]^2$ are represented in the background. (d) Comparison between MLE estimation and the Cramér-Rao bound for γ, as a function of the evolution time, for 500 trajectories. The apparent violation of the bound for small t is due to statistical fluctuations of both the bound (that is computed numerically) and the variance of the estimator. Parameters: $\gamma = 0.4$, $\Omega_A = 1$, $\Omega_B = 1$, $g = 0.01$. Reproduced from [4].

is, therefore,

$$L(\omega_{1:N}) = \log \frac{\mathrm{Tr}\left[\tilde{\rho}^{(1)}_{\omega_{1:N}}\right]}{\mathrm{Tr}\left[\tilde{\rho}^{(0)}_{\omega_{1:N}}\right]}. \tag{IV.51}$$

We consider the evolution equation for the unnormalised conditional state, Eq. (IV.43), for both the null

and the alternative hypotheses, and we obtain an evolution equation for the log-likelihood ratio:

$$L(\omega_{1:N}) = \log \frac{\mathrm{Tr}\left[\mathcal{J}^{(1)}_{k_N} e^{\mathcal{L}^{(1)}_0} \tilde{\rho}^{(1)}_{\omega_{1:N-1}}\right]}{\mathrm{Tr}\left[\mathcal{J}^{(0)}_{k_N} e^{\mathcal{L}^{(0)}_0} \tilde{\rho}^{(0)}_{\omega_{1:N-1}}\right]} = L(\omega_{1:N-1}) + \frac{\mathrm{Tr}\left[\mathcal{J}^{(1)}_{k_N} e^{\mathcal{L}^{(1)}_0} \rho^{(1)}_{\omega_{1:N-1}}\right]}{\mathrm{Tr}\left[\mathcal{J}^{(0)}_{k_N} e^{\mathcal{L}^{(0)}_0} \rho^{(0)}_{\omega_{1:N-1}}\right]}. \tag{IV.52}$$

The relative entropy relevant for the Chernoff-Stein exponent is

$$D_{\mathrm{KL}}\left(\mathrm{Pr}_1(\omega_{1:N})||\mathrm{Pr}_0(\omega_{1:N})\right) = \mathbb{E}_{1,\omega_{1:N}} L(\omega_{1:N}). \tag{IV.53}$$

Therefore, one has to evolve the state according to the dynamics for the alternative hypothesis, but keep track also of the evolution of the conditional state according to the null hypothesis. An efficient way of evolving the log-likelihood ratio is discussed in Chapter V, together with a case study for an application.

Chapter V

Numerical simulations: the quantum Gillespie algorithm

In the previous Chapter, we introduced methods for dealing with estimation theory and hypothesis testing on non-renewal measurement records. Since integrating over the very highly-dimensional space of measurement records has a prohibitive complexity scaling, one is forced to use, instead, quantum trajectories-based methods.

It is, therefore, paramount to have efficient ways of simulating quantum trajectories. In this Chapter, we discuss this problem from a numerical perspective. After reviewing the standard methods, we develop a new simulation technique, termed the quantum Gillespie algorithm. Our method is inspired to the Gillespie algorithm [135, 136], which is routinely used to simulate classical rate equations. Our new algorithm is particularly efficient when one needs a large number of trajectories for a relatively small system. This is very often the regime of interest for metrological applications

To illustrate the practical application of this algorithm, we discuss three case studies in open quantum systems theory and quantum thermodynamics.

References and collaboration information

The working principles of the Gillespie algorithm are discussed in Ref. [3], which is the main reference for this Chapter. I was first author of the paper, written in collaboration with F.C. Binder and G.T. Landi, and developed the Julia implementation of the Gillespie algorithm (with its variants). The Mathematica implementation was developed by G.T. Landi [137].

The single atom maser model is presented in Ref. [4]. The coupled quantum dots in optical cavity will be published in Ref. [6], but it is still unpublished at the thesis submission stage. The section on the connection between the Fisher information and quantum thermodynamics' uncertainty relations is included in in Ref. [5]. The work in Ref. [5] was co-authored with S.V. Moreira, F.C. Binder, A. Candeloro and M.T. Mitchison. As second author, I was responsible for the numerical simulations.

V.1. SIMULATING QUANTUM JUMP TRAJECTORIES

Given a continuously monitored system, we want to generate measurement records $\omega_{1:N}$, sampled with the correct probability. We remark that, in principle, we only care for the measurement record, not for the conditional state along the trajectory at all times. Two methods are commonly employed in the literature to approach this problem.

A. The SS-QJU method

A first possibility is to introduce a very small (theoretically infinitesimal) time interval dt, and simulate the evolution by applying, at each timestep, one of the Kraus operators of Eq. (II.12), with the appropriate probability. This method uses a step-wise approach, and we refer to it as step-by-step QJU (SS-QJU) in the following. Algorithmically, the method proceeds as described in Alg. 1. The SS-QJU

Algorithm 1: Step-by-step quantum jump unraveling

Input: Hamiltonian H and jump operators $\{L_k\}$; initial state of the dynamics ρ_0; length of the infinitesimal time interval dt; number of trajectories
Prepare the Kraus operators:

$$M_0 = \mathbb{I} - iHdt - \frac{1}{2}\sum_k L_k^\dagger L_k dt \qquad M_k = L_k\sqrt{dt}$$

for each trajectory **do**
  Set $\rho = \rho_0$
  Setup an empty list for the measurement record **for** each timestep **do**
    Sample the appropriate Kraus operator to apply, with probability $\Pr(j) = \mathrm{Tr}\left[M_j \rho M_j^\dagger\right]$
    Update the state and renormalise: $\rho \leftarrow M_j \rho M_j^\dagger / \Pr(j)$
    if $j \neq 0$ **then**
      Store the time and the channel in the measurement record
    end
  end
end
Output: the measurement records for all the considered trajectories

method is available in most quantum libraries (`smesolve` in QuTip [138] and `stochastic.master` in QuantumOptics.jl [139]). However, it is not particularly efficient. The probability of getting $j \neq 0$ at the sampling stage vanishes with $dt \to 0$: no jump takes place in the vast majority of steps! The SS-QJU algorithm spends, therefore, most of the time computing the evolution in steps in which no jump is taking

place.

B. The Monte Carlo wavefunction method

A more efficient technique is the Monte Carlo wavefunction method [140–144] (MCW), that allows one to work with pure states, with significant advantages in terms of time and memory complexity. A pseudocode description of the algorithm is given in Alg. 2.

Algorithm 2: Monte Carlo wavefunction

Input: Hamiltonian H and jump operators $\{L_k\}$; initial state of the dynamics $|\psi_0\rangle$;
number of trajectories;
final time T
Prepare the no-jump effective Hamiltonian

$$H_e = H - \frac{i}{2}\sum_k L_k^\dagger L_k$$

for each trajectory **do**
  Set $|\psi\rangle = |\psi_0\rangle$
  Set the total time $t = 0$
  Setup an empty list for the measurement record.
  while $t < T$ **do**
    Sample r uniformly at random from $(0, 1)$
    Solve the equation

$$||e^{-iH_e\tau}|\psi\rangle|| = r$$

    for τ.
    τ is the time of the jump. Evolve the state until τ and renormalise:

$$|\psi\rangle \leftarrow e^{-iH_e\tau}|\psi\rangle / ||e^{-iH_e\tau}|\psi\rangle||$$

    Sample the jump channel with weights $w_j = \langle\psi|L_k^\dagger L_k|\psi\rangle$
    Update $t \leftarrow t + \tau$
    Update the state $|\psi\rangle \leftarrow L|\psi\rangle / ||L|\psi\rangle||$
    Store the time and the channel in the measurement record.
  end
end
Output: the measurement record for all the considered trajectories

The MCW algorithm assumes that the conditional dynamics of the system preserves purity: if the initial state is pure, the conditional state along any trajectory will always be pure as well. This is the case when all the channels are monitored, and the experimenter is able to account for each of them separately. Formally, this corresponds to the finest unraveling of a master equation. Provided that the

dynamics preserve purity, the algorithm can easily be generalised to the case of an initially mixed state by evolving separately the pure components of any ensemble decomposition of the state itself. In this situation, MCW is able to leap directly from a jump to the next one, by exploiting the reduction in norm of the pure state during the no-jump dynamics.

There are, however, situations in which it is not possible to work with pure states alone. This usually happens when one is considering a coarse grained master equation. For instance, an experimenter may not have the possibility to monitor some channels (*partial monitoring*), or may not be able to distinguish among some of the channels (*channel merging*). In the former case, the no-jump evolution superoperator $e^{\mathcal{L}_0\tau}$ has rank larger than one, and purity is lost during the no-jump conditional evolution. In the latter situation, some of the jump superoperators $\mathcal{J}_k$ have rank larger than one, and purity is lost on the corresponding jumps.

In principle, a measurement record for coarse-grainings of a master equation can be obtained by post-processing the measurement record for the finest unraveling. In practice, though, this is not a viable option, as the number of required trajectories scales exponentially with time (for partial monitoring) or number of jumps (for channel merging). In this scenarios, one can use the SS-QJU method detailed above, despite its lower efficiency.

Usually, one fall then back to the SS-QJU method detailed above, that is well able to deal with non-purity-preserving dynamics, but this comes at the expense of efficiency.

V.2. THE GILLESPIE ALGORITHM

To overcome these difficulties, we introduce a new algorithm, that is able to handle non-purity-preserving dynamics, while at the same time being able to leap from one jump to the next, just like MCW does. The algorithm is inspired to the well-known Gillespie algorithm [135, 136], routinely employed to evolve classical master equations in classical statistical mechanics. The algorithm is based on directly sampling the waiting time distribution.

Assume that the system is in state ρ at a given time. The waiting time distribution given ρ represents the probability density that a jump occurs after time τ, in channel k, and can be computed as

$$W(\tau, k|\rho) = \mathrm{Tr}\left[\mathcal{J}_k e^{\mathcal{L}_0\tau}\rho\right]. \tag{V.1}$$

Note that this definition is much more general than the one in Sect. IV.2. Indeed, it is conditioned on a generic state ρ, instead of a jump [145, 146].

In some cases, it may be possible to compute an analytical expression for $W(\tau, k|\rho)$. One can then sample directly from it with the inversion method [98]. Most frequently, though, an analytical expression for the waiting time distribution cannot be obtained. Taking inspiration from the classical Gillespie algorithm [135, 136], we separately sample the jump time and jump channel. Let us decompose $W(\tau, k|\rho) = P(k|\tau, \rho)W(\tau|\rho)$, where

$$W(\tau|\rho) = \mathrm{Tr}\left[\mathcal{J}e^{\mathcal{L}_0\tau}\rho\right], \tag{V.2}$$

where $\mathcal{J} = \sum_k \mathcal{J}_k$. Given $\tilde{\rho} = \mathcal{J}e^{\mathcal{L}_0\tau}\rho$, the conditional probability of picking jump channel k, conditioned on the fact that a jump indeed takes place, is given by

$$P(k|\tau, \rho) = P(k|\tilde{\rho}) = \frac{\mathrm{Tr}\left[\mathcal{J}_k\tilde{\rho}\right]}{\mathrm{Tr}\left[\mathcal{J}\tilde{\rho}\right]}. \tag{V.3}$$

The problem of sampling from the waiting time distribution $W(\tau, k|\rho)$ is, therefore, mapped to the separate sampling of $W(\tau|\rho)$ and $P(k|\tau, \rho)$. While sampling from $P(k|\tau, \rho)$ is straightforward, $W(\tau|\rho)$ warrants some more explanation. We observe that the factor $\mathcal{J}e^{\mathcal{L}_0\tau}$ is trajectory-independent, as it does not contain information about the conditional state. We can precompute the superoperator $\mathcal{J}e^{\mathcal{L}_0\tau} = \mathcal{Q}(\tau)$ for a fine-grained enough set of times `ts` in advance. To reconstruct the probability density $W(\tau|\rho)$, we compute the probability density $\mathrm{Tr}\left[\mathcal{Q}(\tau)\rho\right]$ for each $\tau \in$ `ts`, and sample one of the elements $t \in$ `ts` from it. We give a pseudocode implementation of this algorithm in Alg. 3.

We now discuss the performances of the Gillespie algorithm for the simulation of quantum jump trajectories in more detail. We already mentioned that MCW is to be preferred when the evolution preserves purity. Here, we compare the Gillespie algorithm with SS-QJU in situations in which MCW is not a viable option, e.g. in presence of partial monitoring and/or channel merging.

In most cases of interest, there is a significant timescale difference between the Hamiltonian evolution and the jumps, with the former being much faster. In SS-QJU, the evolution needs to be able to follow the minute details of the Hamiltonian; therefore, one is forced to choose a very small time interval dt to guarantee the convergence of the solution of the differential equation governing the dynamics. In the Gillespie algorithm, convergence is always guaranteed by the form of the exponential propagator.

Algorithm 3: Gillespie algorithm for quantum jump trajectories

Input: Hamiltonian H;
list of monitored jumps operators $\{L_k\}$, list of unmonitored jump operators $\{S_k\}$;
initial state density matrix ρ_0;
final time T; timestep dt;
number of trajectories;
Prepare $J = \sum_k L_k^\dagger L_k$, $\mathcal{J} = \mathrm{Vec}(J)$ and the no-jump Lindbladian

$$\mathrm{Vec}(\mathcal{L}_0) = -i\mathbb{I}\otimes H + iH^T\otimes\mathbb{I} + \sum_l \left(S_l^*\otimes S_l - \frac{1}{2}\mathbb{I}\otimes S_l^\dagger S_l - \frac{1}{2}\left(S_l^\dagger S_l\right)^T\otimes\mathbb{I}\right) + \\ + \sum_k \left(-\frac{1}{2}\mathbb{I}\otimes L_k^\dagger L_k - \frac{1}{2}\left(L_k^\dagger L_k\right)^T\otimes\mathbb{I}\right)$$

$$\mathrm{Vec}(\mathcal{L}_0^\dagger) = i\mathbb{I}\otimes H - iH^T\otimes\mathbb{I} + \sum_l \left(S_l^T\otimes S_l^\dagger - \frac{1}{2}\left(S_l^\dagger S_l\right)^T\otimes\mathbb{I} - \frac{1}{2}\mathbb{I}\otimes\left(S_l^\dagger S_l\right)\right) + \\ + \sum_k \left(-\frac{1}{2}\left(L_k^\dagger L_k\right)^T\otimes\mathbb{I} - \frac{1}{2}L_k^\dagger L_k\right)$$

for all times of the dt-spaced net up to T **do**
  $V_\tau = \exp\left(\mathrm{Vec}(\mathcal{L}_0)\tau\right)$
  $\mathcal{Q}_\tau = \exp\left(\mathrm{Vec}(\mathcal{L}_0^\dagger)\tau\right)\mathcal{J}$
end
for all trajectories **do**
  Fix the initial state $\rho = \rho_0$ and the initial time $t = 0$ **while** $t < T$ **do**
    for every $\mathcal{Q}$ **do**
      compute the WTD component $W(\tau|\rho) = \mathrm{Tr}\left[\mathcal{Q}_\tau\rho\right]$
    end
    sample a time $\bar{\tau}$ for the next jump from $W(\tau|\rho)$;
    increment the total time: $t \leftarrow t + \bar{\tau}$;
    evolve the state to immediately before the jump, $\tilde{\rho} = \mathrm{Unvec}(V_{\bar{\tau}}\mathrm{Vec}(\rho))$.
    for every monitored jump channel **do**
      $P(k|\tilde{\rho}) = \mathrm{Tr}[\mathcal{J}_k\rho]/\mathrm{Tr}[\mathcal{J}\rho]$
    end
    sample a jump channel $\bar{k}$ from $P(k|\tilde{\rho})$;
    evolve the state with the jump and renormalise, $\rho \leftarrow \mathrm{Tr}[\mathcal{J}_k\tilde{\rho}]/\mathrm{Tr}[\mathcal{J}_k\tilde{\rho}]$
    store the time $\bar{\tau}$ and the channel $\bar{k}$ in the measurement record.
  end
end
Output: the measurement record for all the considered trajectories

It therefore suffices to choose a small enough dt that $W(\tau, k|\rho)$ is smooth. The most relevant precision driver for the Gillespie algorithm is, instead, the largest time in the list `ts`. Indeed, the algorithm needs to reconstruct the waiting time distribution as faithfully as possible. If too small a maximal time is

chosen in `ts`, the time separation between the jumps becomes artificially compressed.

Though, the Gillespie algorithm cannot be run for very short final times T. In fact, the algorithm needs to faithfully reproduce the shape of the bulk of the waiting time distribution. Conversely, SS-QJU can be run for arbitrarely short times.

The power of the Gillespie algorithm lies in the precomputation stage. In fact, most of the numerically heaviest quantities are not trajectory-dependent. We can therefore compute them in advance and reuse them on each trajectory. Clearly, precomputation hides a tradeoff between memory and time. Indeed, the Gillespie algorithm is most suited when one needs a very large number of trajectories for relatively small systems. On the contrary, SS-QJU remains the algorithm of choice when dealing with many-body systems, because of the precomputation memory requirements.

By construction, the Gillespie algorithm does not yield the conditional state at all times: only the states immediately before and after the jumps are accessible. On the contrary, SS-QJU easily allows to retrieve the conditional state at all times. Hence, the Gillespie algorithm is particularly appropriate when one wants to only simulate the measurement record, rather than the full state history. In Ref. [3], we suggested a possible implementation to retrieve the state at all times after running the Gillespie algorithm. The algorithm is reproduced in Alg. 4. However, this partially undoes the main advantages of the Gillespie algorithm, as it requires a step-by-step approach.

Algorithm 4: Finding the state at fixed times along a trajectory

Input: $[t^*]$ list of times at which V is known
$[t]$ list of times at which the state has to be computed;
$[V]$ list of evolution superoperators at times specified in $[t^*]$;
$[(j_j, k_j, \rho_j)]$ vector of results and states for a single trajectory;
Set to simplify the notation a final jump at time $+\infty$;
for each jump (t_j, k_j, ρ_j) along the trajectory **do**
  Determine the set $[\bar{t}]$ of times such that $t_j \leq \bar{t} < t_{j+1}$;
  for each $\bar{t}$ in $[\bar{t}]$ **do**
    Find the closest time t_{cl} in $[t^*]$ to $\bar{t} - t_j$;
    $\rho = \frac{\mathrm{Unvec}(V[t_{cl}]\mathrm{Vec}(\rho)}{\mathrm{Tr}[\mathrm{Unvec}(V[t_{cl}]\mathrm{Vec}(\rho)]}$
  end
end
Output: the vector of ρ states at all times in $[t]$

A. Variants: the Gillespie algorithm for estimation and hypothesis testing

Let us now turn, once more, to the original problem of estimation and hypothesis testing on quantum jump trajectories. In Chapter IV, we introduced the stochastic integration approach for the Fisher information (the monitoring operator), as well as for the likelihood ratio. We introduce variants of the Gillespie algorithm that take these into account. We start with hypothesis testing, as the modifications are more straightforward.

To evolve the log-likelihood ratio according to Eq. (IV.52), one needs to keep track of the conditional state according to both the null and the alternative hypothesis. However, if we are interested in an average over the alternative hypothesis, as it is the case, for instance, to compute the relative entropy of Eq. (IV.53), the evolution has to follow the alternative hypothesis. One therefore carries forward the simulation with the standard Gillespie algorithm, selecting jump times and channels according to the alternative hypothesis. One then evolves the conditional state for the null hypothesis using the same jump times and channels.

The generalisation for estimation tasks is more nuanced. To compute the monitoring operator evolution according to Eq. (IV.45), one needs the derivatives both of $\mathcal{L}_0$ and of all the jump operators. We compute the derivatives numerically, by considering the change in the Hamiltonian and the jump operators due to infinitesimal displacements in the parameter. Alongside $Q(\tau)$, for all the times in the list `ts` and all the jump channels k, we precompute the superoperator $\Delta(k,\tau) = \partial_\theta M_k e^{\mathcal{L}_0\tau}$. A pseudocode implementation is given in Alg. 5.

V.3. CASE STUDY: SINGLE ATOM MASER MODEL

The single atom maser [147], or the very similar photon box [148], is one of the simplest systems used to study quantum measurements and feedback. While the first theoretical descriptions and experimental implementations date back to the 1980s [149–151], it is still object of active study [152–154].

A beam of Rydberg atoms transverses a high quality factor cavity. The state of the atom is initially $|\psi\rangle = \alpha\,|e\rangle + \beta\,|g\rangle$, where α and β are amplitudes, such that $|\alpha|^2 + |\beta|^2 = 1$. The atomic beam is very sparse, so that, on average, there is at most one atom in the cavity at any time. The internal degree of

Algorithm 5: Gillespie algorithm variant for the monitoring operator

Input: Same as Alg. 3, and additionally:
displaced Hamiltonians H_+ and H_-, corresponding to parameters $\theta + d\theta$ and $\theta - d\theta$;
displaced monitored jumps operators $\{M_{k,+}\}$ and $\{M_{k,-}\}$;
displaced unmonitored jumps operators $\{S_{k,+}\}$ and $\{S_{k,-}\}$;
derivative increment $d\theta$;
initial monitoring operator ξ_0.
Define $J = \sum_k M_k^\dagger M_k$ and the displaced versions J_+, J_-;
define the no-jump Lindbladian $\mathrm{Vec}(\mathcal{L}_0)$ and its adjoint $\mathrm{Vec}(\mathcal{L}_0^\dagger)$ as in Alg. 3, with their displaced versions;
precompute V_τ and $\mathcal{Q}_\tau$ as in Alg. 3; compute $\dot{V}_\tau$, the derivative of V_τ with respect to θ;
compute $\{\dot{M}_k\}$, the derivative of $\{M_k\}$ w.r.t. θ;
compute $\Delta_{k,\tau}$, the derivative of $M_k V_\tau$ w.r.t. θ.
for all trajectories **do**
  Fix the initial state $\rho = \rho_0$ and the initial monitoring operator $\xi = \xi_0$, set $t = 0$ **while** t ¡ T **do**
    Proceed as in Alg. 3 to compute the evolved state ρ_{new};
    Update the monitoring operator on jump in channel $\bar{k}$ at time $\bar{\tau}$ as

$$\xi \leftarrow \frac{1}{\mathrm{Tr}[\mathcal{J}_{\bar{k}} e^{\mathcal{L}_0 \bar{\tau}} \rho]} \left(M_{\bar{k}} V_{\bar{\tau}} \xi V_{\bar{\tau}}^\dagger M_{\bar{k}}^\dagger + \Delta_{\bar{k},\bar{\tau}} \rho V_{\bar{\tau}}^\dagger M_{\bar{k}}^\dagger + M_{\bar{k}} V_{\bar{\tau}} \rho \Delta_{\bar{k},\bar{\tau}}^\dagger \right)$$

    Update the state, $\rho \leftarrow \mathcal{J}_k e^{\mathcal{L}_0 \tau} \rho / \mathrm{Tr}[\mathcal{J}_k e^{\mathcal{L}_0 \tau} \rho]$
  end
end

freedom of the atom couples with the cavity electromagnetic field via the Jaynes-Cummings Hamiltonian

$$H = \frac{g}{2} \left[a \otimes \sigma_+ + a^\dagger \otimes \sigma_- \right]. \tag{V.4}$$

Here, a is the bosonic operator for the field, $\sigma_+ = |e\rangle\langle g|$, $\sigma_- = \sigma_+^\dagger$, and g is the coupling strength between the field and the internal degree of freedom of the atom flying by. The atom spends time τ inside the cavity and, upon exiting, it is measured in the $\{|e\rangle, |g\rangle\}$ basis. The train of incoming atoms in the cavity acts as a collisional environment. We describe the evolution of the cavity field by tracing out the atoms. We acquire information on the system by performing strong measurements on the atoms, corresponding to weak measurements on the cavity field. In the following, we compute the Fisher information for the estimation of the coupling parameter g.

We start by considering the unitary operator for the joint evolution of atom and field [147]:

$$\begin{aligned} U(\tau) = &\cos\left(g\tau\sqrt{n+1}\right) |e\rangle\langle e| + \cos\left(g\tau\sqrt{n}\right) |g\rangle\langle g| + \\ &- i \frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n}} a\, |e\rangle\langle g| - i a^\dagger \frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}} |g\rangle\langle e|, \end{aligned} \tag{V.5}$$

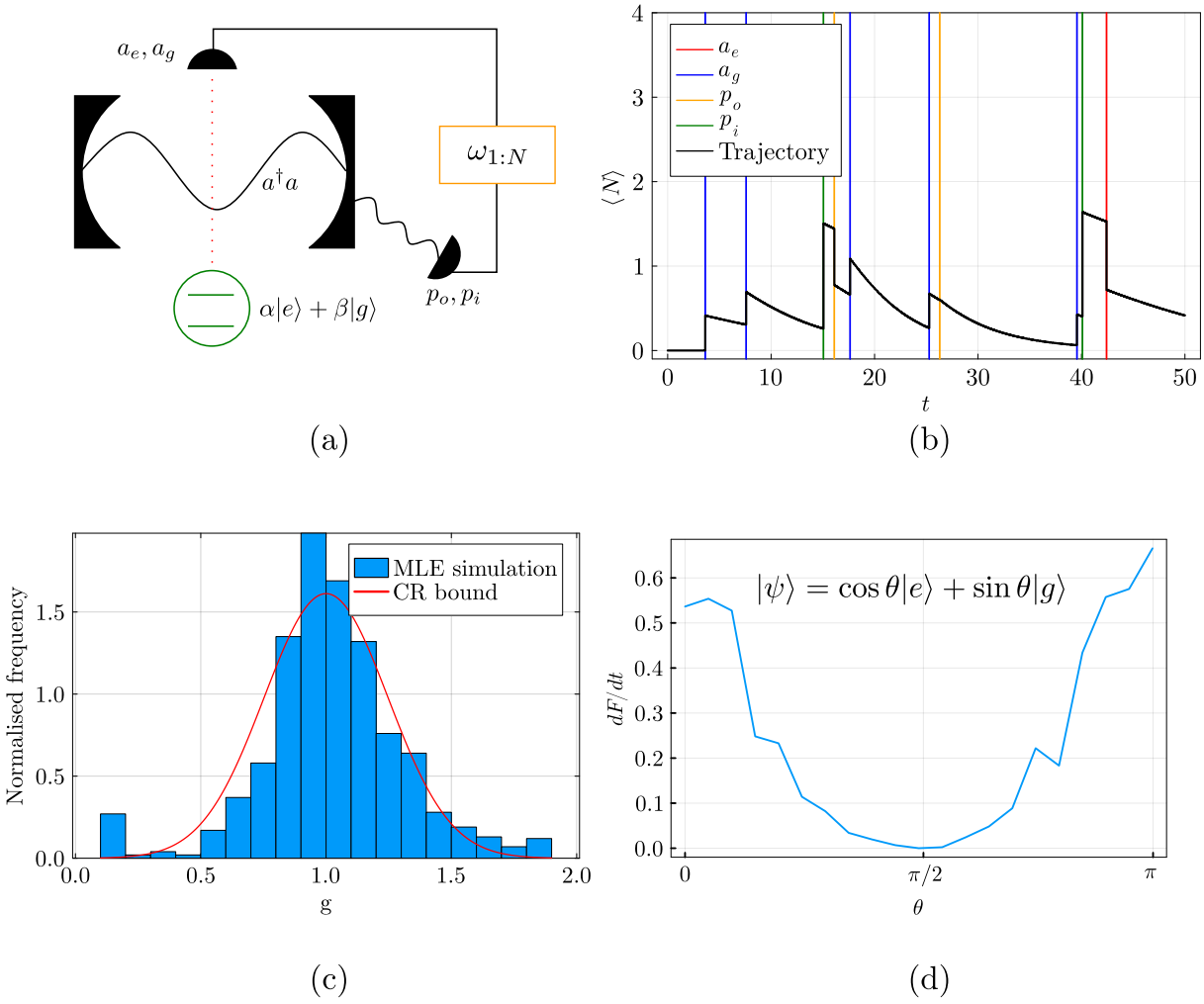

FIG. V.1. (a) scheme of the single atom maser system. (b) example of a trajectory for the expectation value of the number operator of the field, with the jumps highlighted. (c) maximum likelihood estimation task for g, compared with the Cramér-Rao bound. (d) Fisher information rate in time, as a function of the initial state of the atomic beam. Parameters: $g = 1$, $\tau = 1$, $\theta = \pi/4$ for (c), $\bar{n} = 0.1$, e.m. field capped at the first 5 states of the Fock space, 1000 trajectories for (c), 100 trajectories for each value of θ for (d). Reproduced from [4].

where $n = a^\dagger a$ is the number operator for the cavity field. We describe the effect on the cavity field of the strong measurement of the atomic state, when the atom exits the cavity[3]. Assume a factorised initial state between cavity field and atom, $\rho_0 = \rho_{\text{em}} \otimes |\psi\rangle\langle\psi|$. When the atom exits the cavity, but before its measurement, the overall state is $U(\tau)\rho_{\text{em}} \otimes |\psi\rangle\langle\psi| U(\tau)^\dagger$.

Assume now that the measurement on the atom yields that the atom is in the excited state. The

[3] We largely follow Ref. [147] here. However, the derivation for generic initial state of the atom ψ has not been presented, to the best of our knowledge, in the literature, therefore we discuss it in some detail.

backaction on the field state yields the (unnormalised) state

$$\begin{aligned}\rho_{\rm em}^{(e)} &= \mathrm{Tr}_{\rm at}\left[|e\rangle\langle e|\, U(\tau)\rho_{\rm em}\otimes|\psi\rangle\langle\psi|\, U(\tau)^\dagger\right] = \\ &= |\alpha|^2\cos\left(g\tau\sqrt{n+1}\right)\rho_{\rm em}\cos\left(g\tau\sqrt{n+1}\right)+ \\ &+ |\beta|^2\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}a\rho_{\rm em}a^\dagger\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}+ \\ &- i\alpha^*\beta\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}a\rho_{\rm em}\cos\left(g\tau\sqrt{n+1}\right)+ \\ &+ i\beta^*\alpha\cos\left(g\tau\sqrt{n+1}\right)\rho_{\rm em}a^\dagger\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}.\end{aligned} \tag{V.6}$$

If, on the other hand, the atom is measured to be in the ground state, then the state of the field is updated to:

$$\begin{aligned}\rho_{\rm em}^{(g)} &= \mathrm{Tr}_{\rm at}\left[|g\rangle\langle g|\, U(\tau)\rho_{\rm em}\otimes|\psi\rangle\langle\psi|\, U(\tau)^\dagger\right] = \\ &= |\alpha|^2 a^\dagger\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}\rho_{\rm em}\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}a+ \\ &+ |\beta|^2\cos\left(g\tau\sqrt{n}\right)\rho_{\rm em}\cos\left(g\tau\sqrt{n}\right)+ \\ &+ i\alpha^*\beta\cos\left(g\tau\sqrt{n}\right)\rho_{\rm em}\frac{\sin\left(\tau\sqrt{n+1}\right)}{\sqrt{n+1}}a+ \\ &- i\beta^*\alpha a^\dagger\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}\rho_{\rm em}\cos\left(g\tau\sqrt{n}\right).\end{aligned} \tag{V.7}$$

The (strong) measurement on the atom acts a weak measurement on the field [109]. The corresponding jump operators are

$$\begin{aligned}L_{Ae} &= \alpha\cos g\tau\sqrt{n+1} - i\beta\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}}a; \\ L_{Ag} &= \alpha a^\dagger\frac{\sin\left(g\tau\sqrt{n+1}\right)}{\sqrt{n+1}} + i\beta\cos\left(g\tau\sqrt{n}\right),\end{aligned} \tag{V.8}$$

where the former corresponds to having measured an atom in the excited state, and the latter to having measured it in the ground state. Furthermore, the cavity is coupled, with strength γ, to an external thermal bath, with bosonic occupation number $\bar{n}$; the corresponding jump operators are

$$L_{Ci} = \sqrt{\gamma\bar{n}}a^\dagger \quad ; \quad L_{Co} = \sqrt{\gamma\left(\bar{n}+1\right)}a. \tag{V.9}$$

Finally, the field Hamiltonian is simply the free $H = \omega/2n$, where ω is the field frequency. An illustration of the model is given in Fig. V.1a, and an example of a trajectory for the expectation number of the number operator of the cavity in Fig. V.1b.

The single atom maser is not a renewal system, as jumps do not reset the state of the cavity completely. We employ the Gillespie algorithm, in its estimation variant, to compute the Fisher information along trajectories. We also use the monitoring operator formalism to perform maximum likelihood estimation on trajectories. An histogram, showing the saturation of the Cramér-Rao bound, is shown in Fig. V.1c.

It is particularly interesting to consider what is the optimal state $|\psi\rangle$ in which to prepare the Rydberg atoms for optimal estimation of g. We define $\theta = \arccos|\alpha|$, and let $|\psi\rangle = \cos\theta\,|e\rangle + \sin\theta\,|g\rangle$. In Fig. V.1, we show the Fisher information rate for g, computed by linear fitting of the trajectories average, as a function of θ. The highest Fisher information rate is obtained for $\theta \sim 0$ and $\theta \sim \pi$, that is, when the initial state of the atoms in the beam is close to $|e\rangle$. On the contrary, next to no Fisher information about g is observed for $\theta \sim \pi/2$, when the atoms are initially close to the ground state $|g\rangle$. This fact can be understood as follows. The field is such that there is, on average, much less than a photon in the cavity. Assuming a pure state for the sake of simplicity, we can expand it in the Fock basis as

$$|\phi\rangle \sim (1-\varepsilon)\,|0\rangle + \varepsilon\,|1\rangle\,, \tag{V.10}$$

with $\varepsilon \ll 1$. If we apply the Jaynes-Cummings unitary up to first order in τ, and the initial state of the atom is $|\psi\rangle = |g\rangle$, then the overall state is mapped to

$$|\Psi\rangle \sim (1-\varepsilon)\,|0\rangle \otimes |g\rangle + \varepsilon g/2\,|0\rangle \otimes |e\rangle\,, \tag{V.11}$$

where the state bearing the g-dependence is suppressed by the $\epsilon \ll 1$ factor. On the other hand, if the initial state of the atom is $|\psi\rangle = |e\rangle$, then

$$|\Psi\rangle \sim |0\rangle \otimes |e\rangle - i\frac{g}{2}\tau\,|1\rangle \otimes |g\rangle\,, \tag{V.12}$$

where the dependence on g is not suppressed. This corresponds to the following intuitive fact: an interaction between atom and cavity can happen only if at least one of the following conditions is satisfied: (i) the state of the atom is not $|g\rangle$; (ii) the state of the field is not $|0\rangle$. Since the state of the field is close to $|0\rangle$, then the more the atoms are excited, the more interactions will be observed, and each interaction

carries the factor g. If no interaction takes place, then the g-dependence is not picked up.

V.4. CASE STUDY: COUPLED QUANTUM DOTS IN AN OPTICAL CAVITY

We consider in detail a system composed by two quantum dots in an optical cavity. The setup, which is represented in Fig. V.2a, has been studied both in the theoretical [155, 156] and experimental [157, 158] literature. Two coupled quantum dots experience Coulomb repulsion, that ensures single occupancy: the two dots cannot be occupied at the same time. A possible basis of the system is $\{|0\rangle, |L\rangle, |R\rangle\}$, where $|0\rangle$ represents both dots being empty, $|L\rangle$ the left dot being occupied, and $|R\rangle$ the right dot being occupied. It is also convenient to define a rotated basis (*common* basis of the dots), $\{|0\rangle, |g\rangle, |e\rangle\}$, where

$$|g\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \qquad |e\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}. \tag{V.13}$$

The two dots are coupled with an exchange Hamiltonian

$$H_{\mathrm{DQD}} = g_{\mathrm{DQD}} \left(|L\rangle \langle R| + |R\rangle \langle L| \right). \tag{V.14}$$

We assume the energy associated with the occupancy of each of the dots is zero. The parameter g_{DQD} represents the coupling strength. The dots are coupled with an optical cavity. Let a be the bosonic annihilation operator for the cavity electromagnetic field, and assume a Jaynes-Cummings coupling between cavity and dots. The overall Hamiltonian of e.m. field and dots is, therefore:

$$H = \omega_c a^\dagger a + H_{\mathrm{DQD}} + g_{\mathrm{DC}} \left(|g\rangle \langle e| \otimes a^\dagger + |e\rangle \langle g| \otimes a \right), \tag{V.15}$$

where g_{DC} is the strength of the Jaynes-Cummings coupling, and ω_c is the cavity frequency ($\omega_c = 2g_{\mathrm{DQD}}$ if dots and cavity are in perfect resonance with each other). Finally, there can be external coherent driving of the cavity field, of the form

$$H_{\mathrm{dr}} = \epsilon(a + a^\dagger), \tag{V.16}$$

where ϵ is the driving strength. Each of the dots is connected to a separate fermionic thermal bath, with the same coupling strength γ. To simplify the setup, we assume that the left bath is at infinite

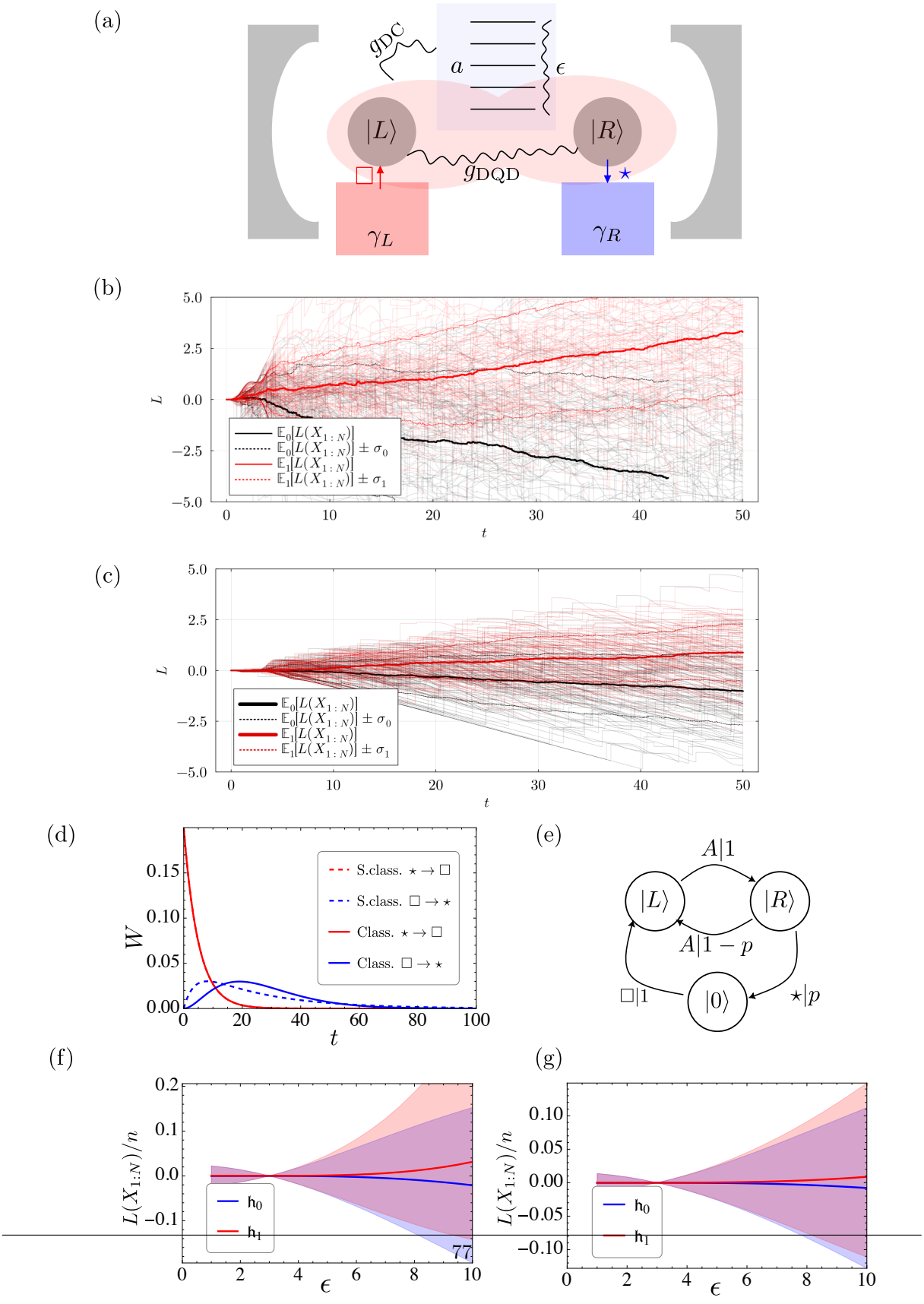

FIG. V.2. (a) schematic of the double quantum dot in optical cavity model (b) expectation of the log-likelihood ratio computed on quantum trajectories, while monitoring all channels. In the background, some of the trajectories for the log-likelihood ratio. (c) same simulation, but this time without monitoring the cavity emission channel κ. (d) waiting time distributions for the semiclassical and the classical models. (e) scheme of the classical model in the language of classical stochastic processes. The probabilities are integrated over time. (f) hypothesis testing diagram, as a function of the driving strength ϵ, for the semiclassical model. (g) the same diagram, but for the classical model. Parameters for (b-c): $g_{\rm DQD} = 0.05$, $\gamma = 0.2$, $\kappa = 0.3$, $g_{\rm DC} = 0.01$, $\epsilon = 0.3$ for the null hypothesis. Simulation with 100 trajectories, e.m. Fock space cutoff $n = 4$. Parameters for (d-g): $g_{\rm DQD} = 0.05$, $g_{\rm DC} = 0.01$, $\kappa = 7$, $\gamma = 0.2$, $\epsilon = 3$ for the alternative hypothesis.

temperature (Fermi occupancy $f_L = 1$), and the right bath is at zero temperature (Fermi occupancy $f_R = 0$). In this way, all excitations are injected via the left dot, and emitted via the right dot. The cavity has quality factor κ. The system evolves, therefore, according to the open dynamics given by Hamiltonian $H + H_{\rm dr}$, with the three jump channels

$$L_\kappa = \sqrt{\kappa} a \quad ; \quad L_\square = \sqrt{\gamma} \left| L \right\rangle \left\langle 0 \right| \quad ; \quad L_\star = \sqrt{\gamma} \left| 0 \right\rangle \left\langle R \right| . \tag{V.17}$$

The system is clearly non-renewal, because the cavity field acts as a quantum memory. In fact, jumps $\square$ and $\star$ do reset the state of the dots, but they do not reset the state of the field.

A problem of relevance is determining whether the driving is present or not. We take the presence of driving $\epsilon \neq 0$ as the null hypothesis, and the absence of it as the alternative hypothesis. We apply the hypothesis testing variant of the Gillespie algorithm to compute the expectation value of the log-likelihood ratio under both the null and the alternative hypothesis. The expectation value of the loglikelihood ratio under the alternative hypothesis on trajectories yields the Chernoff-Stein error rate. The result is shown in Fig. V.2b.

It is not always feasible to monitor all the emission channels. For instance, assume that the one has no way to monitor the field emitted from the cavity. One can expect the hypothesis testing task to become harder, as less information about the dynamics is available. This is evident in the simulation of Fig. V.2c: the log-likelihood ratios for the two averages are closer one with respect to each other.

In all these simulations, one needs to keep track of the state at all times. This is feasible if the driving is small, so that only the very first states of the harmonic oscillator ladder are populated. When the driving is strong, the memory cost of the calculation gets quickly out of hand, as more and more Fock states get populated. In the strong-driving limit, we can introduce a semiclassical model, in which we treat the electromagnetic field classically[4]. In particular, assume that $\epsilon \gg 1$, and $\kappa \gg 1$. Then, we can adiabatically eliminate the electromagnetic field. The effect of the field on the dot is then represented by the new term in the Hamiltonian:

$$H_{\rm s.class.} = H_{\rm DQD} + \left(G^* \left| g \right\rangle \left\langle e \right| + G \left| e \right\rangle \left\langle g \right| \right) , \tag{V.18}$$

[4] To the best of our knowledge, the semiclassical and classical models for the double quantum dot in cavity have not been discussed in the literature. For this reason, we expand on them a little more than what is strictly required for hypothesis testing purposes.

where the (complex) coupling parameter G is

$$G = i\frac{\epsilon g_{\mathrm{DC}}}{i\omega_c + \frac{\kappa}{2}}, \tag{V.19}$$

More details on the derivation will be given in Ref. [6]. The state of the field has been eliminated, there is no quantum memory in the system anymore, and the dynamics become renewal: the results of Sect. IV.2 apply. Of course, also the jump channel κ is not available anymore. Due to the single occupancy constraint, two jumps in the same channel cannot happen consecutively. The only non-trivial waiting time distributions are $\square \to \star$ and $\star \to \square$, as represented in Fig. V.2d. In this case, we consider the task of distinguishing between two hypotheses with different driving strength ϵ. In Fig. V.2, we represent the loglikelihood ratio, as a function of ϵ of the true model, for a fixed ϵ of the alternative hypothesis.

We can ask what is the role played by the coherences in the semiclassical model. We build an equivalent classical model, in the same spirit of the classical three level maser model of Sect. IV.2 D. We replace the Hamiltonian tunnelling term by two unmonitored jumps with the same rate Γ, with operators

$$L_\circ = \sqrt{\Gamma}\,|L\rangle\langle R| \quad ; \quad L_\triangle = \sqrt{\Gamma}\,|R\rangle\langle L|\,. \tag{V.20}$$

We obtain the classical process represented in Fig. V.2e. We have, again, a renewal system, with the waiting time distributions shown in Fig. V.2d. The log-likelihood ratio rate, as a function of ϵ for the null hypothesis, is shown in Fig. V.2g. In this case coherences are helpful for the testing task: by comparing Fig. V.2f and Fig. V.2g, in the former the expectation values of the log-likelihood ratio are more distinct rom each other.

V.5. FISHER INFORMATION AND THE KINETIC UNCERTAINTY RELATION

One of the key results of classical stochastic thermodynamics is the formulation of the thermodynamic uncertainty relation (TUR) [159] and the kinetic uncertainty relation (KUR) [160], linking the fluctuations of thermodynamic currents along a process with dissipation. Assume a system S coupled to an environment E. As the environment dissipates information coming from the system, the dynamics of the latter is irreversible. The degree of irreversibility can be expressed in terms of the entropy production [5, 161]

$$\Sigma = \Delta S_S + \Delta S_E, \tag{V.21}$$

where

$$\Delta S_S = \mathrm{Tr}\left[\rho_0 \log \rho_0\right] - \mathrm{Tr}\left[\rho_\tau \log \rho_\tau\right] \tag{V.22}$$

is the variation in Von Neumann entropy of the system until time τ, and

$$\Delta S_E = \int_0^\tau dt \sum_j \mathrm{Tr}\left[L_{k_j}\rho(t)L_{k_j}^\dagger\right] \Delta s_{k_j}, \tag{V.23}$$

is the change in entropy of the environment. For a jump channel k, Δs_k represents the entropy transferred from the system to the environment when the jump takes place. ΔS_E, then, is the average transferred entropy, taking into account the jump probability. The higher Σ is, the more irreversible the process.

In the same setup, one can also define the dynamical activity of the process [5, 162, 163] as

$$\mathcal{A} = \int_0^\tau dt \sum_j \mathrm{Tr}\left[L_{k_j}\rho(t)L_{k_j}^\dagger\right], \tag{V.24}$$

that is, the average number of jumps until time τ.

The classical thermodynamic uncertainty relation (TUR) [159] lower bounds the variance of an integrated current Θ in terms of the entropy production:

$$\frac{\mathrm{Var}(\Theta)}{\langle\Theta\rangle^2} \geq \frac{2}{\Sigma}. \tag{V.25}$$

Here, $\langle\Theta\rangle$ is the mean of the current, and its variance is $\mathrm{Var}(\Theta)$. In turn, the classical kinetic uncertainty relation (KUR) [160] yields

$$\frac{\mathrm{Var}(\Theta)}{\langle\Theta\rangle^2} \geq \frac{1}{\mathcal{A}}, \tag{V.26}$$

Both of these bounds state that reducing the variance of a current (i.e., making it more precise) comes at the cost of increasing the dissipation of the process. The latter can be expressed in terms of entropy production or dynamical activity. The link between TURs, KURs and estimation theory has been established in Refs. [71, 164, 165]. A fictional parameter is encoded as an unphysical perturbation of the dynamics, and TURs and KURs are obtained from the Cramér-Rao bound for the estimation of those parameters.

An important research effort in quantum thermodynamics has been devoted to formulating a quantum version of the TURs and KURs, after the discovery of a number of violations of the classical ones by coherent systems [130, 166–170]. Even though quantum versions of the bounds have been introduced [171–174], the exact physical reason of the violation of the classical bounds remains, to date, object of open debate. Here, that focusses on estimation theory for correlated processes, we introduce the encoding techniques that are used to map thermodynamic problems on estimation problems. We then give an example of application to the three level maser system, from Ref. [5].

Assume an open quantum system, in contact with M thermal baths. The baths will be labelled by the index k. For each bath, there are two jump channels, in and out of the system, labelled by L_k and $L_{k'}$[5].

We define a vector of real parameters $\vec{\phi}$, with the same number of components as the number of thermal baths (M). We deform the Hamiltonian and the jump operators as follows:

$$\begin{aligned} H_{\vec{\phi}} &= (1+\sum_{\alpha=1}^{M}\phi_\alpha)H; \\ L_{k,\phi_\alpha} &= \sqrt{1+\phi_\alpha}L_k. \end{aligned} \tag{V.27}$$

Counting observables are defined as functions of the measurement record of the form

$$\Phi_\alpha(\omega_{1:N}) = \sum_{j=1}^{N} w_{k_j}^{\alpha}, \tag{V.28}$$

where w_k^α is the weight associated to a jump involving bath k. The key observation here is that $\Phi_\alpha(\omega_{1:N})$ can act as an estimator for the parameter ϕ_α, as it picks up the dependence of the dynamics on ϕ_α. The estimation of the different components of $\vec{\phi}$ gets intertwined, as the precision of one influences all the others. Therefore, one has to use a vectorial form of the Cramér-Rao bound of Eq. (I.14). We define the covariance matrix $[\Xi]_{\alpha\beta} = \mathrm{Cov}[\Phi_\alpha, \Phi_\beta]$. Also, note that the vector function $\vec{\Phi}$ is, in general, a *biased* estimator for the parameters $\vec{\phi}$; this has to be taken into account in the Cramér-Rao bound. The matrix inequality is

$$J_{\vec{\Phi}}^T \Xi^{-1} J_{\vec{\Phi}} \leq F, \tag{V.29}$$

[5] We warn the reader that this is a slight change in notation with respect to the previous sections: here, two jump channels to/from the same bath are labelled with the same index k, but one of them is primed.

where F is the Fisher information matrix, and $J_{\vec{\Phi}}$ is the Jacobian of the vector function $\vec{\Phi}$ with respect to the vector of parameters $\vec{\phi}$. The Jacobian is nontrivial if the estimator is biased.

The Fisher information matrix elements are:

$$F_{ij} = -\mathbb{E}_{\omega_{1:N}}\left[\frac{\partial}{\partial\phi_i}\frac{\partial}{\partial\phi_j}\log\Pr(\omega_{1:N})\right]. \tag{V.30}$$

From the matricial bound of Eq. (V.29), one can infer many different scalar bounds. It is common to take the determinant of both sides of the inequality, yielding

$$\frac{\det(\Xi)}{[\partial_{\phi_1}\langle\Phi_1\rangle]^2\ldots[\partial_{\phi_M}\langle\Phi_M\rangle]^2} \geq \frac{1}{\det(F)}. \tag{V.31}$$

We focus on the case of two baths, with parameters vector $\vec{\phi} = (\phi_1, \phi_2)^T$, and counting operators Φ_1 and Φ_2. Explicitating the determinant, we get:

$$\frac{\mathrm{Var}(\Phi_1)\mathrm{Var}(\Phi_2) - \mathrm{Cov}(\Phi_1,\Phi_2)^2}{\langle\Phi_1\rangle^2\langle\Phi_2\rangle^2(1+\varphi_1)^2(1+\varphi_2)^2} \geq \frac{1}{F_{11}F_{22} - F_{12}^2}, \tag{V.32}$$

where we used $\partial_{\phi_j}\langle\Phi_j\rangle = \langle\Phi_j\rangle(1+\phi_j)$ (proven in the appendices of Ref. [5]). The symbol $\langle\bullet\rangle$ stands for the expectation over all possible trajectories $\mathbb{E}_{\omega_{1:N}}[\bullet]$. φ_j is a constant, whose value is also computed in the appendices of Ref. [5], and which .

Now, we rephrase the right hand side of the inequality of Eq. (V.32) to compare it with the classical KUR of Eq. (V.26). Consider the measurement record probability given in Eq. (IV.6), with the encoding of ϕ_j in Hamiltonian and jumps given in Eq. (V.27). Assuming an initially pure state, and that the evolution preserves purity, the probability of the measurement record $\omega_{1:N}$ factorises as

$$\Pr(\omega_{1:N}) = p_0(1+\phi_1)^{N_1}(1+\phi_2)^{N_2}\left\|\prod_{j=1}^{N} L_{k_j}e^{-iH_e(t_j-t_{j-1})}\left|\psi_0\right\rangle\right\|^2, \tag{V.33}$$

where H_e is the effective non-Hermitian Hamiltonian representing the no-jump evolution, $H_e = H - i/2\sum_k L_k^\dagger L_k$. Here, N_1 and N_2 represent the (stochastic) counters of jumps involving, respectively, baths 1 and 2, with $N_1 + N_2 = N$. Let us focus for the moment on the diagonal components of the Fisher

information matrix:

$$F_{jj} = -\langle \partial^2_{\phi_j} \log(1+\phi_j)^{N_j} \rangle - \langle \partial^2_{\phi_j} \log q_j \rangle, \tag{V.34}$$

where q_j is the squared norm factor of Eq. (V.33). We are interested in the Fisher information for $\phi_j = 0$, as it corresponds to the original dynamics of the system. The first term then is just the dynamical activity $\mathcal{A}_j$, corresponding to the bath j, while the second term, which will be referred to as $\mathcal{Q}_j$, is a genuine quantum contribution, depending on the presence of coherences in the system. The term $\mathcal{Q}_j$ vanishes if the system is fully incoherent. Therefore, the inequality of Eq. (V.32) can be rewritten in its final form

$$\frac{\mathrm{Var}(\Phi_1)}{\langle \Phi_1 \rangle^2} \frac{\mathrm{Var}(\Phi_2)}{\langle \Phi_2 \rangle^2} \geq \frac{(1+\varphi_1)^2(1+\varphi_2)^2}{(\mathcal{A}_1+\mathcal{Q}_1)(\mathcal{A}_2+\mathcal{Q}_2) - F_{12}^2} + \frac{\mathrm{Cov}(\Phi_1,\Phi_2)^2}{\langle \Phi_1 \rangle^2 \langle \Phi_2 \rangle^2}. \tag{V.35}$$

This is a multicurrent version of the KUR first presented in Ref. [173]. Analogous results can be found, with a different encoding choice, for TURs, see Ref. [173] for the single current and Ref. [5] for the multicurrent case.

A. Example: three level maser

A good testing platform for the bound is the three level maser, previously introduced in Sect. IV.2 D. For the four jump channels $\{L_{H,\mathrm{out}}, L_{H,\mathrm{in}}, L_{C,\mathrm{out}}, L_{C,\mathrm{in}}\}$, we introduce the counting observables Φ_1, with weights $w^1 = \{1,1,0,0\}$, and Φ_2, with weights $w^2 = \{0,0,1,1\}$. The system is represented in Fig. V.3a. The computation of the right hand side of the bound in Eq. (V.35) requires the formalism of the Fisher information for correlated processes. As the system is renewal, we use the results of Sect. IV.2 for the Fisher information. The same quantity can also be computed on trajectories, by using the monitoring operator formalism of Sect. IV.4 A and the Gillespie algorithm for the evolution in its estimation theory variant, described in Alg. 5. The latter option works also when the signals of two or more channels are indistinguishable from each other.

The plot of Fig. V.3b illustrates the validity of the inequality in Eq. (V.35) as a function of the Ω/γ ratio. While the bound is not saturated, it is significantly tighter than the bound for the same quantity obtained by considering the currents one at a time, following Ref. [173].

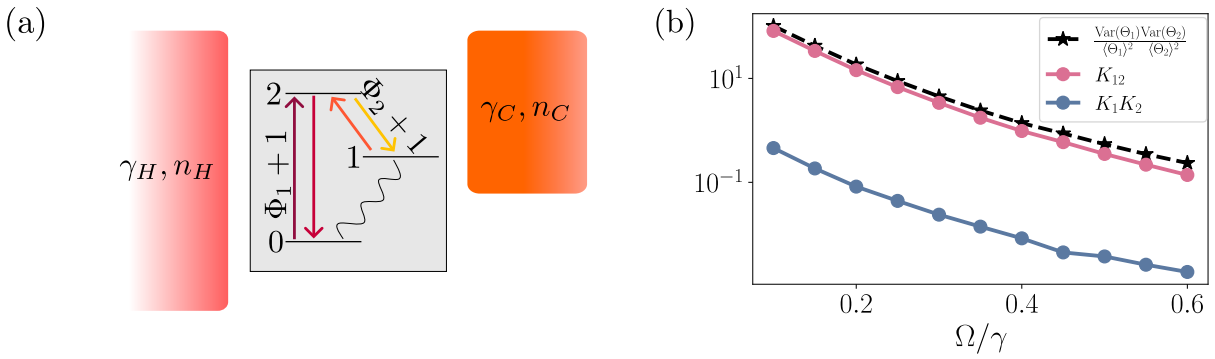

FIG. V.3. (a) schematic of the three level maser system; the update rules for the counting observables in the different jump channels are highlighted. (b) the KUR bounds of Eq. (V.35), keeping into account the interplay of the different currents, compared with the single currents bound of Ref. [173]. Also shown the product of the relative fluctuations on the three level maser, averaged over 5 10^4 trajectories, as a function of Ω/γ (γ assumed to be the same for the two baths). Parameters: $n_1 = 5.0$, $n_2 = 0.01$. Reproduced from [5].

Chapter VI

Collisional learning of a quantum channel

In this Chapter we would like to ask: how much can an experimenter learn about the dynamics of a quantum map, if they can only access a part of it? In the following, we refer to the accessible part as the *system*, and the inaccessible one as the *environment*. We consider a discrete time evolution, with the same CPTP map Λ (on both system and environment) acting at every step. The experimenter feeds in a new system state at every timestep, and performs a local measurement on the system after the action of Λ (in this sense, we talk of a *collisional* protocol). The map Λ couples the information in the system and the environment. The environment acts as a quantum memory. While it is not directly accessible, its effects can indeed be observed in the statistics of measurement outcomes on the system.

The target of the experimenter is to determine, as precisely as possible, the map Λ. In the following, we consider both the case in which the form of Λ is known, and only a few parameters remain to be estimated, and the one in which no prior knowledge about Λ is assumed. We soon observe that not every possible feature of Λ is accessible in this way. Indeed, the problem allows for some gauge degrees of freedom in the characterisation of the map.

The protocol developed in this chapter can be considered as an alternative to the established technique of *process tomography* [175–178]. While our protocol cannot learn all features of a map, it is easier to apply with respect to full process tomography or shadow channel tomography [179]. Through the assumption of an inaccessible environment that perpetuates memory in between measurements, our protocol allows for non-Markovian dynamics. This is important, e.g., for the modeling of errors on real quantum devices [180].

References and collaboration information

The main reference for this section will be [7], co-authored with Simon C. Milz and Felix C. Binder. At the thesis submission stage, the work is, though, still in preparation

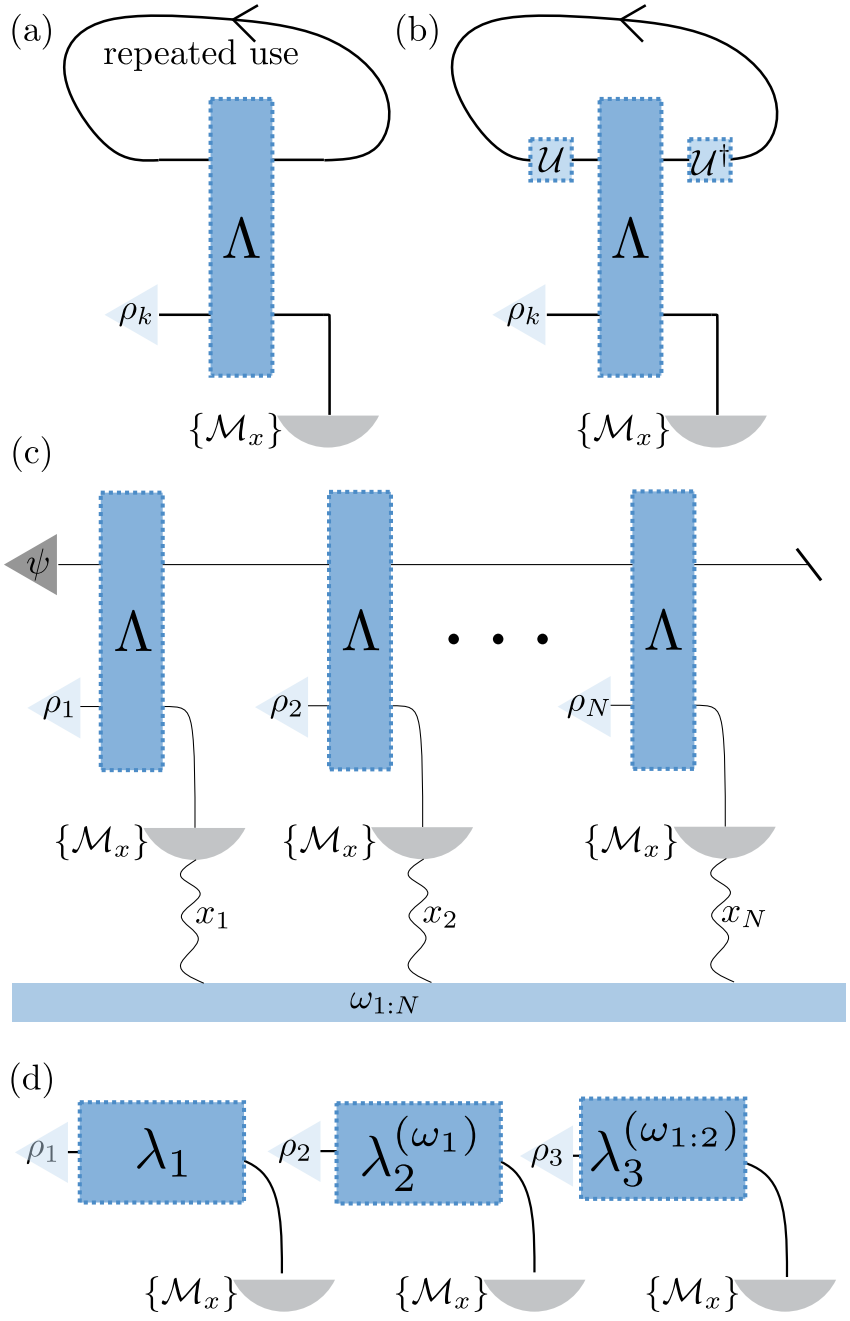

FIG. VI.1. (a) The quantum channel Λ, acting on both environment (above) and system (below), is repeatedly applied. The environment degree of freedom propagates from one step to the next, acting as quantum memory. The experimenter, at each step, feeds a fresh state ρ_k on the system, and after the action of Λ measures the POVM $\{\mathcal{M}_x\}$ (b) Due to the structure of the protocol, some features of the map cannot be learnt. E.g., note that the presence of a unitary map $\mathcal{U}$ before and after Λ on the environment arm leaves no traces on the measurement record. (c) The scheme unravelled in time. (e) From the perspective of the system, the protocol corresponds to the repeated application of new channels λ_k, that are functions of the measurement record up to that point.

VI.1. COLLISIONAL TRAJECTORIES

We consider the protocol represented in Fig. VI.1a. A bipartite CPTP map Λ acts on both system and environment. At each timestep, the experimenter feeds a new state ρ_i into the system branch, and after the action of Λ performs a local measurement on the system, according to the fixed POVM $\{M_x\}$. The outcome x_i of the measurement is recorded, together with the state ρ_i. The measurement record has, therefore, the form

$$\omega_{1:N} = \{(\rho_1, x_1), \dots, (\rho_N, x_N)\}\,. \tag{VI.1}$$

We assume that the POVM $\{M_x\}$ is informationally complete [181, 182]. This means that the POVM elements span the set of self-adjoint operators on the Hilbert space. It is convenient to assume that the states $\{\rho_i\}$ are sampled from a finite and informationally complete set. The measurement outcome of the POVM $\{M_x\}$ is probabilistic: we can associate a probability to a measurement record $\omega_{1:N}$ as

$$\begin{aligned}\Pr(\omega_{1:N}) = \mathrm{Tr}_E \mathrm{Tr}_{S_N}\Big[(\mathcal{M}_{x_N} \otimes \mathbb{I})\, \Lambda \rho_N \otimes \mathrm{Tr}_{S_{N-1}}\Big[\big(\mathcal{M}_{x_{N-1}} \otimes \mathbb{I}\big)\, \Lambda \rho_{N-1} \otimes \dots \\ \dots \otimes \mathrm{Tr}_{S_2}\Big[(\mathcal{M}_{x_2} \otimes \mathbb{I})\, \Lambda \rho_2 \otimes \mathrm{Tr}_{S_1}\Big[(\mathcal{M}_{x_1} \otimes \mathbb{I})\, \Lambda \rho_1 \otimes \psi \Big]\Big] \dots \Big]\Big].\end{aligned} \tag{VI.2}$$

Here and in the following, the space of the environment is labelled as E, while the space of the system at the i-th iteration is S_i. The memory contained in the environment does not allow for a simple factorisation of the probability density of Eq. (VI.2). In this sense, it is similar to non-renewal jump processes, for which the probability density of Eq. (IV.6) does not factorise due to the flow of memory across the jumps.

In our model, the global system-environment dynamics is Markovian, as the global state depends only on the state at the previous step. However, when the environment is traced out, the system effectively experiences a non-Markovian evolution, represented in Fig. VI.1d. In this sense, our model is a special case of a (quantum) Hidden Markov Model (qHMM) [183–186].

It is both of theoretical interest and of numerical use to see the measurement record of Eq. (VI.1) from the perspective of its reduced action on the environment. Let the normalised conditional state of the environment at step $j-1$ be σ_{j-1}. The unnormalised state at step j, upon feeding in of state ρ_j and

measurement of outcome x_j, is

$$\tilde{\sigma}_j = \mathrm{Tr}_{S_j}\left[\left(\mathcal{M}_{x_j}\otimes\mathbb{I}\right)\Lambda\left(\rho_j\otimes\sigma_{j-1}\right)\right]. \tag{VI.3}$$

The evolution $\mathcal{R}^{(\rho_j,x_j)}:\sigma_{j-1}\mapsto\tilde{\sigma}_j$ is a completely positive, trace-non-increasing map, for every choice of ρ_j and for every measurement outcome x_j. Its Kraus operators, which are sub-normalised (with respect to the normalisation of Eq. (II.2)), are given by

$$R_{ilq}^{(\rho_j,x_j)} = \left(\langle i|\, M_{x_j}\otimes\mathbb{I}\right) K_l\left(\sqrt{p_{q_j}}\,|\varphi_{q_j}\rangle\otimes\mathbb{I}\right), \tag{VI.4}$$

where $\{|i\rangle\}$ is an orthonormal basis of the space of the system, M_{x_j} is the operator associated to the POVM element $\mathcal{M}_{x_j}$ (so that $\mathcal{M}_{x_j}\rho = M_{x_j}\rho M_{x_j}^\dagger$), K_l is a Kraus operator of Λ, and $\rho_j=\sum_{q_j}p_{q_j}\,|\varphi_{q_j}\rangle\langle\varphi_{q_j}|$ is the eigendecomposition of the input state ρ_j.

For a finite set of input states, and a fixed POVM, all the maps $\mathcal{R}^{(\rho_j,x_j)}$ can be precomputed, as they do not depend on the conditional state along the trajectory. From a numerical perspective, this allows for a significant speed-up in the calculation of the probability (or the log-likelihood, see below), when compared to the much slower direct evaluation of Eq. (VI.2).

VI.2. GAUGE DEGREES OF FREEDOM AND PROCESS DISTANCE

Not all the possible features of a map Λ can be learnt in our collisional paradigm, if we only have access to the measurement record. Assume that two maps Λ and Λ' yield the same measurement record statistics:

$$\Pr^{(\Lambda)}(\omega_{1:N}) = \Pr^{(\Lambda')}(\omega_{1:N}), \qquad \text{for all } \omega_{1:N}, \tag{VI.5}$$

where $\Pr^{(\Lambda)}$ indicates the probability of a measurement record with respect to the map Λ. Then, no statistical technique can possibly be used to set the maps apart from each other. This induces an equivalence relation among maps: two maps are equivalent if they yield the same probability density for the measurement record. In the language of computational mechanics, two equivalent maps *generate* the same process [187]. Since the difference between two equivalent maps cannot be determined with our protocol, it is only able to learn the quotient set of this equivalence relation. The choice of which map to

pick among the equivalent ones is arbitrary. We refer to this as a *gauge freedom*.

For any possible map Λ, there is always a degree of freedom given by the application of a unitary map $\mathcal{U}$ on the environmental branch, as:

$$\tilde{\Lambda} = (\mathbb{I} \otimes \mathcal{U})\, \Lambda \left(\mathbb{I} \otimes \mathcal{U}^\dagger\right). \tag{VI.6}$$

It is trivial to see that this unitary gauge transformation preserves the statistics of the measurement outcomes. This unitary freedom is not the only possible gauge freedom. However, it is the only one that holds regardless of the structure of the map Λ. Therefore, it is the only one that can be assumed if no prior knowledge of Λ is available.

We argued that the collisional approach cannot learn all features of a quantum map Λ, because of the gauge freedom. Therefore, we need a metric for how close a map Ξ is to the equivalence class of the map Λ, factoring out every possible degree of freedom. In other words, we need to define a distance between Ξ and Λ in terms of the difference between the corresponding outcome statistics $\Pr^{(\Lambda)}(\omega_{1:N})$ and $\Pr^{(\Xi)}(\omega_{1:N})$, for all measurement records $\omega_{1:N}$.

In the following, we introduce a distance measure between processes, that is based on the distance defined for Hidden Markov Models in Ref. [188], and recently generalised to quantum HMM in Ref. [187]. Let $\mathcal{P}_{1:N}$ and $\mathcal{Q}_{1:N}$ be the Choi matrices of, respectively, the process generated by the map Λ and the process generated by Ξ, for N timesteps (they are, in effect, N-teethed quantum combs [106]). Let us define the *quantum coemission divergence rate*

$$R(\Lambda, \Xi) = -\lim_{N\to\infty} \frac{1}{N} \log \left(\frac{\mathrm{Tr}\left[\mathcal{P}_{1:N}\mathcal{Q}_{1:N}\right]}{\sqrt{\mathrm{Tr}\left[\mathcal{P}_{1:N}^2\right]\mathrm{Tr}\left[\mathcal{Q}_{1:N}^2\right]}} \right). \tag{VI.7}$$

In practice, the coemission divergence rate tells the average rate at which the probabilities that the two maps yield for the same measurement record drift apart. If the two maps Λ and Ξ belong to the same equivalence class, i.e. their differences can be gauged out, then $R(\Lambda, \Xi) = 1$.

The explicit calculation of the coemission divergence rate of Eq. (VI.7) appears to be a daunting task at first sight, and in general it is. However, we can exploit the fact that, for each tooth of the comb $\mathcal{P}$ the same map Λ is applied, and same for $\mathcal{Q}$ and Ξ. The inner product $\mathrm{Tr}\left[\mathcal{P}_{1:N}\mathcal{Q}_{1:N}\right]$ is represented in Fig. VI.2a, where $\{L_\alpha\}$ is the collection of Kraus operators for $\mathcal{P}$, and $\{K_\beta\}$ for $\mathcal{Q}$. The network can be

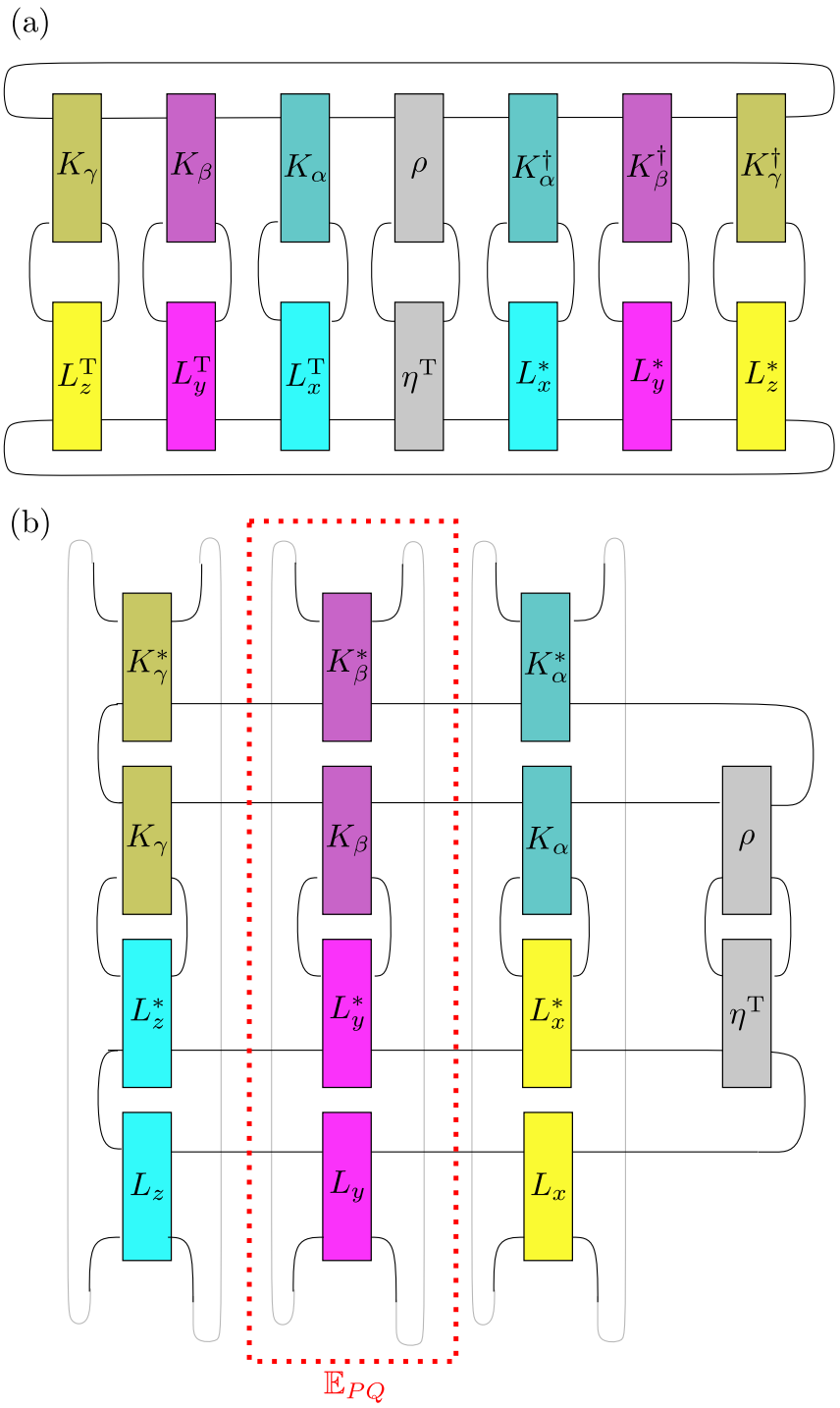

FIG. VI.2. (a) Tensor network calculation of the Hilbert-Schmidt product between the combs $\mathcal{P}_{1:3}$ and $\mathcal{Q}_{1:3}$, numerator of the quantum coemission divergence rate. $\{K_j\}$ are the Kraus operators for Λ, and $\{L_j\}$ the Kraus operators for Ξ. ρ and η are the initial states (on system and environment) of the two processes. (b) Same expression, after rearrangement. Note the definition of the transfer matrix $\mathbb{E}_{\mathcal{PQ}}$. Summation over repeated indices is always assumed.

rearranged as in Fig. VI.2b. Let us define the tensor

$$\mathbb{E}_{\mathcal{PQ}} = \langle \phi^+_{18}| \otimes \langle \phi^+_{45}| \left(\sum_{\alpha\beta} K^\star_\alpha \otimes K_\alpha \otimes L^\star_\beta \otimes L_\beta \right) |\phi^+_{18}\rangle \otimes |\phi^+_{45}\rangle, \tag{VI.8}$$

called *transfer tensor*, where the state $|\phi^+_{ij}\rangle$ is the unnormalised maximally entangled state on Hilbert spaces ij. The numbering of the spaces follows Fig. VI.2b, from top to bottom. The inner product $\mathrm{Tr}\left[\mathcal{P}_{1:N}\mathcal{Q}_{1:N}\right]$ can then be written as

$$\langle \mathcal{P}_{1:N}\mathcal{Q}_{1:N}\rangle_{\mathrm{HS}} = \mathrm{Tr}\left[\mathcal{P}_{1:N}\mathcal{Q}_{1:N}\right] = \langle \phi^+ \otimes \phi^+ | \mathbb{E}^N_{\mathcal{PQ}} | \rho \otimes \eta^{\mathrm{T}} \rangle\rangle, \tag{VI.9}$$

where the notation $|\bullet\rangle\rangle$ represents the vectorisation of the matrix $\bullet$. In the limit $N \to \infty$, only the leading eigenvalue $\mu_{\mathcal{PQ}}$ of $\mathbb{E}_{\mathcal{PQ}}$ plays a role. Using the same logic for the denominator of Eq. (VI.7), we get:

$$R_{\mathrm{HS}}(\Lambda, \Xi) = -\log\left(\frac{\mu_{\mathcal{PQ}}}{\mu_{\mathcal{PP}}\mu_{\mathcal{QQ}}}\right), \tag{VI.10}$$

where $\mu_{\mathcal{PP}}$ and $\mu_{\mathcal{QQ}}$ are, in turn, the leading eigenvalues of the transfer matrices $\mathbb{E}_{\mathcal{PP}}$ and $\mathbb{E}_{\mathcal{QQ}}$. The independence of the quantum coemission divergence rate on the gauge, and its computability, make it perfectly suited for the evaluation of the results of a learning task.

VI.3. A TRADEOFF BETWEEN PRECISION AND EXPRESSIVITY

In the previous section, we established that it is not possible to learn all the possible features of Λ in a collisional way. In this section, we consider the (much simpler) problem of learning a limited set of parameters in an otherwise fully specified Λ. If these parameters cannot be gauged out, they can be collisionally learnt. Let us then define a vector of parameters $\vec{\theta}$, and an estimator $\Phi(\omega_{1:N}) = \hat{\vec{\theta}}$. Let us assume that the estimator is unbiased: $\mathbb{E}_{\omega_{1:N}}\left[\hat{\vec{\theta}}\right] = \vec{\theta}$.

Since the problem is multiparametric in nature, the estimation precision is bounded by the matrix version of the Cramér-Rao bound:

$$\Gamma[\Phi(\omega_{1:N})] - \mathbb{F}^{-1}_{\boldsymbol{\theta}}(\omega_{1:N}) \geq 0, \tag{VI.11}$$

where Γ is the covariance matrix, yielding the interplay of the various parameters to be estimated, and $\mathbb{F}$

is the Fisher information matrix (see, e.g., Eq. (V.30) above).

The computation of the Fisher information matrix can be perfomed with the method of the monitoring operator introduced in Sect. IV.4 A [4], adapted for the collisional framework. If at step j the state ρ_j is fed into the system, and outcome x_j is observed, the unnormalised state of the environment becomes

$$\tilde{\sigma}_j = \sum_{ilq} R_{ilq}^{(\rho_j,x_j)} \sigma_{j-1} \left(R_{ilq}^{(\rho_j,x_j)}\right)^\dagger, \tag{VI.12}$$

using the Kraus operators

$$R_{ilq}^{(\rho_j,x_j)} = \left(\langle i| M_{x_j} \otimes \mathbb{I}\right) K_l \left(\sqrt{p_{q_j}} |\varphi_{q_j}\rangle \otimes \mathbb{I}\right). \tag{VI.4 revisited}$$

Let $p_{x_j} = \mathrm{Tr}\left[\tilde{\sigma}_j\right]$. Note that all the parametric dependence is contained in the Kraus operators of the map Λ, $\{K_l\}$. We define a different monitoring operator $\xi_{\theta,j}$ for each considered parameter. At step j, we update the monitoring operator for each parameter as

$$\begin{aligned}\xi_{\theta,j} = \frac{1}{p_{x_j}} \Bigg[&\sum_{ilq} \left(\partial_\theta R_{ilq}^{(\rho_j,x_j)}\right) \sigma_{j-1} \left(R_{ilq}^{(\rho_j,x_j)}\right)^\dagger + \\ &+ R_{ilq}^{(\rho_j,x_j)} \xi_{\theta,j-1} \left(R_{ilq}^{(\rho_j,x_j)}\right)^\dagger + \\ &+ R_{ilq}^{(\rho_j,x_j)} \sigma_{j-1} \left(\partial_\theta R_{ilq}^{(\rho_j,x_j)}\right)^\dagger \Bigg].\end{aligned} \tag{VI.13}$$

The component ab of the Fisher information matrix $\mathbb{F}$ is then given by

$$\mathbb{F}_{ab} = \mathbb{E}_{\omega_{1:N}} \left[\mathrm{Tr}\left[\xi_a\right] \mathrm{Tr}\left[\xi_b\right]\right], \tag{VI.14}$$

where the expectation value is taken over a large set of trajectories.

A model with more free parameters can represent a larger set of maps. The width of the class of systems that a model can represent is called *expressivity* of the model. More expressivity usually comes at the cost of lower precision: since the parameters interact with each other, the precision in their estimation decreases with their number. Formally, assume that an experimenter actually only cares for a single parameter γ, but is unsure about some other parameters of the physical model, and decides to leave

them free. Then, the precision in the estimation of γ is bounded by the *effective* Fisher information

$$F_{\text{eff},\gamma}(\omega_{1:N}) = \frac{1}{(\mathbb{F}(\omega_{1:N})^{-1})_{\gamma\gamma}}. \tag{VI.15}$$

Note that the inversion of the Fisher information matrix accounts for the mixing of the influences of all the parameters, and that $F_{\text{eff},\gamma} \leq F_\gamma$ for any possible choice of the model. These additional parasitic parameters are known as *nuisance parameters* [189].

A. Example: a correlated noise model

Let us consider the correlated noise model recently introduced in Ref. [45]. A parameter θ is encoded in a qubit via the unitary $R_{\vec{z}}(\theta)$. The rotation operator around an axis $\vec{n}$ is defined as

$$R_{\vec{n}}(\varphi) = \exp\left(-\frac{i}{2}\varphi\vec{n}\cdot\vec{\sigma}\right), \tag{VI.16}$$

with $\vec{\sigma}$ being the vector of Pauli matrices. On the qubit, some dephasing noise is also applied. Such noise is represented by the two operators

$$K_\pm = \frac{R_{\vec{n}}(\pm\varepsilon)}{\sqrt{2}}, \tag{VI.17}$$

where $\vec{n}$ is a fixed axis, and ε is the fixed angle of dephasing. To introduce correlations, we assume that the noise is such that consecutive perturbations form a Markov chain. Let us introduce an ancillary qubit, playing the role of an environment, with basis $\{|-1\rangle, |1\rangle\}$. The Markov chain is enacted by the action of four Kraus operators on the environment, of the form

$$T_{xy} = \sqrt{\frac{1+xyC}{2}}\,|x\rangle\langle y|, \tag{VI.18}$$

for $x, y = \pm 1$, where $C \in [-1, 1]$ is a parameter representing strength and sign of the correlations. The operators of Eq. (VI.17) are conditionally applied on the environment state, yielding a step unitary

$$V_\theta = \sqrt{2}R_{\vec{z}}K_+ \otimes |1\rangle\langle 1| + \sqrt{2}R_{\vec{z}}(\theta)K_- \otimes |-1\rangle\langle -1|. \tag{VI.19}$$

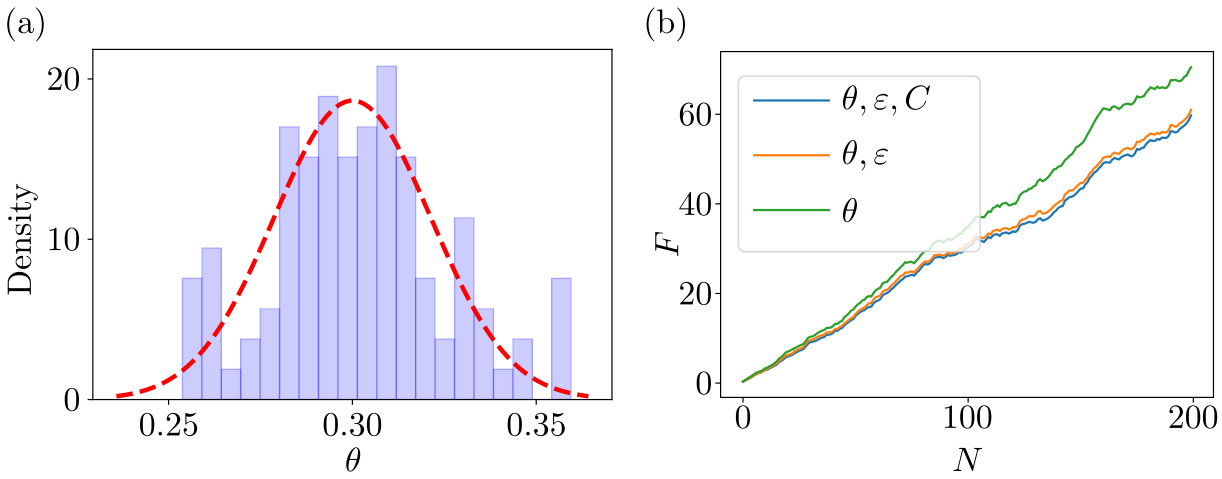

FIG. VI.3. (a) distribution of the outcomes of the parameter θ in the model introduced in Sect. VI.3 A, when the other parameters are assumed to be fixed. The result is compared with the Cramér-Rao bound, represented by the dashed red line, computed numerically with the monitoring operator method. Parameters: $\theta_{\text{true}} = 0.3$, 5 input states, string length 10 000, 100 estimation attempts. Cramér-Rao bound extrapolated via average and fit on 100 trajectories of 100 time steps each. (b) Fisher information for the parameter θ as a function of the number of steps for the same model. The free parameters in each case are represented in the legend. Note that unknown parameters act as nuisance parameters, and decrease the overall estimation precision. Parameters: $\theta_{\text{true}} = 0.3, \varepsilon_{\text{true}} = 0.1, C_{\text{true}} = 0.5$, 5 input states, string length 200, 400 trajectories.

The overall CPTP map Λ has, therefore, the four Kraus operators

$$L_{xy} = (\mathbb{I} \otimes T_{xy})\, V_\theta. \tag{VI.20}$$

The free parameters in the model are, beyond θ, the dephasing angle ε and the correlation strength C. In Fig. VI.3a, we use maximum likelihood estimation, as introduced in Eq. I.21, to estimate θ when all the other parameters are fixed. As expected, the Cramér-Rao bound is saturated. In Fig. VI.3b, we show that the precision is significantly reduced if the parameters ε and C are not fixed, but left to estimate.

VI.4. LEARNING A HAMILTONIAN

In this Section, we assume that the map Λ is guaranteed to be unitary. We refer to it as the super-operator $\mathcal{V}$, such that $\mathcal{V}\bullet = V \bullet V^\dagger$, for a unitary operator V. The map $\mathcal{V}$ is fully characterised by the Hamiltonian H, such that $V = \exp(-iH)$. Learning $\mathcal{V}$ is, therefore, equivalent to learning H. In the literature, different approaches have been presented for Hamiltonian learning, mostly in the framework of measure-and-reprepare metrology [190–194].

As discussed in detail above, collisional protocols are able to learn maps up to a number of gauge freedoms. If the evolution is guaranteed to be unitary, there is only the unitary freedom on the environment of Eq. VI.6, $\mathcal{V} \mapsto (\mathbb{I} \otimes \mathcal{U})\mathcal{V}(\mathbb{I} \otimes \mathcal{U}^\dagger)$.

We focus on a 2-qubit unitary $\mathcal{V}$, which is fully determined by the 16 real parameters $\vec{\theta}$ of its Hamiltonian ($d_S^2 d_E^2$, due to the Hermiticity constraint). The collisional protocol generates a measurement record $\omega_{1:N}$, as in Eq. (VI.1). We then apply maximum likelihood estimation to estimate all 16 parameters, optimising via gradient descent. We obtain an estimate for the parameters $\hat{\vec{\theta}}$, corresponding to a learnt map $\hat{\mathcal{V}}$. The learning trajectory is represented in Fig. VI.4. The loglikelihood of $\hat{\mathcal{V}}$ (solid blue line) converges to the loglikelihood of $\mathcal{V}$ (dotted blue line). As expected, $\hat{\mathcal{V}}$ represents well the dynamics at the level of the system: the process similarity converges to 1 (red line). However, $\hat{\mathcal{V}}$ differs from $\mathcal{V}$ on the environment degree of freedom. To show this fact, we plot the normalised Hilbert-Schmidt product of the Choi states of the two maps (green line). This measure does not converge to 1. Instead, the distance between the Choi state of $\hat{\mathcal{V}}$ and the Choi state of $\mathcal{V}$ is comparable to the distance between two random Choi states. This is a consequence of the gauge freedom. The environment behaves in different ways in the two models while, though, yielding the same dynamics at the level of the system.

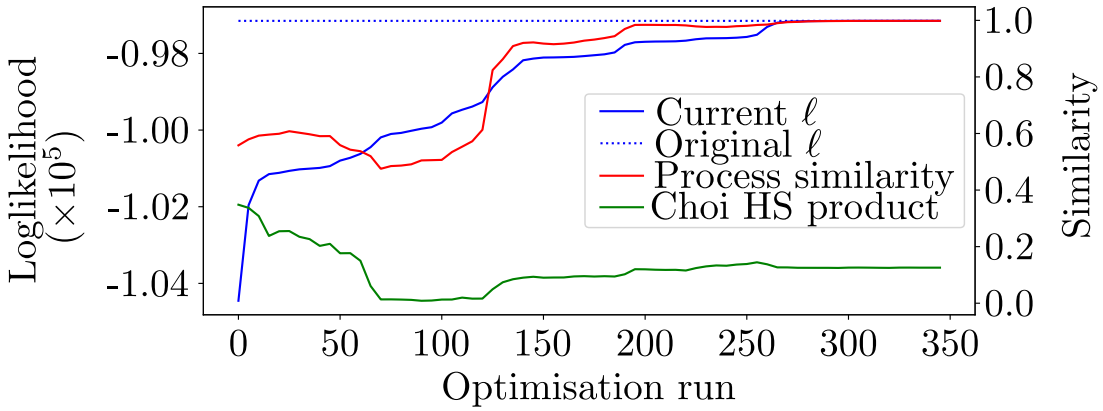

FIG. VI.4. Results of the Hamiltonian learning protocol described in Sect. VI.4, for a single trajectory $\omega_{1:N}$ of 75 000 symbols, where the original Hamiltonian is chosen at random. The log-likelihood curve of the trajectory for the learnt model, as a function of the learning epochs, is shown as the solid blue line. It converges to the log-likelihood of the string for the original model (blue dotted line). The reference axis for the log-likelihoods is on the left. The red line yields the R_{HS} distance, defined in Eq. (VI.10), between the original and the learnt maps (reference axis on the right). The green line represents the (normalised) Hilbert-Schmidt product between the Choi states of the original and the learnt map (reference axis on the right). Parameters: 8 randomly picked input states on the system, 4-outputs symmetric and informationally complete POVM.

The gauge freedom leaves traces at the level of the Fisher information matrix $\mathbb{F}$ for the 16 model parameters. Due to the gauge freedom the model is in fact *overparametrised*. This means that the

number of parameters in the model is larger than what can be learnt from the data. The matrix $\mathbb{F}$ is, therefore, not full-rank, and the Cramér-Rao bound of Eq. (VI.11) diverges. The rank deficiency of $\mathbb{F}$ is, then, a witness of how much the model is overparametrised. In Fig. VI.5, we plot the eigenvalues of

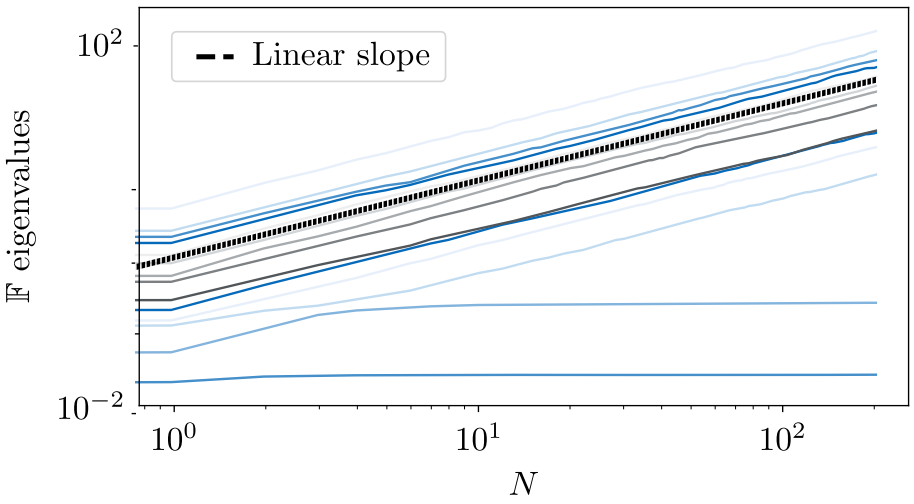

FIG. VI.5. Eigenvalues of the Fisher information matrix, as a function of the length of the measurement record $\omega_{1;N}$, for the same"true" unitary as in Fig. VI.4. Two eigenvalues, that are numerically zero at all times, are not represented. The values are computed with the monitoring operator formalism, and averaged over 1000 trajectories

the Fisher information as a function of the number of steps N of the collisional protocol. The matrix is computed by averaging on trajectories, with a variant of the monitoring operator formalism, see Sect. IV.4 A. We expect all the non-zero eigenvalues to grow linearly with the number of steps, according to the shot-noise scaling. Two eigenvalues are vanishing all along the evolution, and two other initially grow, but then saturate. The latter are, therefore, negligible when $N \gg 1$. The rank deficiency of $\mathbb{F}$ is then 4. Parametrising the gauge freedom at the level of the Hamiltonian is tricky, but we can give the following counting argument to interpret the result. The qubit unitary freedom on the environment $\mathcal{U}$ is characterised by the three Euler angles [95]. Another gauge parameter comes from the general rescaling of the Hamiltonian, according to the mapping $H \mapsto H + \lambda\mathbb{I}$, for any $\lambda \in \mathbb{R}$, that gives a non-physical global phase in the evolution.

The method of the Fisher information matrix we just introduced is a useful tool to diagnose gauge freedoms for any model. This is especially the case when the gauge freedom is hard to isolate and parametrise explicitly. It has, though, two main shortcomings, that we discuss in the following. First, the

Fisher information matrix is defined in terms of the second derivative of the probability distribution. In principle, it is not possible to exclude a dependence on a parameter for derivatives of higher order [195], even though there is none at second order. According to *singular learning theory* [196, 197], this situation is, indeed, the typical one for very large machine learning models.

On the other hand, the Fisher information matrix is defined locally in the space of parameters. In general, there is no guarantee that its rank should remain the same in the entire space. In some pathological cases, it is then possible that the Fisher information matrix has lower rank for the specific set of the "true" parameters. We argue that this behaviour, while theoretically possible, is quite uncommon in real-world data.

Chapter VII

Estimation with classical postprocessing

So far, in this thesis, we have assumed that estimation had access to the entire measurement record. The optimal estimator, whose precision is bound by the Fisher information, is, in general, a complicated function of measurement record, taking its intricate correlations into account. In many situations, the assumption of full access to the measurement record is unrealistic. For example, if an experiment collects a large amount of data per unit time, storing the raw information may prove too expensive. In this situation, one stores, instead, summary statistics. In this section, we prove precision bounds for parameter estimation based on the most common summary statistics. The starting point is the Fisher information data-processing inequality, that gives the reduction in Fisher information due to a post-processing of the raw data.

References and collaboration information

The main reference for this section is [2], co-authored with J.A. Smiga, F.C. Binder and G.T. Landi. As second author of the paper, I worked mainly on the spin thermometry example of Sect. VII.2 A, which is correspondingly given ample space in this thesis. A discussion of information loss and compression in estimation on quantum jump trajectories is given in Sect. IV of Ref. [4].

VII.1. DATA-PROCESSING INEQUALITY

Let us consider a random variable X, distributed according to a parameter-dependent probability distribution $\Pr_\theta(X)$. Let $\Phi : X \to Y$ be a deterministic function of the random variable. It can be proven [86, 198, 199] that the Fisher information about θ contained in Y is less or equal than the Fisher information in X, provided that the mapping Φ does not depend on θ. This is the data-processing inequality for the Fisher information:

$$F_\theta(\Phi(X)) \leq F_\theta(X). \tag{VII.1}$$

It can be proven [200–202] that equality yields if and only if the mapping Φ is a sufficient statistic for θ. This means that $\Phi(X)$ contains all the information in X which is relevant to the estimation taks for θ[6].

In the following, we consider specific mappings Φ, and discuss their impact on estimation performances. We first focus on very general mappings, that can be performed on any process with a discrete set of outcomes. Then, we consider specific post-processings of measurement outcomes of quantum jump processes.

VII.2. EMPIRICAL DISTRIBUTION AND SAMPLE AVERAGE

As introduced in Chapter I, a stochastic process is fully characterised by its joint probability distribution $\Pr(X_{1:N})$. We define its *empirical distribution* as

$$Q_x = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i,x}, \tag{VII.2}$$

where $\delta_{x,y}$ is Kronecker's δ. The random variable Q_x counts the number of occurrences of the symbol x in the random string $X_{1:N}$. Given a realisation of the process $x_{1:N}$, the realisation of the empirical distribution q_x represents the histogram of the outcomes of the process. It is also possible to define *higher-order* empirical distributions, counting transitions between symbols [205].The empirical distribution contains less information than the full realisation of the process, as every information about the order in which the symbols appeared is lost. We will see how this impacts estimation performances in the following.

A further compression of the data is given by the *sample mean*:

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \sum_x Q_x, \tag{VII.3}$$

where the outcomes are assumed to be numeric (so that summing over them is a meaningful operation).

Note that the empirical distribution can be obtained by post-processing the outcomes of a process, and the sample mean by further post-processing the empirical distribution. Hence:

$$F_\theta^{(\mathrm{avg})} \leq F_\theta^{(\mathrm{emp})} \leq F_\theta^{(\mathrm{all})}, \tag{VII.4}$$

[6] An interesting line of research has very recently emerged, investigating how to generalise this proposition to the quantum Fisher information case [203]. The problem is connected to the issue of the reversibility of quantum operations and the Petz recovery map [204].

where the shorthands $F^{(\text{avg})}$, $F^{(\text{emp})}$ and $F^{(\text{all})}$ represent, respectively, the Fisher information of the sample mean, the empirical distribution, and the entire process.

Let us start by considering the Fisher information contained in the empirical distribution. The empirical distribution is constituted by a random vector $\vec{Q}$, where each component Q_x corresponds to a possible outcome x in the alphabet of the process. The expectation value of the empirical distribution yields the steady state of the process $\vec{p}$, as

$$\mathbb{E}[Q_x] = \sum_{x_{1:N}} \Pr(x_{1:N}) \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i,x} = \Pr(x) = p_x. \tag{VII.5}$$

We assume that the random vector $\vec{Q}$ is distributed as a multivariate normal around the steady state[7]. The covariance matrix of such multivariate normal is proven in the appendices of Ref. [2] to be

$$\Sigma_{xy} = \text{Cov}(Q_x, Q_y) = \frac{1}{N} \left(p_x(\delta_{xy} - p_y) + \Psi_{xy} p_y + \Psi_{yx} p_x \right) = \frac{\tilde{\Sigma}_{xy}}{N}, \tag{VII.6}$$

where

$$\Psi_{xy} = \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \left[\Pr(X_{i+k} = x | X_i = y) - p_x\right]. \tag{VII.7}$$

In practice, this means that the covariance matrix of the Q distribution depends on the correlations between the measurements, encoded in the matrix Ψ. We use the Fisher information of the multivariate Gaussian distribution [15] to obtain

$$F_\theta^{(\text{emp})}(X_{1:N}) = N \left(\partial_\theta \vec{p}\right)^T \tilde{\Sigma}^{-1} \left(\partial_\theta \vec{p}\right), \tag{VII.8}$$

where p is the steady-state vector, and we neglected higher order terms in the limit $N \to \infty$. Via the manipulations in the appendices of Ref. [2], we explicitate the role of the correlations matrix Ψ, as

$$F_\theta^{(\text{emp})}(X_{1:N}) = N (\partial_\theta p)^T \left(\mathbb{P} + \Psi\mathbb{P} + \mathbb{P}\Psi^T\right)^{-1} (\partial_\theta p), \tag{VII.9}$$

where $\mathbb{P}$ is the diagonal matrix whose entries are the elements of $\vec{p}$. In the absence of correlations, $\Psi = 0$

[7] We were not able to prove general conditions in which this requirement holds. However, we were also unable to find a counterexample for a stationary process. See the notes in [2] for an example in which this is violated for a non-stationary process.

and one recovers the ordinary expression of the i.i.d. Fisher information. This expression is fully general with respect to the structure of the process, which is fully hidden in the matrix Ψ. For instance, for a Markov process ($\mathcal{M} = 1$), one gets $\Psi = T(\mathbb{I} - T)^+$, where T is the transition matrix of the process, and the superscript $+$ represents the Drazin pseudo-inverse [206, 207]. The Fisher information of the empirical distribution for a Markov process gives

$$F_\theta^{(\mathrm{emp})}(X_{1:N}) = N(\partial_\theta p)^T \left(\mathbb{P} + T(\mathbb{I} - T)^+\mathbb{P} + \mathbb{P}(\mathbb{I} - T^T)^+T^T\right)^{-1}(\partial_\theta p). \tag{VII.10}$$

Let us now focus on the sample mean. For a large number of realisations of the process, the expectation value of the sample mean converges to the average value of the process $\mu = \mathbb{E}[\bar{X}] = J^T p$. Here, the vector J contains the numeric values associated to the elements of the alphabet. The variance of the sample mean is $\sigma^2_{\bar{X}} = J^T \Sigma J$, from the rule for the propagation of uncertainties for a linear transformation, where Σ is the covariance matrix for the empirical distribution defined above. We assume that the sample mean is normally distributed around μ. Using the Fisher information of the normal distribution, we obtain the expression for the Fisher information of the sample mean:

$$F_\theta^{(\mathrm{avg})}(X_{1:N}) = N\frac{(\partial_\theta \mu)^2}{J^T(\mathbb{P} + \Psi\mathbb{P} + \mathbb{P}\Psi^T)J}. \tag{VII.11}$$

A. Example: spin-one chain thermometry

We consider a similar example to the one of Sect. III.4, but for spin-one. Also in this case, we set to estimate the temperature T. The Fisher information data-processing inequality fixes a hierarchy of

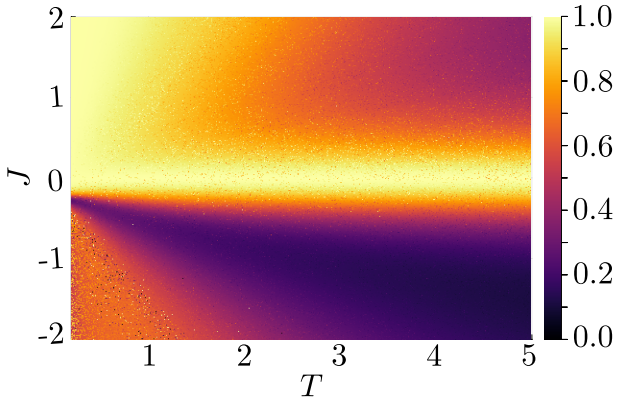

FIG. VII.1. Heatmap of $F_T^{(\mathrm{emp})}/F_T^{(\mathrm{all})}$ for different values of T and J, and fixed $B = 0.5$. Reproduced from [2].

informational content at different levels of post-processing, given by Eq. (VII.4). However, the ratios can vary widely, according to the specifics of the system at hand. This is illustrated in Fig. VII.1 in terms of the ratio $F^{(\mathrm{emp})}/F^{(\mathrm{all})}$. The empirical distribution only collects information about the one-site marginal, while the full distribution also considers the correlations and the order in which the elements appear. In a ferromagnetic spin chain, most of the information is contained in the marginals, in an anti-ferromagnetic chain in the correlations. Consequently, the empirical distribution retains almost the entire Fisher information for a ferromagnetic spin chain, but it has next to no Fisher information for an anti-ferromagnetic chain. In the strongly ferromagnetic regime $J \gg 1$ at low temperature, the vast majority of the spins is aligned along the magnetic field. The empirical distribution becomes then almost a sufficient statistic for T, as $F_T^{(\mathrm{emp})} \sim F_T^{(\mathrm{all})}$.

B. Sample mean and quantum jump trajectories

Let us now focus again on quantum jump trajectories. We obtain an analytical expression for the Fisher information of the sample mean of waiting times, for renewal processes with a single jump channel. Let us consider an estimator of a parameter θ, $\hat{\theta}(\bar{\tau})$ that only depends on the sample mean $\bar{\tau}$ of the waiting times:

$$\bar{\tau} = \frac{\tau_1 + \tau_2 + \ldots \tau_N}{N}. \tag{VII.12}$$

Note that the process $\tau_{1:N}$ is i.i.d., if the system is renewal and there is a single jump channel. The Fisher information for the sample mean follows from Eq. (VII.10), as:

$$F_\theta(\bar{\tau}) = N \frac{(\partial_\theta \mu)^2}{\sigma^2}, \tag{VII.13}$$

for $\mu = \mathbb{E}[\tau_j]$, and $\sigma^2 = \mathrm{Var}[\tau_j]$. Since the process is renewal, the single jump superoperator $\mathcal{J}$ has rank 1, $\mathcal{J}\bullet = \mathrm{Tr}[\bullet]\eta$, for some output state η. The waiting time distribution is then $W(\tau) = \mathrm{Tr}\left[\mathcal{J}e^{\mathcal{L}_0\tau}\eta\right]$. The values of μ and σ follow from the Laplace transform of the WTD [145]

$$\tilde{W}(s) = \int_0^\infty W(\tau)e^{-s\tau} = \mathrm{Tr}\left[\mathcal{J}\left(s - \mathcal{L}_0\right)^{-1}\eta\right], \tag{VII.14}$$

as $\mu = -\partial_s \tilde{W}(0)$, and $\sigma^2 = \partial_s^2 \tilde{W}(0)$. We express the Fisher information contained in the sample mean of the waiting times as

$$F_\theta(\bar{\tau}) = N \frac{\left(\partial_\theta \partial_s \tilde{W}(0)\right)^2}{\partial_s^2 \tilde{W}(0)}. \tag{VII.15}$$

C. Example: resonant fluorescence

We consider a similar setup to the one of Sect. IV.2 B, a single qubit with Hamiltonian

$$H = \frac{\omega}{2}\sigma_z + \frac{\Omega}{2}\sigma_x, \tag{VII.16}$$

but this time with a single jump operator $L = \sqrt{\Gamma}\sigma_-$. The system is renewal, as the entire information on the state is lost on jump. This setup is typical of *resonant fluorescence* experiments. The Fisher information for the Rabi frequency Ω has been computed in Ref. [36]:

$$F_\Omega^{(\mathrm{all})}(\omega_{1:N}) = N\left(\frac{8}{\Gamma^2} + \frac{4}{\Omega^2}\right) \tag{VII.17}$$

in resonance $\omega = 0$. The computation of the Fisher information of the sample mean for Ω as per Eq. (VII.15) gives

$$F_\Omega^{(\mathrm{avg})}(\omega_{1:N}) = \frac{N}{\Omega^2}\frac{4}{1 - 2(\Omega/\Gamma)^2 + 4(\Omega/\Gamma)^4}. \tag{VII.18}$$

As expected from Eq. (VII.4), $F_\Omega^{(\mathrm{avg})}$ lower-bounds $F_\Omega^{(\mathrm{all})}$. Their behaviour, as a function of the emission rate Γ, is represented in Fig. VII.2. The two expressions for the Fisher information are, in general, far from tight. They show a particularly diverging behaviour for small Γ, while for large Γ they converge to the same value. Indeed, for $\Gamma \gg 1$, the systems behaves very similarly to a classical system, with almost no evolution of the state in between a jump and the next, as jumps happen very frequently. Consequently, the complex structure of the quantum WTD, which is not captured by the sample mean, does not play a role, and only the classical exponential decay is relevant.

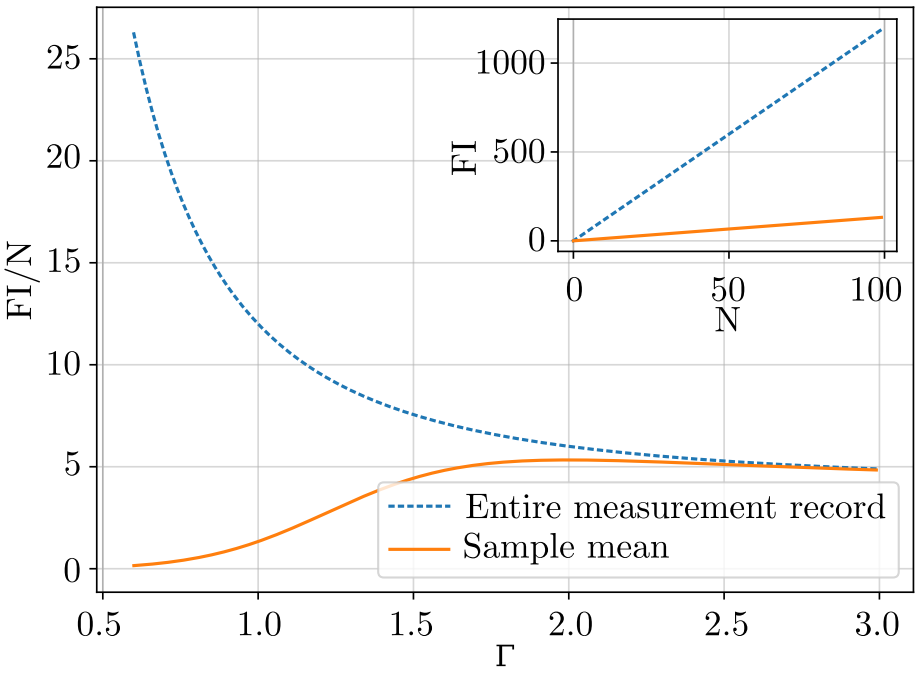

FIG. VII.2. Fisher information rate for Ω of the whole measurement record, compared to the Fisher information contained in the sample mean, as a function of the rate Γ. In the main panel, for fixed $\Omega = 1$, as a function of Γ. In the inset plot, for fixed $\Gamma = \Omega = 1$, as a function of the number of jumps. Reproduced from [4].

VII.3. PARTIAL MONITORING AND CHANNEL MERGING

Let us finally focus on specific functions Φ, that are specific to quantum jump trajectories. Assume the usual form for the measurement record

$$\omega_{1:N} = \{(\tau_1, k_1), (\tau_2, k_2), \ldots, (\tau_N, k_N)\}. \qquad \text{(IV.1 revisited)}$$

We introduce two classes of classical post-processings:

- *partial monitoring* means that some of the channels are not monitored. The function Φ acts by ignoring all the tuples involving some specific channels in $\omega_{1:N}$.

- *channel merging* means that some of the channels cannot be distinguished from each other. As an extreme case, the experimenter may be completely blind to the different channels, and only retrieve the sequence of jump times.

Partial monitoring modifies the subdivision of the Lindbladian $\mathcal{L}$ between jumps $\{\mathcal{J}_k\}$ and no-jump $\mathcal{L}_0$. However, partial monitoring does not, in general, change the renewal or non-renewal nature of the process.

The case of channel merging is different, as it can change the nature of the process. Let us consider, for instance, a renewal process with two jump operators, having different output states σ_1 and σ_2. If the two jumps are merged, the state after the jump will be given by a statistical mixture of σ_1 and σ_2, with weights depending on the state before the jump: the state dependence reappears!

A. Example: partial monitoring and channel merging on the three level maser

As an example of postprocessing on quantum jump trajectories, we consider again the three level maser model, introduced in Sect. IV.2 D. The model is reproduced in Fig. VII.3a. This time, we consider the

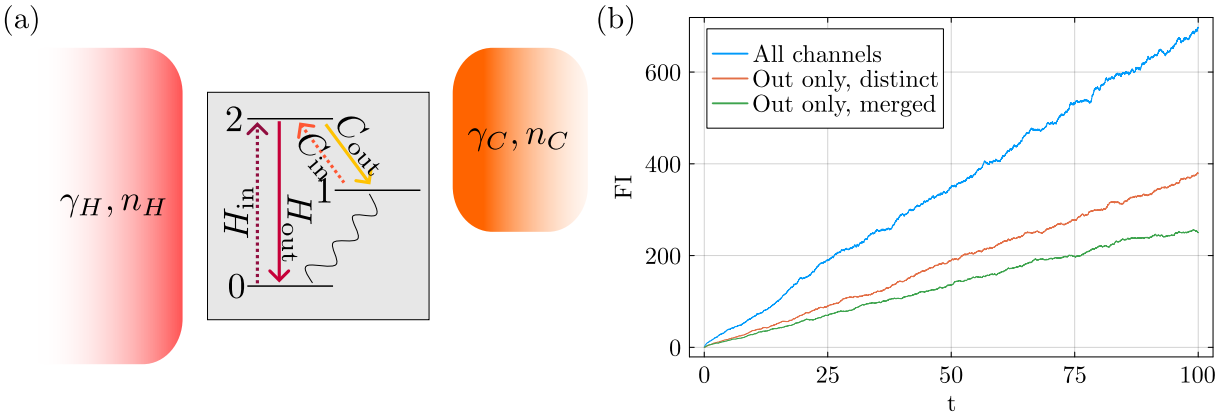

FIG. VII.3. (a) Scheme of the three level maser model. The unmonitored jump channels are dotted. (b) Fisher information for the γ_H coupling, computed on trajectories, for different monitoring schemes: all channels are monitored (blue), only $H_{\rm out}$ and $C_{\rm out}$ are monitored but distinguishable from each other (orange), only $H_{\rm out}$ and $C_{\rm out}$ are monitored and undistinguishable (green). Parameters: $\Omega = 1$, $\Delta = 1$, $\bar{n}_H = 8$, $\bar{n}_C = 2$, $\gamma_H = 0.3$, $\gamma_C = 0.1$, 400 trajectories.

effects of post-processing on the Fisher information for the jump rate γ_H.

When all the jumps are separately monitored, the three level maser is renewal, as discussed above. Physically, detecting injections is usually trickier than detecting emissions. Let us then assume that only the two emission channels $H_{\rm out}$ and $C_{\rm out}$ are monitored. This is a post-processing of the measurement record, and yields a reduction in the Fisher information rate, see Fig. VII.3. The process, though, remains renewal, as both the remaining jumps satisfy the condition of Eq. (IV.8).

Let us now assume that the two emission channels cannot be distinguished from each other. In this case, the process is not renewal anymore. In fact, after a jump the system can be either in $|0\rangle$ or in $|1\rangle$, with a

probability that depends on the state before the jump. As a further post-processing of the measurement record, also channel merging comes at the expense of a reduction in the Fisher information.

Discussion and outlook

In this thesis, we approached the problem of learning from correlated data of quantum origin from a variety of perspectives. Let us now discuss some of our choices, and give some insights into possible future research avenues.

The first main result of this thesis is a comprehensive framework for correlated estimation on processes with finite Markov order. We gave a general expression for the Fisher information rate. We also heuristically connected the sign of the correlations to the estimation advantage or disadvantage they bring.

We then focussed on classical data generated by quantum processes. We both considered a continuous-time framework, the quantum jump unraveling, and a discrete-time one, collisional learning. The learning is always classical. Indeed, while the data are created by a quantum system, we do not assume the freedom to optimise the measurement scheme. In a sense, the quantumness of the system always remains on the other side of a barrier that the experimenter cannot trespass.

This is not the only possible choice. One could consider optimisation over the measurement scheme, or even for the application of feedback to the system itself, depending on previous measurement outcomes. Given the invasive nature of quantum measurements, this perspective requires a fully quantum description of the learning task. For correlated processes, this has recently been attempted using the formalism of quantum combs [42–45]. While this formalism is numerically extremely powerful, the understanding of the results is still very much open. For example, it is unclear if and how our results regarding the estimation role of correlations can be generalised to a full quantum setting. In Ref. [87], the authors proved that the chain rule for the Fisher information of Eq. (I.17) does not generalise to the quantum Fisher information. Furthemore, bounding the memory length for quantum systems is a challenging problem. In Ref. [208], in fact, the authors showed that a quantum version of the Markov order can only take zero, one or infinite values. From a more practical perspective, the memory of (most) quantum systems vanishes in time, so it would make sense to define a notion of an *effective* (or approximate) quantum Markov order. Whether this may lead to meaningful metrological insights, it is still unclear. Another possibility is to bound the memory about the past in terms of the dimension of the space that is available to store such information [209]: smaller environments should, in some way, correspond to a shorter memory length. In particular, it would be interesting to understand how the amount of memory available about the past constrains the estimation performances.

In the framework of continuously monitored systems, we gave analytical results for estimation and hypothesis testing on renewal processes. For non-renewal processes, we generalised the previously introduced monitoring operator method. In particular, we showed its direct connection to maximum likelihood estimation. We introduced a new algorithm, termed the quantum Gillespie algorithm, to efficiently simulate quantum jump trajectories.

There are many other possible unravellings of the GKSL master equation [98, 99]. In particular, the diffusive unravelling corresponding to homodyne detection is a natural alternative to the jump unravelling. The definition of renewal process of Eq. (IV.8) does not make sense in the diffusive unravelling, where the experimenter is assumed to get an outcome at each point in time. Similarly, a WTD cannot be defined. On the other side, the monitoring operator formalism can be used in the homodyne detection case. This has been done in Refs. [35, 38]. Furthemore, one can optimise over all possible unravellings, see Ref. [40]. For any unravelling, one can write a Fisher information rate similar to Eq. (IV.48), but the task of analytically averaging it seems to be quite hopeless. In this sense, some insight may come from recent research in the statistical structure of the space of quantum trajectories, see e.g. Ref. [210]. In particular, it may be possible to approximate the Fisher information rate in terms of the typical set of quantum trajectories on the specific system.

From a more applied perspective, the ideal model of the quantum jump unravelling falls short of taking into account some unavoidable experimental uncertanties, e.g. on jump times [211–213].

We introduced the framework of collisional learning. In particular, we explored the gauge freedoms in the learning of a general channel, and introduced a new definition of quantum process distance that allows to deal with them. We introduced a method, based on the Fisher information matrix, to diagnose the gauge freedom. We presented a collisional solution to the Hamiltonian learning problem.

In the last chapter, we gave analytical results for the Fisher information rate of summary statistics of correlated stationary processes, particularly for the empirical distribution and the sample mean. We then discussed the effect of partial monitoring and/or channel merging on estimation tasks based on quantum jump trajectories.

More in general, this thesis is about learning problems. Two different formalisms are usually employed to tackle such problems: a more physics-oriented one, based on estimation theory, and one more common in machine learning, centred around the concepts of statistical learning risk [214, 215]. The interplay of these two formalisms is still largely unexplored. An in-depth exploration of their connection could also shed light on the performances of neural networks as estimators, for which mostly aneddotical evidence

is available at the moment [216, 217].

Bibliography

[1] M. Radaelli, G. T. Landi, K. Modi, and F. C. Binder, New J. Phys. **25**, 053037 (2023).

[2] J. A. Smiga, M. Radaelli, F. C. Binder, and G. T. Landi, Phys. Rev. Res. **5**, 033150 (2023).

[3] M. Radaelli, G. T. Landi, and F. C. Binder, Phys. Rev. A **110**, 062212 (2024).

[4] M. Radaelli, J. A. Smiga, G. T. Landi, and F. C. Binder, arXiv:2402.06556 (2024).

[5] S. V. Moreira, M. Radaelli, A. Candeloro, F. C. Binder, and M. T. Mitchison, arXiv:2411.09088 (2024).

[6] M. Radaelli, J. A. Smiga, G. T. Landi, and F. C. Binder, In preparation (2025).

[7] M. Radaelli, S. C. Milz, and F. C. Binder, In preparation (2025).

[8] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[9] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photonics **5**, 222 (2011).

[10] P. Christoffersen, *Elements of Financial Risk Management* (Academic Press, 2011).

[11] J. E. Ingersoll, *Theory of Financial Decision Making*, Vol. 3 (Rowman & Littlefield, 1987).

[12] M. E. Blume, J. Finance **26**, 1 (1972).

[13] R. Zhang, Y. Li, A. L. Zhang, Y. Wang, and M. J. Molina, Proc. Natl. Acad. Sci. **117**, 14857 (2020).

[14] C. W. Helstrom, J. Stat. Phys. **1**, 231 (1969).

[15] S. M. Kay, *Statistical Signal Processing: Estimation Theory*, Vol. 1 (Prentice Hall, 1993).

[16] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory* (John Wiley & Sons, 2004).

[17] M. Thomas and A. T. Joy, *Elements of Information Theory* (Wiley-Interscience, 2006).

[18] R. A. Fisher, Philos. Trans. R. Soc. **222**, 309 (1922).

[19] H. Cramér, *Mathematical Methods of Statistics*, Vol. 26 (Princeton University Press, 1999).

[20] C. W. Helstrom, Information and Control **10**, 254 (1967).

[21] C. W. Helstrom, Int. J. Theor. Phys. **1**, 37 (1968).

[22] C. W. Helstrom, Inform. Control **13**, 156 (1968).

[23] C. Helstrom, Phys. Rev. Lett. **25**, 1012 (1967).

[24] H. Yuen and M. Lax, IEEE Trans. Inf. Theory **19**, 740 (1973).

[25] H. P. Yuen, Phys. Rev. Lett. **51**, 719 (1983).

[26] C. M. Caves, K. S. Thorne, R. W. Drever, V. D. Sandberg, and M. Zimmermann, Rev. Mod. Phys. **52**, 341 (1980).

[27] O. E. Barndorff-Nielsen and R. D. Gill, J. Phys. A **33**, 4481 (2000).

[28] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[29] S. L. Braunstein, C. M. Caves, and G. J. Milburn, Ann. Phys. **247**, 135 (1996).

[30] S. L. Braunstein, Phys. Lett. A **219**, 169 (1996).

[31] M. G. Paris, Int. J. Quantum Inf. **7**, 125 (2009).

[32] L. Pezze and A. Smerzi, in *Atom interferometry* (IOS Press, 2014) pp. 691–741.

[33] L. Pezze, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Rev. Mod. Phys. **90**, 035005 (2018).

[34] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola, Rev. Mod. Phys. **90**, 035006 (2018).

[35] S. Gammelmark and K. Mølmer, Phys. Rev. A **87**, 032115 (2013).

[36] A. H. Kiilerich and K. Mølmer, Phys. Rev. A **89**, 052110 (2014).

[37] J. Boeyens, B. Annby-Andersson, P. Bakhshinezhad, G. Haack, M. Perarnau-Llobet, S. Nimmrichter, P. P. Potts, and M. Mehboudi, New J. Phys. **25**, 123009 (2023).

[38] F. Albarelli, M. A. Rossi, D. Tamascelli, and M. G. Genoni, Quantum **2**, 110 (2018).

[39] M. Tsang, H. M. Wiseman, and C. M. Caves, Phys. Rev. Lett. **106**, 090401 (2011).

[40] S. Gammelmark and K. Mølmer, Phys. Rev. Lett. **112**, 170401 (2014).

[41] A. Godley and M. Guta, Quantum **7**, 973 (2023).

[42] A. Altherr and Y. Yang, Phys. Rev. Lett. **127**, 060501 (2021).

[43] Q. Liu and Y. Yang, Quantum **8**, 1571 (2024).

[44] S. Kurdzialek, P. Dulian, J. Majsak, S. Chakraborty, and R. Demkowicz-Dobrzanski, New J. Phys. **27**, 013019 (2025).

[45] S. Kurdzialek, F. Albarelli, and R. Demkowicz-Dobrzanski (2024).

[46] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing Statistical Hypotheses*, Vol. 3 (Springer, 1986).

[47] M. G. Kendall, *The Advanced Theory of Statistics.* (John Wiley and Sons, 1943).

[48] J. Neyman and E. S. Pearson, Phil. Trans. R. Soc. **231**, 289 (1933).

[49] J. Neyman and E. S. Pearson, Math. Proc. Camb. Philos. Soc. **29**, 492 (1933).

[50] D. Middleton, I. of Electrical, and E. Engineers, *An Introduction to Statistical Communication Theory*, Vol. 960 (McGraw-Hill New York, 1960).

[51] S. Lloyd, Science **321**, 1463 (2008).

[52] A. Chefles, Contemp. Phys. **41**, 401 (2000).

[53] J. Bae and L.-C. Kwek, J. Phys. A **48**, 083001 (2015).

[54] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **46**, 2428 (2000).

[55] K. M. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, Commun. Math. Phys. **279**, 251 (2008).

[56] F. G. Brandao and M. B. Plenio, Commun. Math. Phys. **295**, 791 (2010).

[57] M. Berta, F. G. Brandão, G. Gour, L. Lami, M. B. Plenio, B. Regula, and M. Tomamichel, Quantum **7**, 1103 (2023).

[58] L. Lami, arXiv:2408.06410 (2024).

[59] M. Hayashi and H. Yamasaki, arXiv:2408.02722 (2024).

[60] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, Npj Quantum Inf. **5**, 50 (2019).

[61] M. Tsang, Phys. Rev. Lett. **108**, 170502 (2012).

[62] K. Mølmer, Phys. Rev. Lett. **114**, 040401 (2015).

[63] A. H. Kiilerich and K. Mølmer, Phys. Rev. A **98**, 022103 (2018).

[64] A. H. Kiilerich and K. Mølmer, Phys. Rev. A **97**, 052113 (2018).

[65] E. Martínez Vargas, C. Hirche, G. Sentís, M. Skotiniotis, M. Carrizo, R. Muñoz-Tapia, and J. Calsamiglia, Phys. Rev. Lett. **126**, 180502 (2021).

[66] G. Gasbarri, M. Bilkis, E. Roda-Salichs, and J. Calsamiglia, Quantum **8**, 1289 (2024).

[67] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 180501 (2008).

[68] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, Phys. Rev. A **81**, 032339 (2010).

[69] Q. Zhuang and S. Pirandola, Phys. Rev. Lett. **125**, 080505 (2020).

[70] J. Bavaresco, M. Murao, and M. T. Quintino, Phys. Rev. Lett. **127**, 200504 (2021).

[71] Y. Hasegawa and T. Van Vu, Phys. Rev. E **99**, 062126 (2019).

[72] R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield, Phys. Rev. E **89**, 042135 (2014).

[73] J. P. Crutchfield and K. Young, Phys. Rev. Lett. **63**, 105 (1989).

[74] C. R. Shalizi and J. P. Crutchfield, J. Stat. Phys. **104**, 817 (2001).

[75] J. P. Crutchfield and D. P. Feldman, Chaos **13**, 25 (2003).

[76] J. P. Crutchfield, Nat. Phys. **8**, 17 (2012).

[77] J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield, J. Phys. A **42**, 362002 (2009).

[78] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield, J. Stat. Phys. **136**, 1005 (2009).

[79] N. F. Travers and J. P. Crutchfield, J. Stat. Phys. **145**, 1181 (2011).

[80] R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield, Phys. Rev. E **89**, 042135 (2014).

[81] N. F. Travers and J. P. Crutchfield, J. Stat. Phys. **145**, 1202 (2011).

[82] K. Dietz, Stat. Methods Med. Res. **2**, 23 (1993).

[83] N. G. Becker, *Analysis of Infectious Disease Data* (Chapman and Hall/CRC, 2017).

[84] D. Sivia and J. Skilling, *Data Analysis: a Bayesian Tutorial* (Oxford University Press, 2006).

[85] Y. Li, L. Pezzè, M. Gessner, Z. Ren, W. Li, and A. Smerzi, Entropy **20**, 628 (2018).

[86] P. Zegers, Entropy **17**, 4918 (2015).

[87] K. Micadei, D. Rowlands, F. Pollock, L. Céleri, R. Serra, and K. Modi, New J. Phys. **17**, 023057 (2015).

[88] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[89] A. W. Van der Vaart, *Asymptotic Statistics*, Vol. 3 (Cambridge University Press, 2000).

[90] E. Sudarshan, P. Mathews, and J. Rau, Phys. Rev. **121**, 920 (1961).

[91] T. F. Jordan, M. A. Pinsky, and E. Sudarshan, J. Math. Phys. **3**, 848 (1962).

[92] K. Kraus, A. Böhm, J. D. Dollard, and W. Wootters, *States, Effects, and Operations* (Springer, 1983).

[93] K. Kraus, Annals of Physics **64**, 311 (1971).

[94] A. Peres, *Quantum Theory: Concepts and Methods* (Springer, 2002).

[95] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2001).

[96] G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).

[97] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, J. Math. Phys. **17**, 821 (1976).

[98] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, 2002).

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

[100] M. M. Wilde, arXiv:1106.1445 (2011).

[101] A. Monras, A. Beige, and K. Wiesner, arXiv:1002.2337 (2010).

[102] J. de Pillis, Pac. J. Math. **23**, 129 (1967).

[103] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).

[104] M.-D. Choi, Linear Algebra its Appl. **10**, 285 (1975).

[105] S. Milz and K. Modi, PRX Quantum **2**, 030201 (2021).

[106] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).

[107] A. Gilchrist, D. R. Terno, and C. J. Wood, arXiv:0911.2539 (2009).

[108] S. Milz, F. A. Pollock, and K. Modi, Open Syst. Inf. Dyn. **24**, 1740016 (2017).

[109] A. N. Jordan and I. A. Siddiqi, *Quantum Measurement: Theory and Practice* (Cambridge University Press, 2024).

[110] W. Nagourney, J. Sandberg, and H. Dehmelt, Phys. Rev. Lett. **56**, 2797 (1986).

[111] J. C. Bergquist, R. G. Hulet, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. **57**, 1699 (1986).

[112] T. Sauter, R. Blatt, W. Neuhauser, and P. Toschek, Opt. Commun. **60**, 287 (1986).

[113] R. Vijay, D. Slichter, and I. Siddiqi, Phys. Rev. Lett. **106**, 110502 (2011).

[114] K. Siva, G. Koolstra, J. Steinmetz, W. P. Livingston, D. Das, L. Chen, J. M. Kreikebaum, N. Stevenson, C. Jünger, D. I. Santiago, *et al.*, PRX Quantum **4**, 040324 (2023).

[115] A. Kubanek, M. Koch, C. Sames, A. Ourjoumtsev, P. W. Pinkse, K. Murr, and G. Rempe, Nature **462**, 898 (2009).

[116] S. Gleyzes, S. Kuhr, C. Guerlin, J. Bernu, S. Deleglise, U. Busk Hoff, M. Brune, J.-M. Raimond, and S. Haroche, Nature **446**, 297 (2007).

[117] T. Rybarczyk, B. Peaudecerf, M. Penasa, S. Gerlich, B. Julsgaard, K. Mölmer, S. Gleyzes, M. Brune, J. Raimond, S. Haroche, *et al.*, Phys. Rev. A **91**, 062116 (2015).

[118] P. M. Riechers, arXiv:2310.03968 (2023).

[119] P. Billingsley, *Probability and Measure* (John Wiley and Sons, 1995).

[120] S. Kurdziałek and R. Demkowicz-Dobrzański, J. Opt. **23**, 075701 (2021).

[121] E. Ising, *Beitrag zur Theorie des Ferro-und Paramagnetismus*, Ph.D. thesis, Hamburg University (1924).

[122] J. M. Yeomans, *Statistical Mechanics of Phase Transitions* (Clarendon Press, 1992).

[123] D. P. Feldman, *Computational Mechanics of Classical Spin Systems*, Ph.D. thesis, University of California, Davis (1998).

[124] D. P. Feldman and J. P. Crutchfield, Entropy **24**, 1282 (2022).

[125] Q. Liu, Z. Hu, H. Yuan, and Y. Yang, Phys. Rev. Lett. **130**, 070803 (2023).

[126] L. Bouten, M. Guta, and H. Maassen, J. Phys. A **37**, 3189 (2004).

[127] B. Vacchini, Sci. Rep. **10**, 5592 (2020).

[128] E. Geva and R. Kosloff, Phys. Rev. E **49**, 3903 (1994).

[129] H. E. Scovil and E. O. Schulz-DuBois, Phys. Rev. Lett. **2**, 262 (1959).

[130] A. A. S. Kalaee, A. Wacker, and P. P. Potts, Phys. Rev. E **104**, L012103 (2021).

[131] G. T. Landi, arXiv:2305.07957 (2023).

[132] A. J. Daley, Adv. Phys. **63**, 77 (2014).

[133] D. J. C. Mackay, in *Learning in Graphical Models* (Springer, 1998) pp. 175–204.

[134] J. Hammersley, *Monte Carlo Methods* (Springer Science & Business Media, 2013).

[135] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).

[136] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).

[137] G. T. Landi, Melt.

[138] J. R. Johansson, P. D. Nation, and F. Nori, Comput. Phys. Commun. **183**, 1760 (2012).

[139] S. Krämer, D. Plankensteiner, L. Ostermann, and H. Ritsch, Comput. Phys. Commun. **227**, 109 (2018).

[140] C. Cohen-Tannoudji and J. Dalibard, Europhys. Lett. **1**, 441 (1986).

[141] J. Dalibard, Y. Castin, and K. Mølmer, Phys. Rev. Lett. **68**, 580 (1992).

[142] K. Mølmer and Y. Castin, J. Eur. Opt. Soc.: Part B **8**, 49 (1996).

[143] R. Dum, P. Zoller, and H. Ritsch, Phys. Rev. A **45**, 4879 (1992).

[144] C. Gardiner and P. Zoller, *Quantum Noise* (Springer Science & Business Media, 2004).

[145] T. Brandes, Ann. Phys. **520**, 477 (2008).

[146] G. T. Landi, M. J. Kewming, M. T. Mitchison, and P. P. Potts, PRX Quantum **5**, 020201 (2024).

[147] M. O. Scully and M. S. Zubairy, *Quantum Optics* (Cambridge University Press, 1997).

[148] S. Haroche, Rev. Mod. Phys. **85**, 1083 (2013).

[149] D. Meschede, H. Walther, and G. Müller, Phys. Rev. Lett. **54**, 551 (1985).

[150] P. Filipowicz, J. Javanainen, and P. Meystre, J. Opt. Soc. Am. B **3**, 906 (1986).

[151] L. Lugiato, M. Scully, and H. Walther, Phys. Rev. A **36**, 740 (1987).

[152] U. Korkmaz and D. Türkpençe, Phys. Rev. A **107**, 012432 (2023).

[153] C. G. Feyisa and H. Jen, New J. Phys. **26**, 033038 (2024).

[154] A. Mikhalychev, S. Vlasenko, and S. Y. Kilin, Phys. Rev. A **105**, 063723 (2022).

[155] L. P. Bettmann, M. J. Kewming, and J. Goold, Phys. Rev. E **107**, 044102 (2023).

[156] L. F. Santos and G. T. Landi, arXiv:2404.13775 (2024).

[157] W. Khan, P. P. Potts, S. Lehmann, C. Thelander, K. A. Dick, P. Samuelsson, and V. F. Maisi, Nat. Commun. **12**, 5130 (2021).

[158] D. Zenelaj, P. Samuelsson, and P. P. Potts, arXiv:2410.14278 (2024).

[159] A. C. Barato and U. Seifert, Phys. Rev. Lett. **114**, 158101 (2015).

[160] I. Di Terlizzi and M. Baiesi, J. Phys. A **52**, 02LT03 (2018).

[161] G. T. Landi and M. Paternostro, Rev. Mod. Phys. **93**, 035008 (2021).

[162] N. Shiraishi, K. Funo, and K. Saito, Phys. Rev. Lett. **121**, 070601 (2018).

[163] T. Van Vu and Y. Hasegawa, Phys. Rev. Lett. **126**, 010601 (2021).

[164] A. Dechant, J. Phys. A **52**, 035001 (2018).

[165] N. Shiraishi, J. Stat. Phys. **185**, 19 (2021).

[166] K. Ptaszyński, Phys. Rev. B **98**, 085425 (2018).

[167] B. K. Agarwalla and D. Segal, Phys. Rev. B **98**, 155438 (2018).

[168] J. Liu and D. Segal, Phys. Rev. E **99**, 062141 (2019).

[169] K. Prech, P. Johansson, E. Nyholm, G. T. Landi, C. Verdozzi, P. Samuelsson, and P. P. Potts, Phys. Rev. Res. **5**, 023155 (2023).

[170] A. Rignon-Bret, G. Guarnieri, J. Goold, and M. T. Mitchison, Phys. Rev. E **103**, 012133 (2021).

[171] G. Guarnieri, G. T. Landi, S. R. Clark, and J. Goold, Phys. Rev. Res. **1**, 033021 (2019).

[172] Y. Hasegawa, Phys. Rev. Lett. **125**, 050601 (2020).

[173] T. Van Vu and K. Saito, Phys. Rev. Lett. **128**, 140602 (2022).

[174] K. Prech, P. P. Potts, and G. T. Landi, arXiv:2407.14147 (2024).

[175] M. Branderhorst, J. Nunn, I. Walmsley, and R. Kosut, New J. Phys. **11**, 115010 (2009).

[176] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).

[177] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, New J. Phys. **14**, 095022 (2012).

[178] A. Shabani, R. Kosut, M. Mohseni, H. Rabitz, M. Broome, M. Almeida, A. Fedrizzi, and A. White, Phys. Rev. Lett. **106**, 100401 (2011).

[179] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, Phys. Rev. A **107**, 042403 (2023).

[180] X. Zhang, Z. Wu, G. A. White, Z. Xiang, S. Hu, Z. Peng, Y. Liu, D. Zheng, X. Fu, A. Huang, *et al.*, arXiv:2312.06062 (2023).

[181] P. Busch, Int. J. Theor. Phys. **30**, 1217 (1991).

[182] G. D'Ariano, P. Perinotti, and M. Sacchi, J. Opt. B: Quantum Semiclass. Opt. **6**, S487 (2004).

[183] L. R. Rabiner, Proc. IEEE **77**, 257 (1989).

[184] D. R. Upper, *Theory and Algorithms for Hidden Markov Models and Generalized Hidden Markov Models*, Ph.D. thesis, University of California, Berkeley (1997).

[185] M. Gu, K. Wiesner, E. Rieper, and V. Vedral, Nat. Commun. **3**, 762 (2012).

[186] M. Cholewa, P. Gawron, P. Głomb, and D. Kurzyk, Quantum Inf. Process. **16**, 1 (2017).

[187] M. Zonnios, A. Boyd, and F. C. Binder, arXiv:2412.12812 (2024).

[188] C. Yang, F. C. Binder, M. Gu, and T. J. Elliott, Phys. Rev. E **101**, 062137 (2020).

[189] R. A. Fisher, J. R. Stat. Soc. **98**, 39 (1935).

[190] D. Hangleiter, I. Roth, J. Fuksa, J. Eisert, and P. Roushan, Nat. Commun. **15**, 9595 (2024).

[191] W. Yu, J. Sun, Z. Han, and X. Yuan, Quantum **7**, 1045 (2023).

[192] H.-Y. Hu, M. Ma, W. Gong, Q. Ye, Y. Tong, S. T. Flammia, and S. F. Yelin, arXiv:2502.11900 (2025).

[193] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O'Brien, J. G. Rarity, A. Laing, *et al.*, Nat. Phys. **13**, 551 (2017).

[194] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, New J. Phys. **14**, 103013 (2012).

[195] J. Hervas, A. Goldberg, A. Sanz, Z. Hradil, J. Řeháček, and L. Sánchez-Soto, Phys. Rev. Lett. **134**, 010804 (2025).

[196] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, Vol. 25 (Cambridge university press, 2009).

[197] S. Watanabe, in *2007 IEEE Symposium on Foundations of Computational Intelligence* (IEEE, 2007) pp. 383–388.

[198] R. Zamir, IEEE Trans. Inf. Theory **44**, 1246 (1998).

[199] C. Ferrie, Phys. Rev. A **90**, 014101 (2014).

[200] M. J. Schervish, *Theory of Statistics* (Springer Science & Business Media, 2012).

[201] G. Casella and R. Berger, *Statistical Inference* (CRC press, 2024).

[202] S.-i. Amari and H. Nagaoka, *Methods of Information Geometry*, Vol. 191 (American Mathematical Soc., 2000).

[203] L. Gao, H. Li, I. Marvian, and C. Rouzé, Commun. Math. Phys. **405**, 180 (2024).

[204] D. Petz, Q. J. Math. **39**, 97 (1988).

[205] J. A. Smiga and G. T. Landi, arXiv:2411.08214 (2024).

[206] M. Drazin, Am. Math. Mon. **65**, 506 (1958).

[207] M. J. Kewming, M. T. Mitchison, and G. T. Landi, Phys. Rev. A **106**, 033707 (2022).

[208] P. Taranto, F. A. Pollock, S. Milz, M. Tomamichel, and K. Modi, Phys. Rev. Lett. **122**, 140401 (2019).

[209] M. M. Wolf and D. Pérez-García, Phys. Rev. A **75**, 012303 (2007).

[210] T. Benoist, C. Pellegrini, and A. Szczepanek, arXiv:2409.18655 (2024).

[211] S. Wallace, Y. Altmann, B. D. Gerardot, E. M. Gauger, and C. Bonato, arXiv:2410.17942 (2024).

[212] S. M. Vermeulen, T. Cullen, D. Grass, I. A. MacMillan, A. J. Ramirez, J. Wack, B. Korzh, V. S. Lee, K. M. Zurek, C. Stoughton, *et al.*, Phys. Rev. X **15**, 011034 (2025).

[213] P. Guilmin, P. Rouchon, and A. Tilloy, arXiv:2410.11955 (2024).

[214] S. Jerbi, J. Gibbs, M. S. Rudolph, M. C. Caro, P. J. Coles, H.-Y. Huang, and Z. Holmes, arXiv:2303.12834 (2023).

[215] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, arXiv:2409.01706 (2024).

[216] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, and M. Barbieri, Phys. Rev. Lett. **123**, 230502 (2019).

[217] I. Gianani and C. Benedetti, AVS Quantum Sci. **5**, 014405 (2023).