

---

# New concepts in quantum-metrology: From coherent averaging to Hamiltonian extensions

---

## Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät  
der Eberhard Karls Universität Tübingen  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)

vorgelegt von

Julien Mathieu Elias Fraïsse  
aus Castres, Frankreich

Tübingen  
2017

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät  
der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 20.07.2017

Dekan: Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter: Prof. Dr. Daniel Braun

2. Berichterstatter: Dr. habil. Olivier Giraud





## Abstract

This thesis is dedicated to the understanding of the metrology of quantum systems by using the tools of quantum parameter estimation, in particular the quantum Fisher information (QFI).

**Our first project** deals with a specific protocol of quantum enhanced measurement known as *coherent averaging* [Braun and Martin, 2011]. This protocol is based on a star topology, with one central object, the so-called *quantum bus*, connected to  $N$  extra subsystems, called *probes*. For the estimation of a parameter characteristic of the interaction between the quantum bus and the probes, coherent averaging leads to a Heisenberg limited (HL) scaling for the QFI (QFI proportional to  $N^2$ ). Importantly this HL scaling can be obtained while starting with a separable state. This provides an advantage as generally one needs to use entangled states to achieve this scaling. Another important aspect in coherent averaging is the possibility to obtain the HL scaling by performing a measurement on the quantum bus only. These results were obtained using perturbation theory in the regime of weak interactions.

In this thesis we go one step further in the study of the coherent averaging protocol. We extend the formalism of perturbation theory to encompass the possibility of estimating any parameter, in the regimes of strong and weak interactions. To illustrate the validity of our results, we introduce two models as examples for a coherent averaging scheme. In these models both the quantum bus and all the probes are qubits. In the ZZXX model, the free Hamiltonians do not commute with the interaction Hamiltonians and we have to rely on numerics to find non-perturbative solutions. In the ZZZZ model the free evolution Hamiltonians commute with the interaction Hamiltonians and we can find the exact solution analytically.

Perturbation theory shows that in the strong interaction regime and starting with a separable state, we can estimate the parameter of the free evolution of the probes with a HL scaling if the free Hamiltonians do not commute with the interaction Hamiltonians. This is confirmed by the non-perturbative numerical results for the ZZXX model. In the weak interaction regime we only obtain a standard quantum limit (SQL) scaling for the parameter of the free evolution of the probes (QFI proportional to  $N$ ). When one has only access to the quantum bus, we show that the HL scaling found using the perturbation theory does not necessarily survive outside the regime of validity of the perturbation. This is especially the case as  $N$  becomes large. It is shown by comparing the exact analytical result to the perturbative result with the ZZZZ model. The same behaviour is observed with the ZZXX model using the non-perturbative numerical results.

**In our second project** we investigate the estimation of the depolarizing channel and the phase-flip channel under non-ideal conditions. It is known that using an ancilla can lead to an improvement of the channel QFI (QFI maximized over input states feeding the channel) even if we act with the identity on the ancilla. This method is known as channel extension. In all generality the maximal channel QFI can be obtained using an ancilla whose Hilbert space has the same dimension as the dimension of the Hilbert space of the original system. In this ideal scenario using multiple ancillas — or one ancilla with a larger Hilbert space dimension — is useless.

To go beyond this ideal result we take into account the possibility of losing either the probe or a finite number of ancillas. The input states considered are GHZ and W states with  $n + 1$  qubits (the probe plus  $n$  ancillas). We show that for any channel, when the probe is lost then all the information is lost, and the use of ancillas cannot help. For the phase-flip channel the introduction of ancillas never improves the channel QFI and ancillas are useless.

For the depolarizing channel the maximal channel QFI can be reached using one ancilla and feeding the extended channel with a Bell state, but if the ancilla is lost then all the advantage is lost. We show that the GHZ states do not help to fight the loss of ancillas: If one ancilla or more are lost all the advantage provided by the use of ancillas is lost. More interestingly, we show that the W states with more than one ancilla are robust against loss. For a given number of lost ancillas, there always exists an initial number of ancillas for which a W state provides a higher QFI than the one obtained without ancillas.

**Our last project** is about Hamiltonian parameter estimation for arbitrary Hamiltonians. It is known that channel extension does not help for unitary channels. Instead we apply the idea of extension to the Hamiltonian itself and not to the channel. This is done by adding to the Hamiltonian an extra term, which is independent of the parameter and which possibly encompasses interactions with an ancilla. We call this technique Hamiltonian extension. We show that for arbitrary Hamiltonians there exists an upper bound to the channel QFI that is in general not saturated. This result is known in the context of non-linear metrology. Here we show explicitly the conditions to saturate the bound.

We provide two methods for Hamiltonian extensions, called *signal flooding* and *Hamiltonian subtraction*, that allow one to saturate the upper bound for any Hamiltonian. We also introduce a third method which does not saturate the upper bound but provides the possibility to restore the quadratic time scaling in the channel QFI when the original Hamiltonian leads only to a periodic time scaling of the channel QFI.

We finally show how these methods work using two different examples. We study the estimation of the strength of a magnetic field using a NV center, and show how using signal flooding we saturate the channel QFI. We also consider the estimation of a direction of a magnetic field using a spin-1. We show how using signal flooding or Hamiltonian subtraction we saturate the channel QFI. We also show how by adding an arbitrary magnetic field we restore the quadratic time scaling in the channel QFI. Eventually we explain how coherent averaging can be scrutinized in the formalism of Hamiltonian extensions.

## Zusammenfassung

Diese Arbeit ist der Metrologie von Quantensystemen gewidmet, speziell der Schätzung von Quantenparametern und der Quanten Fisher Information (QFI).

**In unserem ersten Projekt** betrachten wir „kohärentes Mitteln“ [Braun and Martin, 2011], ein spezifisches Protokoll für optimierte Quantenmessungen („quantum enhanced measurement“). Dieses Protokoll basiert auf einer Sterntopologie: Das zentrale Objekt, der sogenannte *Quanten-Bus*, ist mit  $N$  Untersystemen, den sogenannten Sonden, verknüpft. Für einen Parameter, der die Wechselwirkung zwischen dem Quanten-Bus und den Sonden beschreibt, führt kohärentes Mitteln zu einer Heisenberg-Skalierung (HL) der QFI (das heißt, dass die QFI proportional zu  $N^2$  ist), wenn man mit einem separierbaren Zustand beginnt. Dies ist ein Vorteil gegenüber anderen Messungen, wo man im Allgemeinen diese Skalierung nur mit einem verschränkten Zustand erreichen kann. Zudem ist es mit kohärentem Mitteln möglich, diese Skalierung zu erhalten, wenn man nur den Quanten-Bus misst. Zu diesen Erkenntnissen kam man, indem man Störungstheorie im Bereich der schwachen Wechselwirkung angewendet hat.

In dieser Arbeit gehen wir einen Schritt weiter. Wir erweitern den Formalismus der Störungstheorie, sodass wir sie für einen beliebigen Parameter des Protokolls im Bereich der schwachen sowie der starken Wechselwirkung anwenden können. Um die Gültigkeit unserer Resultate zu untermauern, betrachten wir zwei Beispiele, das ZZXX- und das ZZZZ-Modell. In beiden Modellen sind sowohl der Quanten-Bus als auch die Sonden qubits. Im ZZXX-Modell kommutieren die freien Hamilton-Operatoren nicht mit den Wechselwirkungs-Hamilton-Operatoren. Es stellt sich heraus, dass man in diesem Fall nicht-störungstheoretische Resultate nur numerisch finden kann. Im ZZZZ-Modell kommutieren die freien mit den Wechselwirkungs-Hamilton-Operatoren, und man findet exakte Lösungen analytisch.

Mithilfe der Störungstheorie im Bereich der starken Wechselwirkung kann man zeigen, dass für die Schätzung eines Parameters der freien Zeitentwicklung der Sonden eine HL-Skalierung erreicht werden kann, falls man mit einem separierbaren Zustand beginnt. Diese Erkenntnis können wir mit numerischen Rechnungen ohne Verwendung der Störungstheorie bestätigen. Im Bereich der schwachen Wechselwirkung hingegen findet man in dieser Situation nur eine SQL-Skalierung (QFI proportional zu  $N$ ). Falls wir nur Zugang zum Quanten-Bus und nicht zu den Sonden haben, finden wir, dass die HL-Skalierung ausserhalb des angegebenen Gültigkeitsbereichs der Störungstheorie nicht unbedingt erhalten bleibt, vor allem falls die Anzahl der Sonden  $N$  groß wird. Dies zeigen wir anhand des ZZZZ-Modells, indem wir die mit der Störungstheorie gefundenen Resultate mit den analytischen Ausdrücken vergleichen. Anhand des ZZXX-Modells können wir dies ebenfalls illustrieren, indem wir die mit der Störungstheorie gefundenen Resultate mit numerischen Werten der exakten Lösungen vergleichen.

**In unserem zweiten Projekt** betrachten wir die Schätzung des „Depolarisierungs-Kanals“ und des „Phasenflip-Kanals“ unter nicht-idealen Bedingungen. Bekanntlich kann mit einem Hilfsystem („ancilla“) eine verbesserte Kanal-QFI (QFI maximiert über alle Anfangszustände) erreicht werden. Diese Methode nennt man auch Kanalerweiterung. Generell kann man die maximale QFI mit einem Hilfssystem erreichen, dessen Hilbertraum die gleiche Dimension hat wie der Hilbertraum des ursprünglichen Systems. In diesem Fall nützt es nichts, mehrere Hilfssysteme oder ein Hilfssystem mit einem höher-dimensionalen Hilbertraum zu verwenden.

Um dies zu verallgemeinern, betrachten wir zusätzlich die Möglichkeit, entweder die Sonde oder eine gewisse Anzahl der Hilfssysteme zu verlieren. Die betrachteten Anfangszustände sind GHZ- und W-Zustände mit  $n + 1$  qubits (die Sonde plus  $n$  Hilfssysteme). Wir zeigen für einen beliebigen Kanal, dass alle Information verloren geht, wenn man die Sonde verliert. Es hilft in dem Fall auch nicht, Hilfssysteme zu verwenden. Für den Phasenflip-Kanal führen Hilfssysteme nie zu einer verbesserten Kanal-QFI.

Für den Depolarisierungs-Kanal kann man die maximale QFI mit einem Hilfssystem und einem Bell-Zustand als Anfangszustand erreichen. Geht hingegen ein qubit verloren, hat man keinen Vorteil. Wir zeigen, dass die GHZ-Zustände nicht helfen, um robuster gegen die Verluste zu werden: Wenn ein oder mehrere Hilfssysteme verloren gehen, dann verschwindet der gesamte Vorteil. W-Zustände hingegen sind, wie wir zeigen, robust gegenüber Verlusten von Hilfssystemen: Gehen eine beliebige Anzahl Hilfssysteme verloren, existiert immer eine gewisse Anzahl Hilfssysteme zu Beginn, sodass man eine höhere QFI erhält, als man ohne Hilfssysteme erhalten würde.

**In unserem letzten Projekt** untersuchen wir Hamiltonsche Parameterschätzung für beliebige Hamilton-Operatoren. Es ist bekannt, dass Kanalerweiterung für unitäre Kanäle keinen Vorteil bringen. Hier erweitern wir nicht den Kanal, sondern den Hamilton-Operator selbst: Wir addieren zum Hamilton-Operator einen zusätzlichen Term, der unabhängig vom Parameter ist und eventuell eine Wechselwirkung mit einem Hilfssystem enthält. Dieses Vorgehen nennen wir Hamiltonsche Erweiterung. Wir zeigen für einen beliebigen Hamilton-Operator, dass es eine obere Schranke für die Kanal-QFI gibt, die in der Regel nicht erreicht wird. Dies ist bekannt im Zusammenhang mit nicht-linearer Metrologie. Hier geben wir explizite Bedingungen an, unter denen die Schranke erreicht wird.

Mit zwei Varianten einer Hamiltonschen Erweiterung, genannt „Signalfutung“ und „Hamiltonsche Subtraktion“ lässt sich die obere Schranke für jeden Hamilton-Operator erreichen. Mit einer dritten Variante kann man zwar die obere Schranke nicht erreichen, aber man kann eine quadratische Zeitskalierung in der Kanal-QFI erhalten, was mit dem ursprünglichen Hamilton-Operator nicht möglich ist.

Die Methoden illustrieren wir anhand von Beispielen: Wir betrachten die Schätzung der magnetischen Feldstärke mittels eines NV-Zentrums. Wir zeigen, wie man mit der Signalfutung die maximale Kanal-QFI erreichen kann. Weiter untersuchen wir die Schätzung der Richtung eines Magnetfeldes anhand eines Spin-1. Mit der Signalfutung sowie mit der Hamiltonschen Subtraktion können wir die QFI maximieren. Wir zeigen auch, wie wir die quadratische Zeitskalierung erreichen können, indem wir ein zusätzliches Magnetfeld betrachten. Zuletzt erklären wir, wie man die Hamiltonsche Erweiterung im kohärenten Mitteln verwenden kann.

# Summary

|  |            |
|--|------------|
| Abstract   | i          |
| Zusammenfassung  | ii         |
| Summary  | iv         |
| Abbreviations  | vi         |
| Introduction   | 1          |
| <b>I Theoretical Framework</b>                                       | <b>7</b>   |
| 1 Parameter estimation theory  | 9          |
| 2 Foundation of quantum theory                                       | 33         |
| 3 Quantum parameter estimation theory                                | 51         |
| 4 Quantum metrology and quantum enhanced measurements                | 83         |
| <b>II Coherent averaging</b>   | <b>119</b> |
| 5 Coherent averaging protocol and metrology                          | 121        |
| 6 Qubit model: ZZZZ Hamiltonian                                      | 133        |
| 7 Qubit model: ZZXX Hamiltonian                                      | 147        |
| <b>III Extensions for metrology</b>                                  | <b>163</b> |
| 8 Channel extension for qubit errors: robustness of W and GHZ states | 165        |
| 9 Hamiltonian engineering  | 185        |
| Conclusion   | 207        |

|   |            |
|---|------------|
| <b>Appendices</b>   | <b>211</b> |
| <b>A Probability theory</b>   | <b>211</b> |
| <b>B Perturbation theory</b>  | <b>219</b> |
| <b>C Perturbative result for coherent averaging</b>   | <b>223</b> |
| <b>D Diagonalization of the matrix <math>K</math> and <math>G</math> for channel estimation</b> | <b>229</b> |
| <b>Bibliography</b>   | <b>241</b> |
| <b>List of figures</b>  | <b>244</b> |
| <b>List of tables</b>   | <b>245</b> |
| <b>Contents</b>   | <b>252</b> |

# Introduction

*La Convention nationale, convaincue que l'uniformité des poids et mesures est un des plus grands bienfaits qu'elle puisse offrir à tous les citoyens français [...]*

Convention nationale, Décret du 1er août 1793

1.1 (1.1)

## **quantity**

*property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference*

2.1 (2.1)

## **measurement**

*process of experimentally obtaining one or more quantity values that can reasonably be attributed to a **quantity***

2.2 (2.2)

## **metrology**

*science of **measurement** and its application*

BIPM, *International Vocabulary of Metrology — Basic and General Concepts and Associated Terms* (VIM 3rd edition)

## Metrology and physics

The question of measurement is clearly a cornerstone of physics. Indeed physics aims to explain the laws of nature, which implies that we should observe, and thus measure, the world. Physical theories are validated or invalidated by comparing the predictions of a theory to the experimental results. This trivial statement is there to remind us how metrology, even beyond the question of defining units, is crucial in physics. For this reason, from the birth of modern physics in the XVII century to nowadays, physicists tried constantly to design measurements that were more and more precise. From this point of view the discovery and then the growth of quantum mechanics played a double role. On one side quantum physicist carried out experiments to better understand their own field. While at the beginning of the century experimentalist had access only to macroscopic objects, at the end of the century researchers were able to manipulate individual quantum objects. This pushed the theoreticians to better understand the concept of measurement and played a big role in the raise of quantum information.

On the other side the technologies resulting from quantum mechanics, and among them the laser in particular, allowed to design devices way more sensitive than what was used before. This pushed the community to understand metrology at a quantum level. The main motivation was the measurement of weak classical forces for the detection of gravitational waves. The level of sensitivity needed to detect them is so high that one should take into account the quantum behaviour

## Introduction

of the detector. This field led to many controversies and debates about the quantum limits — meaning the limits derived from quantum mechanics — on the sensitivity. These controversies were mainly due to the lack of a unified framework. Luckily, in the mid-sixties the theory of parameter estimation was transposed to the quantum world. Quantum parameter estimation theory, also known as q-pet, led to two main advances: First it allowed to go beyond the idea that all what is measurable in quantum mechanics are observables. To be more precise it made clear that the estimation stage is fundamentally different from the measurement stage. Second, q-pet provides fundamental figures of merit to assess the best sensitivity that one can reach given a state or a process (quantum channel).

Eventually, at the turn of the millennium, the use of the tools of q-pet in the context of quantum metrology helped to clarify the debates raised by gravitational wave detection. It allowed to define more precisely the concepts of the standard quantum limit (SQL) and the Heisenberg limit (HL). The former represents usually the best sensitivity achievable by classical means, while the latter corresponds to the best sensitivity offered using all the power of quantum mechanics. Behind this dichotomy a change of paradigm appears: Quantum mechanics is not only viewed as a fundamental topic to investigate, but also as a tool that one can use to its own benefit. In the context of metrology, quantum enhanced measurements illustrate this change of point of view.

While it is now clear that quantum mechanics can bring an advantage in terms of metrology, there are still many pending questions. Among them the question of the necessity of entanglement for obtaining an enhancement appears to be particularly important. In the usual derivation of the SQL and of the HL, a linear Hamiltonian is used, and the SQL is obtained by starting with a separable state while the HL is obtained by starting with an entangled state. It turns out that by using non-linear Hamiltonians, entanglement is not necessary. In this context, the coherent averaging scheme allows to reach the HL while starting with a separable state. The coherent averaging scheme is based on a central system called the quantum bus, connected to  $N$  independent subsystems, and it was introduced to reach a HL for a parameter characterizing the coupling between the bus and the subsystems. The drawback with this analysis is that this parameter apparently does not characterize neither the bus nor the ancillas. Then it is not clear if the coherent averaging scheme is really useful when it goes to the estimation of some parameter that exists independently of the scheme itself. We investigated this question by extending the previous perturbative results already known and by introducing a toy model composed of qubits. This allowed us to give hints that the coherent averaging scheme can also be of use for estimating parameters characterizing the subsystems.

A crucial point for quantum technologies is the question of noise or decoherence, modelled by non-unitary quantum channels. In general a quantum system in contact with an environment will suffer decoherence, and this decoherence may ruin all the advantage provided by quantum mechanics. Then it is important to check how a given protocol — metrological to our concern — is resistant against decoherence, and eventually how we can fight the deleterious effects of noise. This implies as a prerequisite to identify the noise properly. In particular it is important to identify the noise affecting qubits (two-level systems), as they are at the base of many quantum informational or computational tasks. This can be done using the tools of quantum metrology itself. It turns out that the optimal protocol for an important class of channels — called Pauli channels — is to entangle the qubit with an ancilla. The fact that we estimate a non-unitary quantum channel does not mean that this estimation procedure does not suffer from noise itself. In the worst case one of the two qubits can be lost. In this context we study how the use of multiple qubit ancillas can help to keep an advantage in comparison of the situation where no ancilla is used.

Eventually the investigation on the use of ancillas in the context of estimation of the noise

affecting a qubit led us to consider the use of ancillas for the estimation of parameters encoded in Hamiltonians (unitary channels). In this field most of the research previously done focused on phase-shift Hamiltonians. We address the question of the use of ancillas and in general of the maximization of the channel QFI for arbitrary Hamiltonians.

## Thesis overview

This thesis is divided into three parts, organized as follows:

- A first part where all the theoretical framework is derived.
- A second part dedicated to the study of the coherent averaging scheme.
- A third part dealing with channel extensions and Hamiltonian extensions.

## Description of the thesis

**Part I** The part I is devoted to the study of the theoretical framework underpinning this thesis. In Chapter 1 we review the basics of parameter estimation theory. We start from the very definition of the problem of parameter estimation before presenting some estimators. The most important notions there are undoubtedly the Fisher information (FI) and the Cramér-Rao theorem, as they are at the bedrock of quantum parameter estimation. The analysis of moment estimators helps to give later on a precise meaning to the signal-to-noise ratio used in quantum parameter estimation theory.

Chapter 2 contains a summary of quantum mechanics which helps to settle the notation. As it plays a big role in our studies, we pay special attention to the study of two-level systems. We then turn to the study of quantum parameter estimation in Chapter 3. We present the extremely important quantum Cramér-Rao theorem, provide two proofs and study extensively the quantum Fisher information. Some recent advances in channel estimation and Hamiltonian parameter estimation are also presented.

Finally the part I ends with a presentation of quantum metrology and quantum enhanced measurement in Chapter 4. While the first three chapters were rather formal and mathematical — as they present the tools used in our research — Chapter 4 emphasizes on more historical aspects and is written in a review spirit. We do not insist on calculations and rather try to give an overview of the large field that quantum metrology became. Still, in this chapter the standard quantum limit and the Heisenberg limit — which play a crucial role in the study of the coherent averaging scheme — are presented.

**Part II** The part II of the thesis is entirely devoted to the study of coherent averaging. The coherent averaging scheme is formally introduced in Chapter 5. We also provide the formalism of perturbative parameter estimation for the three different kinds of parameters encountered in the coherent averaging scheme. In Chapter 6 we introduce the ZZZZ Hamiltonian for coherent averaging. Exact and perturbative solutions are provided. Chapter 7 focuses on the ZZXX Hamiltonian. No exact analytical solution can be provided and we thus use numerics and perturbation theory to analyze the sensitivity for the different parameters.

**Part III** The part III is concerned with channel extensions and Hamiltonian extensions. Chapter 8 starts with a review of the estimation of Pauli channels — with a focus on the depolarizing and phase-flip channel— in both the ideal and non-ideal situations. We then turn our attention to the effect of using GHZ and W states with a large number of ancillas, before looking at the consequences of a loss of these ancillas.

Chapter 9 deals with Hamiltonian extensions. After formalizing the notion of Hamiltonian extension and making clear its difference with channel extension we derive an upper bound for the estimation of Hamiltonian parameters. Using two different methods of Hamiltonian extensions we show how we can saturate the upper bound. We also discuss the time scaling in Hamiltonian parameter estimation and we provide a Hamiltonian extension to improve it under certain conditions. Eventually we conclude this topic by studying two examples, namely the magnetometry with NV centers and the estimation of the direction of a magnetic field using a spin-1.

**Appendices** The appendix A presents an introduction to probability theory and statistics. It is there mainly to formalize some notions and result used in chapter 1. The appendix B is a reminder about perturbation theory in the interaction picture. Finally the appendix C and D give some technical results related respectively to coherent averaging and to the estimation of depolarizing and phase-flip channels.

## Structure of the thesis

We made the choice to rather strictly separate our own results from the theoretical background for at least two reasons. First, due to the variety of the topics studied it would have been hard, but moreover counter-productive, to try to write this thesis as a long, unified story. Indeed the common denominator of our work lies more in the tools used than in a specific question that motivates all the research. As it was then clear that we have to present in a clearly divided way our results, it appeared natural to group the presentation of the tools and of the known results in a separate part.

The second motivation for this separation is related to the question of the potential readers of this thesis. We are interested in quantum metrology, a field based mainly on two pillars: Quantum mechanics in the formalism of quantum information and parameter estimation theory. At the same time this field can potentially interest readers from various communities of physics: all the researchers dealing with quantum mechanics in general, from solid state physics to quantum optics, but also researchers from quite different fields, like researchers from general relativity (cf. gravitational wave detection), or engineers who are interested in developing extremely sensitive detectors.

The drawback is that parameter estimation theory is usually not taught during physics studies. Moreover, quantum mechanics as presented in quantum information can appear exotic to researchers from other fields of quantum mechanics (POVMs, quantum channels, distance between states). For all these reasons, we wanted to provide all the content necessary for the understanding of this thesis to a reader neither familiar to parameter estimation theory nor to quantum information. Strictly speaking we go far beyond what is necessary to know to understand our work. Still we believe that having a wider view allows to go deeper into the topic.

## How to read this thesis

For the reader who is not familiar with parameter estimation theory nor with quantum metrology there can be different ways to read this thesis. If the goal is to have the very basic knowledge to

understand our work it should be enough to read the summary of the first four chapters. This would provide only the minimum knowledge, and it will still be advised to have a deeper look at the chapter 3, potentially skipping the proofs. In general the chapter 1 will interest the readers that wonder from where the concept of QFI or signal-to-noise ratio originates. The reader interested in quantum metrology but not in the history of the field can read the chapter 4 but skip the first section of it.

In every chapter, we introduce boxes that are separated from the text. They are there to either formulate something more precisely, give an example or present some result that goes beyond the scope of the discussion.

Concerning the results, each topic can be read independently: The chapters 5, 6 and 7 for coherent averaging; the chapter 7 for the estimation of depolarizing and phase-flip channels under presence of loss; the chapter 8 for Hamiltonian extensions.



Part I

**Theoretical Framework**



# Chapter 1

## Parameter estimation theory

This chapter is devoted to parameter estimation theory. We start by presenting the basic tools for parameter estimation, introducing estimators and their desirable properties. After presenting the moment estimator and the maximum likelihood estimator we look at the Cramér-Rao bound, which lies at the core of all our studies in this thesis. We then discuss in detail the question of locality in estimation and of saturation of the bound.

Most of the results presented in this chapter are textbook standards. One of the first book written on the topic is the one from H. Cramer [Cramér, 1946]. A more modern presentation can be found in [Kay, 1993; Roussas, 1997; Trees, 2013; George Casella, 1998]. Finally the standard reference of information geometry is [Amari and Nagaoka, 2007].

### 1.1 The problem of parameter estimation

The idea in parameter estimation is pretty simple: we consider a random process which is known, but depending on one or several parameters which are *unknown*. By repeating the process, we acquire a set of data and from it we try to *infer* the value of the unknown parameters.

Involving random processes, the proper formalization of parameter estimation theory is done by using the theory of probability. The modern treatment of probabilities is based on the axiomatic approach due to Kolmogorov. Based on measure theory, this approach proved powerful during the past century. For the purpose of our thesis, it is not indispensable to present this axiomatic and formalized approach. It can still be helpful for certain readers or when we want to refer to some specific results clearly stated without making the text too cluttered. Therefore we present this axiomatic approach in the appendix A.

#### 1.1.1 Parametric families, $n$ -samples and estimators

##### Random variables

A random processes is modelled by a real random variable  $X$ . We call a *realization* of  $X$  the value observed when we implement the random process. We also sometimes refer to a realization as an *observation*. When the result of the process is a real number we have a real scalar random variable and when the random process has multiple real outcomes we have a real multivariate random variable. Since we will deal only with real random variables we will drop the "real", and

when we do not need to specify if the random variable is scalar or multivariate we will only use the term "random variable" alone.

When the set of possible results of the process is countable we say that the random variable is *discrete*. We denote the probability that  $X$  takes the value  $x_i$  as  $\mu(x_i)$  or  $\mu(X = x_i)$  and we say that the random variable  $X$  is distributed according to  $\mu$ . When the set of possible results of the process is not countable the random variable is said to be continuous. We then introduce the probability density  $\mu(x)$  (see Appendix A for a formal treatment of those notions). The normalization condition of the probability is written as

$$\sum_i \mu(x_i) = 1 \quad \text{or} \quad \int dx \mu(x) = 1, \quad (1.1)$$

respectively for a discrete or continuous random variable.

In general we will not specify if a random variable is continuous or discrete. We will use the term "probability distribution" to refer to both probabilities and probability densities, and we always use the notation  $X \sim \mu$  to say that a random variable is distributed according to  $\mu$ . Notwithstanding, when writing equations we need to choose the symbols depending if the random variable is continuous or discrete. For concision we will use the notations for continuous random variables, assuming that it applies also to the discrete case, and making it explicit if it is not the case.

### Parametric families and $n$ -samples

Let us now introduce the exact framework of classical parameter estimation theory. We are given a random variable  $X$  and a parametric family of distributions  $\{\mu_\theta\}_{\theta \in \Theta}$ . This means that we know that  $X$  is distributed according to a certain  $\mu_\theta$  but we ignore the value of  $\theta$ . Here,  $\theta$  is the parameter that we want to *estimate*. When the parameter space  $\Theta$  is a subset of  $\mathbf{R}$  we say that it is a *scalar* parameter, while when it is a subset of  $\mathbf{R}^n$  we refer to it as a *vector parameter*. All the problems that we will deal with in the subsequent chapters on quantum metrology correspond to scalar parameter, and therefore we will focus on the theory of estimation for scalar parameters.

Importantly, we assume that the parameter  $\theta$  is a deterministic parameter, and not a random parameter. This means that the parameter  $\theta$  has a fixed value which does not vary from one realization of the process to another one. The case where  $\theta$  is considered itself as a random variable is treated by Bayesian parameter estimation (see Box 1).

Often the estimation of the parameter is done by realizing several times the process, collecting all the outcomes (which constitutes the data) and then trying to infer a value of  $\theta$  using this data. This situation is best described by using  *$n$ -samples*. In the language of probability theory an  $n$ -sample is a multivariate random variable  $X^{(n)}$  composed of  $n$  times the original random variable  $X$ ,

$$X^{(n)} := (X_1, X_2, \dots, X_n) \equiv \{X_i\} \quad \text{with} \quad X_i \sim \mu_\theta. \quad (1.2)$$

A  $n$ -sample is thus a sample of size  $n$  of identically and independently distributed (*i.i.d.*) random variables. In statistics the  $n$ -sample gets sometimes confused with the data, which strictly speaking is the realization of the  $n$ -sample and not the  $n$ -sample itself. The confusion between random variables and their realization is indeed something that one has to pay attention to. Since the  $X_i$ s are independent it is easy to write the joint distribution  $\mu_\theta^{(n)}$  of the  $n$ -sample as it simply factorizes:

$$\mu_\theta^{(n)}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mu_\theta(X_i = x_i). \quad (1.3)$$

We are now able to formulate the main question of parameter estimation theory : *knowing the form of a probability distribution and having a realization of an  $n$ -sample of it, what can we say about the value of  $\theta$ ?*

## Estimators

Although more precise, this question remains vague and we have to define more clearly the goals of the estimation. The very first constraint is that our guess of the value has to be realistic. Say we want to estimate a temperature given in Kelvin. If after realizing the process  $n$  times we infer from the data a negative value we can clearly say that it is a pretty bad guess. Our first obvious requirement will thus be that the guessed value of  $\theta$  belongs to the parameter space  $\Theta$ . Indeed this requirement defines the concept of an estimator, which we present formally now.

Suppose a given random variable  $X$  along with a parametric family of distributions  $\{\mu_\theta\}_{\theta \in \Theta}$ . Assume that  $X$  takes values in  $\mathcal{A}$ . A first way to define an estimator is by saying that an estimator is a function  $\hat{\theta}_{\text{est}}$  that takes data as input and gives as output a guess, called *estimate*, which belong to the parameter space  $\Theta$ . Schematically we have

$$(X_1, X_2, \dots, X_n) \xrightarrow{\text{realization}} (x_1, x_2, \dots, x_n) \xrightarrow{\hat{\theta}_{\text{est}}} \hat{\theta}_{\text{est}}(x_1, x_2, \dots, x_n) \in \Theta. \quad (1.4)$$

A second and somehow more useful way to define an estimator is the following: We define an *estimator* of the parameter  $\theta \in \Theta$  as a function  $\hat{\theta}_{\text{est}}$  of the  $n$ -sample of  $X$  whose realization takes values in  $\Theta$ . Being a function of random variables, the estimator itself is a random variable. Schematically we have

$$(X_1, X_2, \dots, X_n) \xrightarrow{\hat{\theta}_{\text{est}}} \hat{\theta}_{\text{est}}(X_1, X_2, \dots, X_n) \xrightarrow{\text{realization}} \hat{\theta}_{\text{est}}(x_1, x_2, \dots, x_n) \in \Theta. \quad (1.5)$$

In both cases the form of the function is the same but we do not make it acting on the same objects. In the first case it acts on a vector of real numbers, and in the second case it acts on a random variable. We will adopt the second definition, and we will confuse the function and its result, meaning that we call estimator both  $\hat{\theta}_{\text{est}}$  and  $\hat{\theta}_{\text{est}}(X_1, X_2, \dots, X_n)$ . It is important to distinguish the estimator and a realization of it, an estimate.

To summarize, it will be enough to remember that an estimator is a function going from the  $n$ -sample to the space of parameters. The  $n$ -sample being a random variable, the estimator is also a random variable. Importantly, this means that we can define the mean or the variance of an estimator.

### 1.1.2 Good and bad estimators

#### Risk function

Now that we defined formally what an estimator is, we would like to have some way to characterize it. Indeed we can build many different estimators but not all of them perform equally well. For example we can always build a constant estimator, *i.e.* an estimator that gives back always the same value. If this value belongs to the parameter space then our estimator fulfils the definition, but we intuitively feel that it is a pretty bad estimator. To go beyond intuition we need some figure of merit to say if an estimator is good or not.

In general we can build a figure of merit by designing a cost function and a risk function. The cost function  $c(\hat{\theta}_{\text{est}}(\{x_i\}), \theta)$  is a function that is positive and equal to 0 when the estimate equals

## Box 1: Bayesian parameter estimation

The parameter estimation theory that we present in this thesis takes the so-called *frequentist* approach of parameter estimation theory. The other famous approach is the *Bayesian* parameter estimation theory. The main difference between them is how we interpret the parameter. In frequentist approach the parameter, obviously unknown, is assumed to have a defined value, fixed, that in principle we could guess arbitrarily precisely. In Bayesian estimation, the parameter itself is a random variable. And what we want to learn is its distribution. Then not only the data matters, but also the *prior* information on the parameter. The typical procedure is then to use the prior and the observed data (likelihood) to produce the updated distribution, called the posterior:

$$\text{posterior} \propto \text{prior} \times \text{likelihood} . \quad (1.6)$$

the true value of the parameter:  $c(\theta, \theta) = 0$ . The risk function is defined as the average of the cost function  $r(\hat{\theta}_{\text{est}}(\{X_i\}), \theta) = E[c(\hat{\theta}_{\text{est}}(\{X_i\}), \theta)]$ . The optimal estimator will be the estimator that minimizes the risk function.

A common cost function is given by the squared difference between the estimate and the parameter which leads to a risk function called Means Square Error (MSE) and defined as

$$\text{MSE}[\hat{\theta}_{\text{est}}] := E[(\hat{\theta}_{\text{est}} - \theta)^2] . \quad (1.7)$$

We see that the MSE is an intuitive figure of merit that we seek to minimize. We want that on average the distance between the estimate and the true value of the parameter is as small as possible. Unfortunately there is in general no estimator that minimizes the MSE for all values of  $\theta$ . Therefore, instead of looking for the estimator that minimizes the MSE, or any other risk function, we will first add some arbitrary, but justified, constraint to the estimators before coming back to the risk function.

### Variance of estimators

A reasonable constraint to impose on estimators is the unbiasedness condition. The bias  $b(\theta)$  of an estimator is defined as the difference between its expectation value and the true value of the parameter

$$b(\hat{\theta}_{\text{est}}) := E[\hat{\theta}_{\text{est}}] - \theta . \quad (1.8)$$

Notice that the bias is not a random variable and that in general it is a function of the parameter. An estimator is said to be *unbiased* when its bias vanishes for all values of the parameter,  $b(\hat{\theta}_{\text{est}}) = 0$ , which means that the expectation value of the estimator equals the true value of the parameter,  $E[\hat{\theta}_{\text{est}}] = \theta$ . The unbiasedness condition, when fulfilled, ensures us through the frequentist interpretation of probabilities that by making a large number of repetitions of the estimation process we will end up with the true value of the parameter.

In general the MSE can be written as a function of the variance and the bias,

$$\text{MSE}[\hat{\theta}_{\text{est}}] = \text{Var}[\hat{\theta}_{\text{est}}] + b(\hat{\theta}_{\text{est}})^2 , \quad (1.9)$$

where the variance  $\text{Var}[\hat{\theta}_{\text{est}}]$ , is defined as

$$\text{Var}[\hat{\theta}_{\text{est}}] := \text{E}[(\hat{\theta}_{\text{est}} - \text{E}[\theta])^2]. \quad (1.10)$$

The variance measures the average distance of a random variable from its expectation value, and thus for unbiased estimators to the true value of the parameter. Especially, when the estimator is unbiased, we have by definition  $b(\hat{\theta}_{\text{est}}) = 0$ , and thus the MSE becomes equal to the variance.

Our quest of the best estimators lead us to two properties:

- i) We look exclusively for estimators that are unbiased, as it guarantees us to find in average the true value of the parameter.
- ii) Among those unbiased estimators we look for the one with minimal variance — it will be the one that gives the closest result to the true value of the parameter for a finite sample.

The estimator satisfying these two properties is called the Minimum-Variance Unbiased (MVU) estimator and we denote it as  $\hat{\theta}_{\text{mvu}}$ .

### Asymptotic behaviour and consistency

The discussion on bias and variance of estimators showed that it is in general important to distinguish between properties that hold in general and properties that are fulfilled only in the limit of large samples.

A useful asymptotic property of estimators is the consistency. We say that an operator is *consistent* when it asymptotically converges to the true value of the parameter,

$$\lim_{n \rightarrow \infty} \hat{\theta}_{\text{est}} = \theta. \quad (1.11)$$

Here convergence refers to a convergence in probability, meaning that the distribution of the estimator should converge to a Dirac distribution centred in  $\theta$ .

The requirement of consistency for an estimator is less demanding than the unbiasedness condition. Biased estimators can be consistent, as long as their bias decreases with the size of the  $n$ -sample. Similarly, consistency does not impose a condition on the speed at which the variance goes to zero, but just ask for a convergence of the variance to zero with increasing  $n$ .

### Empirical moments of a $n$ -sample

To define the bias and the dispersion of an estimator we used the first moment and the second centred moment. These moments characterize the estimator as a random variable, but are not themselves random variables. When dealing with  $n$ -samples, we can also define empirical moments. These empirical moments are still random variables, and thus take different values at each realization. The *empirical moment*  $\tilde{m}_k$  of order  $k$  for an  $n$ -sample  $\{X_i\}$  is defined as

$$\tilde{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k. \quad (1.12)$$

Being random variables, we can compute the moments of the empirical moments. For example the first moment of the first empirical moment reads

$$\text{E}[\tilde{m}_1] = \frac{1}{n} \sum_{i=1}^n \text{E}[X_i^k] = \text{E}[X] = m_1, \quad (1.13)$$

meaning that in average the empirical mean of a sample equals the original mean of the random variable. This is indeed more general, the expectation value of the empirical moment of order  $k$  equals the moment of order  $k$  of the original random variable

$$\mathbb{E}[\tilde{m}_k] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^k] = \mathbb{E}[X^k] = m_k. \quad (1.14)$$

## 1.2 Some estimators

To make the concept of estimators clearer we present three specific estimators.

### 1.2.1 Constant estimator

This is the simplest estimator one could think of. A constant estimator is defined by  $\hat{\theta}_{\text{est}} = c$  with  $c \in \Theta$ . We see that it fulfils the definition of an estimator, being a function from the  $n$ -sample to the parameter space. But it is an extremely poor estimator since it gives always the same value, whatever are the observations, and is clearly not unbiased for all values of  $\theta$ .

### 1.2.2 Moment estimator

The method of moment estimators is probably the most intuitive method for estimating a parameter, and is historically one of the first methods developed. The idea is to use the dependency of the moments on the parameter. The estimator is obtained by equating one theoretical moment of the random variable to the corresponding empirical moment of the  $n$ -sample.

#### Definition

Consider a random variable  $X \sim \mu_\theta$ . The moment  $m_k$  of order  $k$  is in general a function of  $\theta$ ,  $m_k = m_k(\theta)$ . The  $k^{\text{th}}$  moment estimator  $\hat{\theta}_{m_k}$  is obtained by solving the equation

$$m_k(\hat{\theta}_{m_k}) = \tilde{m}_k. \quad (1.15)$$

Assuming that the function  $m_k$  has an inverse  $m_k^{-1}$ , we can write  $\hat{\theta}_{m_k}$  as

$$\hat{\theta}_{m_k} := m_k^{-1}(\tilde{m}_k). \quad (1.16)$$

Obviously, a necessary condition for using this method is that the moment chosen depends on the parameter to estimate. In general, although a distribution depends on a parameter  $\theta$ , not all of the moments depend on this parameter.

When the parameter to estimate is itself a moment, then the estimator is directly given by the empirical moment, as in this case the function  $m_k(\theta) = \theta$  is the identity, and the inverse of it is obviously the identity too. As we already noticed (Eq. (1.14)), the expectation value of the empirical moment is equal to the theoretical moment. Therefore the empirical moments are unbiased estimators of the theoretical moments. But is that true in general? We will explore this question in the specific case of the first moment estimator.

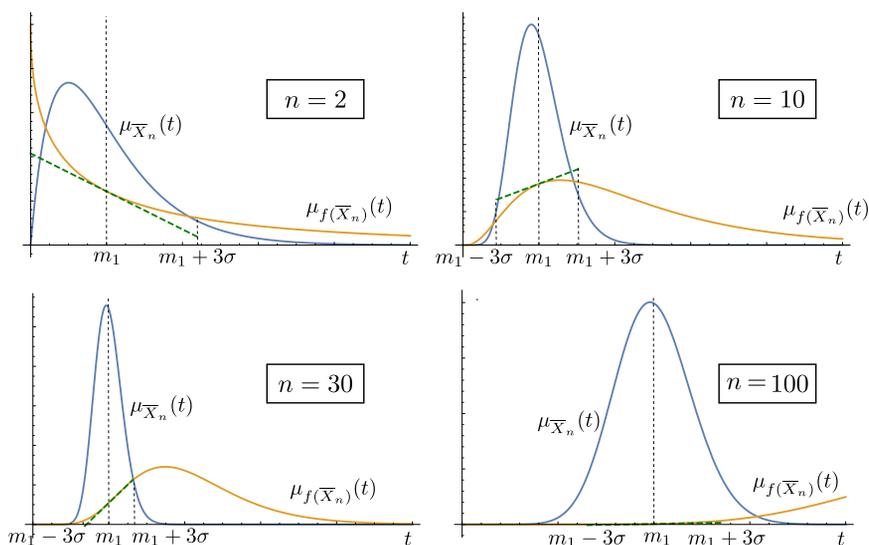


Figure 1.1: Statistical linearity for the first empirical moment of an exponential distribution. The plots show the distribution of the empirical mean (blue line) and the distribution of the square of the empirical mean function ( $f(x) = x^2$ ). The green dotted lines represent the linearisation of  $f(\bar{X}_N)$  at the theoretical mean value, displayed over an interval of  $\pm 3\sigma$ . We see that for increasing values of  $n$  the linearisation gets more and more accurate. Notice that the scale of the horizontal axis in the bottom right plot differs from the scale of the three others plots by a factor five.

### First moment estimator

We introduce a specific notation for the first empirical moment, also called empirical expectation value or empirical mean. For a random variable  $X$ , the empirical mean associated to an  $n$ -sample is given by

$$\bar{X}_n \equiv \tilde{m}_1 = \frac{1}{n} \sum_{i=1}^n X_i . \quad (1.17)$$

With this notation the first moment estimator is written as  $\hat{\theta}_{m_1} = m_1^{-1}(\bar{X}_n)$ . Is this estimator unbiased for a general parameter  $\theta$ ? To check it we have to calculate the expectation value  $E[\hat{\theta}_{m_1}]$ . To calculate it we use the transfer formula for the expectation value given in Appendix A (Eq. (A.15)), and we obtain

$$E[\hat{\theta}_{m_1}] = \int_{-\infty}^{+\infty} dt m_1^{-1}(t) \mu_{\bar{X}_n}(t) , \quad (1.18)$$

where  $\mu_{\bar{X}_n}$  is the probability distribution of the empirical mean. The problem is that this distribution is unknown. We know its asymptotic behaviour, given by the central limit theorem, but not its exact form. Therefore we cannot calculate the bias of the first moment estimator in general.

If we want to calculate the variance  $\text{Var}[\hat{\theta}_{m_1}]$  of  $\hat{\theta}_{m_1}$  the situation gets worse than for the calculation of the bias. We still do not know the distribution of  $\bar{X}_n$ , but we even do not have a transfer formula to calculate the variance of  $m_1^{-1}(\bar{X}_n)$ .

**Linearisation**

The statistical linearity of an estimator refers to its behaviour for large size of the  $n$ -sample (see Fig. 1.1). We will not make a strict mathematical proof, which would require theorems on convergence and quantification of the error, but just give the idea of the procedure.

The law of large numbers ensures that for a large  $n$ -sample, the empirical first moment converges to the expectation value of the original random variable. Moreover, the central limit theorem insures that asymptotically the empirical mean is normally distributed around  $m_1$ . The idea is to use this property to expand to first order (linearity) the estimator around the theoretical moment  $m_1(\theta)$ . Writing  $m_1^{-1} \equiv f$  for shortening the notations we have

$$f(\bar{X}_n) = f(m_1) + f'(m_1)(m_1 - \bar{X}_n) + \mathcal{O}((m_1 - \bar{X}_n)^2), \quad (1.19)$$

where  $f'$  is the derivative of the function  $f$ ,  $f'(t) = \frac{\partial f(t)}{\partial t}$ .

**Approximation of the mean** With this expansion we can calculate the approximate expectation value of the first moment estimator

$$\mathbb{E}[\hat{\theta}_{m_1}] = f(m_1) + f'(m_1) \mathbb{E}[m_1 - \bar{X}_n] + \mathcal{O}(\mathbb{E}[(m_1 - \bar{X}_n)^2]). \quad (1.20)$$

Using Eq. (1.13) and the properties of the expectation value we obtain

$$\mathbb{E}[\hat{\theta}_{m_1}] = f(m_1) + \mathcal{O}(\mathbb{E}[(m_1 - \bar{X}_n)^2]), \quad (1.21)$$

and since by definition  $f(m_1) = m_1^{-1}(m_1(\theta)) = \theta$  we are left with

$$\mathbb{E}[\hat{\theta}_{m_1}] = \theta + \mathcal{O}(\mathbb{E}[(m_1 - \bar{X}_n)^2]). \quad (1.22)$$

This means that up to the second order in  $(m_1 - \bar{X}_n)$  the first moment estimator is an unbiased estimator.

**Approximation of the variance** We now apply the statistical linearity for calculating the variance of the first moment estimator. Using Eq. (1.19) we obtain

$$\text{Var}[\hat{\theta}_{m_1}] = \text{Var}[f(m_1) + f'(m_1)(m_1 - \bar{X}_n) + \mathcal{O}((m_1 - \bar{X}_n)^2)]. \quad (1.23)$$

At this stage we are still facing the lack of simple formulas to express the variance of a function of  $X$ . Still we can make the approximation to neglect completely the terms of order two and more in the variance to obtain

$$\text{Var}[\hat{\theta}_{m_1}] \simeq \text{Var}[f(m_1) + f'(m_1)(m_1 - \bar{X}_n)]. \quad (1.24)$$

This last equation corresponds to an affine transformation of the empirical mean, and therefore we can use Eq. (A.17) to compute the variance

$$\text{Var}[\hat{\theta}_{m_1}] \simeq f'(m_1)^2 \text{Var}[\bar{X}_n]. \quad (1.25)$$

Using the fact that the variables  $X_i$  are independent, the variance of the empirical mean is  $\text{Var}[\bar{X}_n] = n(\text{Var}[X]/n^2)$  and we get

$$\text{Var}[\hat{\theta}_{m_1}] \simeq \frac{f'(m_1)^2}{n} \text{Var}[X]. \quad (1.26)$$

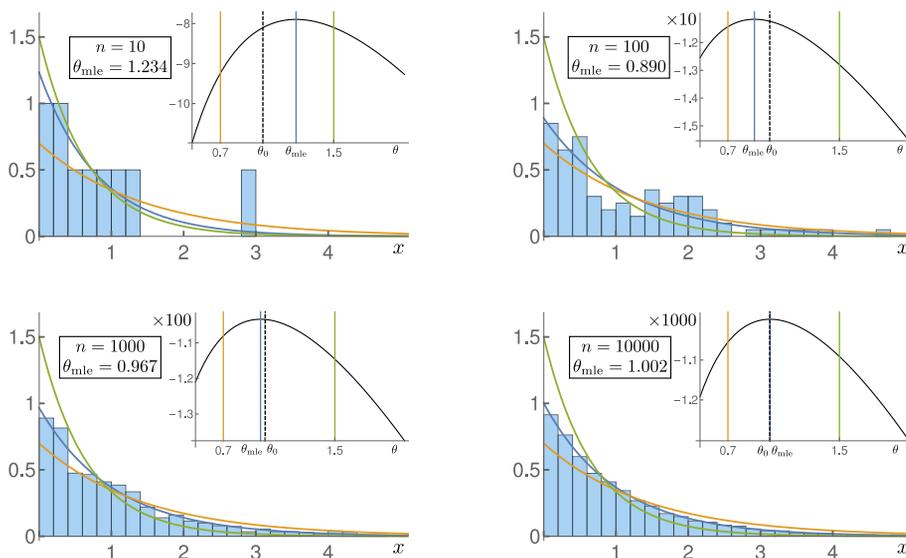


Figure 1.2: Maximum likelihood estimation for the mean value  $1/\theta$  of the exponential distribution for different sizes of the  $n$ -sample. The main plots represent the histogram of a realization of the  $n$ -sample, along with the probability distribution of the  $n$ -sample for different values of  $\theta$  (green:  $\theta = 1.5$  ; orange  $\theta = 0.7$  ; blue  $\theta = \hat{\theta}_{\text{mle}}$ , the maximum likelihood estimate). The insets represent the log-likelihood for the realization of the  $n$ -sample.

Now we can just use the standard formula for the derivative of the inverse of a function to calculate  $f'$

$$f'(m_1(\theta)) = (m_1^{-1})'(m_1(\theta)) = \frac{1}{m_1'(\theta)}. \quad (1.27)$$

Thus we are left with

$$\text{Var}[\hat{\theta}_{m_1}] \simeq \frac{\text{Var}[X]}{n \left( \frac{\partial m_1(\theta)}{\partial \theta} \right)^2}. \quad (1.28)$$

Under this approximation we see that the first two moments of  $\hat{\theta}_{m_1}$  depend only on the first two moments of  $X$ .

### 1.2.3 Maximum likelihood estimator

We now turn back to another very popular estimator, which will play an important role in the interpretation of the Cramér-Rao theorem: the *maximum likelihood estimator*.

#### Probably or likely?

First let us have a look on the concept of likelihood. As usual our starting point is a random process modelled by a random variable  $X \sim \mu_\theta$ . The likelihood  $l$  of  $X$  is defined as

$$l(\theta; x) := \mu_\theta(x), \quad (1.29)$$

*i.e.* the likelihood equals the probability distribution. So why do we introduce a new notation? The reason has to do with the *interpretation* of  $\theta$  and  $X$ . In estimation theory  $X$  represents the experiment, and is then a truly random quantity whose realizations fluctuate from experiment to experiment, while  $\theta$  is a fixed parameter that does not fluctuate. The notation  $\mu_\theta(X)$  is used to emphasize the fact that it is related to the random variable  $X$ . When reading  $\mu_\theta(X = x)$  we really think of the probability, the chance that we observe  $x$  as a result. The notation  $l(\theta; x)$  is there to emphasize the dependency on  $\theta$ . The interpretation being that we already observed  $x$ , which is therefore considered fixed, and ask ourselves how *likely* it was that the parameter had the value  $\theta$ . Formally we can write

$$\begin{aligned} \mu_\theta : x \in \mathcal{B}(\mathbf{R}) &\mapsto \mu_\theta(x) \in [0, 1] && \text{with } \theta \in \Theta \text{ a parameter} \\ l(\theta; x) : \theta \in \Theta &\mapsto l(\theta; x) \in [0, 1] && \text{with } x \in \mathcal{B}(\mathbf{R}) \text{ a parameter .} \end{aligned}$$

From a chronological point of view, the probability corresponds to a statement before the experiment, while the likelihood corresponds to a statement once that the experiment was realized and we observed the value  $x$ . For practical reasons, we also define the log-likelihood  $L$  as

$$L(\theta; x) := \ln(l(\theta; x)) . \tag{1.30}$$

which is the logarithm of the likelihood.

### Maximizing the likelihood

We start with the  $n$ -sample  $\{X_i\}$  and a specific realization of it noted  $\{x_i\}$ . Given this realization, we can ask ourselves *what is the most likely value of  $\theta$  such that the result is  $\{x_i\}$* ? The answer is the value of  $\theta$  which maximizes the likelihood. Since the logarithm is a monotonously increasing function, the maximum of the likelihood is also the maximum of the log-likelihood. Therefore and for practical reasons we define the *maximum likelihood estimator* (MLE)  $\hat{\theta}_{\text{mle}}$  of  $\theta$  as the value which maximizes the log-likelihood

$$\begin{aligned} \hat{\theta}_{\text{mle}} : \mathcal{B}(\mathbf{R}^n) &\rightarrow \Theta \\ x &\mapsto \operatorname{argmax}_{\theta \in \Theta} L(\theta; x) . \end{aligned}$$

Often we will use an  $n$ -sample  $X^{(n)}$  in the estimation procedure. The maximum likelihood estimator becomes  $\hat{\theta}_{\text{mle}} = \operatorname{argmax}_{\theta \in \Theta} L(\theta; \{x_i\})$ . Using Eq. (1.3) for the joint distribution of the  $n$ -sample we obtain for the likelihood  $l(\theta, \{x_i\}) = \prod_i l(\theta; x_i)$ . We see the advantage of introducing and using the log-likelihood, as the logarithm transforms the product into a sum to give

$$L(\theta, \{x_i\}) = \sum_i \ln(\mu_\theta(X_i = x_i)) . \tag{1.31}$$

### Invariance property of the MLE

It is interesting to note that the MLE obeys the so-called invariance property. This means that if we have calculated the MLE  $\hat{\theta}_{\text{mle}}$  for a parameter  $\theta$ , then the MLE for a parameter  $g = g(\theta)$ , where  $g$  is a bijective function, is given by

$$\hat{g}_{\text{mle}} = g(\hat{\theta}_{\text{mle}}) . \tag{1.32}$$

## 1.3 Cramér-Rao theorem

We have seen in Sec. 1.1 two important properties of estimators: their bias and their MSE which equals the variance for unbiased estimators. The quest for the best estimator leads to imposing a vanishing bias as a first condition, and, among the estimators that fulfil this criteria, to try to find the estimator that minimize the variance (MVU estimator). The problem is that MVU estimators do not always exist because usually not the same estimator minimizes the variance for different values of the parameter. Then an alternative solution would be to look for the estimator that locally minimizes the variance. The corresponding variance would be a good bound to check how good a given estimator is. The problem being that the calculation of this variance is in general a complicated task. What we will rather do is to derive a lower bound for the variance, and then to study how tight it is.

### 1.3.1 A lower bound for the variance

The famous Cramér-Rao theorem, named<sup>1</sup> after Harald Cramér and Calyampudi Radhakrishna Rao, provides a lower bound, the Cramér-Rao bound, on the variance of any unbiased estimator [Fréchet, 1943; Rao, 1945; Cramér, 1946]. Simple to prove mathematically, this theorem has been widely used in estimation theory and constitutes a central element in quantum metrology. It is also known as the "Information inequality".

**Theorem 1.1 — CRAMÉR-RAO BOUND (INFORMATION INEQUALITY).**

*Consider a random variable  $X$  distributed according to a family of probability distribution  $\{\mu_\theta(X)\}$  indexed by a parameter  $\theta \in \Theta$ . We assume the regularity condition to hold:*

$$\mathbb{E}\left[\frac{\partial \ln(\mu_\theta(x))}{\partial \theta}\right] = 0, \quad \forall \theta. \quad (1.33)$$

*We define an estimator  $\hat{\theta}_{\text{est}}$  of  $\theta$  as a function from the space of the outcome of  $X$  to  $\Theta$  and we assume that this estimator is unbiased:*

$$\mathbb{E}[\hat{\theta}_{\text{est}}] = \theta. \quad (1.34)$$

*Then the variance  $\text{Var}[\hat{\theta}_{\text{est}}]$  is bounded by the inverse of the so-called Fisher information  $J(\mu_\theta; \theta)$ :*

$$\text{Var}[\hat{\theta}_{\text{est}}] \geq \frac{1}{J(\mu_\theta; \theta)}, \quad (1.35)$$

*with*

$$J(\mu_\theta; \theta) := \mathbb{E}\left[\left(\frac{\partial \ln(\mu_\theta(x))}{\partial \theta}\right)^2\right]. \quad (1.36)$$

*An estimator saturating the Cramér-Rao bound is said to be efficient.*

### 1.3.2 Change of parameters

Suppose we are not interested anymore in the estimation of  $\theta$  but rather in the estimation of a function  $g(\theta)$  of  $\theta$ . What is then the Cramér-Rao bound for this new parameter? The following theorem answers the question:

---

<sup>1</sup>This theorem is also known as the Fréchet-Darmois-Cramér-Rao theorem, accounting for the contribution of Maurice Fréchet and Georges Darmois.

## Box 2: Vector parameter estimation

In vector parameter estimation, also known as multi-parameter estimation, one seeks to estimate a vector parameter  $\boldsymbol{\theta} := \{\theta_i\}$ . The formalism is the same as in scalar parameter estimation. The Cramér-Rao bound corresponds to an inequality for the covariance matrix:

$$\text{Cov}[\hat{\boldsymbol{\theta}}_{\text{est}}] - \mathbf{J}(\mu_{\boldsymbol{\theta}}; \boldsymbol{\theta})^{-1} \geq 0. \quad (1.37)$$

The elements of the Fisher information matrix  $\mathbf{J}(\mu_{\boldsymbol{\theta}}; \boldsymbol{\theta})$  are defined as

$$(\mathbf{J}(\mu_{\boldsymbol{\theta}}; \boldsymbol{\theta}))_{ij} := \mathbb{E}\left[\frac{\partial \ln(\mu_{\boldsymbol{\theta}})}{\partial \theta_i} \frac{\partial \ln(\mu_{\boldsymbol{\theta}})}{\partial \theta_j}\right]. \quad (1.38)$$

Under a change of parameter  $\mathbf{g}(\boldsymbol{\theta})$  the inverse of the new Fisher information matrix is given by

$$\mathbf{J}(\mu_{\boldsymbol{\theta}}; \mathbf{g})^{-1} = \mathbf{Jac}[\mathbf{g}, \{\theta_i\}] \cdot \mathbf{J}(\mu_{\boldsymbol{\theta}}; \boldsymbol{\theta})^{-1} \cdot \mathbf{Jac}[\mathbf{g}, \{\theta_i\}]^t, \quad (1.39)$$

where  $\mathbf{Jac}[\mathbf{g}, \{\theta_i\}]$  is the Jacobian of the transformation with elements  $[\mathbf{Jac}[\mathbf{g}, \{\theta_i\}]]_{ij} := \frac{\partial g_i(\boldsymbol{\theta})}{\partial \theta_j}$ .

The geometric interpretation is similar to the scalar case: The Fisher information matrix  $\mathbf{J}(\mu_{\boldsymbol{\theta}}; \mathbf{g})$  is a metric for the manifold of the probability distribution in the coordinate system  $\{\theta_i\}$ , and the change of coordinates is done through the Jacobian. Notice that we gave the inverse of the new Fisher information matrix rather than the original as after changing coordinates the matrix may not be invertible anymore (we could for example have less new variables than old ones).

When the FI matrix is diagonal the estimation of a parameter  $\theta_i$  is not affected by the lack of knowledge on the other parameters (the multi-parameter estimation is just equal to multiple scalar parameter estimation). Being a metric, we can always locally define a set of coordinates such that the FI matrix is diagonal. In general, for more than three parameters, it is not possible to find a global set of coordinate such that the FI matrix is diagonal.

**Theorem 1.2 — CHANGE OF PARAMETER.**

Let  $X$  be a random variable and  $\{\mu_{\boldsymbol{\theta}}(X)\}$  a parametric family with  $X \sim \mu_{\boldsymbol{\theta}}$ , and assume that the condition of Theorem 1.1 fulfilled. Let  $g(\boldsymbol{\theta})$  be a smooth function from  $\Theta$  to  $G$  and  $\hat{g}_{\text{est}}$  an unbiased estimator of  $g(\boldsymbol{\theta})$ . Then the variance of  $\hat{g}_{\text{est}}$  is lower bounded as

$$\text{Var}[\hat{g}_{\text{est}}] \geq \frac{\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2}{J(\mu_{\boldsymbol{\theta}}; \boldsymbol{\theta})}. \quad (1.40)$$

This theorem shows that when we re-parametrize the probability distribution the lower bound is rescaled by  $\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2$ . This result is to be put in relation with the usual formula of error propagation.

Equally we can use Theorem 1.2 to express the Fisher information for the new parameter  $g(\theta)$ :

$$J(\mu_\theta; g) = \frac{J(\mu_\theta; \theta)}{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}. \quad (1.41)$$

Especially the Fisher information for an affine transformation  $g(\theta) = a + b\theta$  of the parameter becomes  $J(\mu_\theta; g) = J(\mu_\theta; \theta)/b^2$ .

The change of parameters destroys the efficiency of an estimator. If  $\hat{\theta}_{\text{est}}$  is an efficient estimator of  $\theta$  then in general  $g(\hat{\theta}_{\text{est}})$  is not an efficient estimator of  $g(\theta)$ . In the asymptotic case the efficiency is typically conserved by any transformation (as we will see when studying the asymptotic properties of the MLE).

### 1.3.3 Proof of the Cramér-Rao theorem

We consider the general case where we want to estimate not directly the parameter  $\theta$  but a function of it,  $g(\theta)$ . This is done by an estimator  $\hat{g}_{\text{est}}$ . If  $\theta \in \Theta$  then we require that  $\hat{g}_{\text{est}}$  takes its values in  $g(\Theta)$ . The Cramér-Rao theorem is mainly based on the Cauchy-Schwarz inequality and indeed simple to prove.

Let us start by investigating the regularity condition (Eq. (1.33)) which is assumed to hold in the theorem. The left hand side of the equation reads

$$\text{E}\left[\frac{\partial \ln(\mu_\theta(x))}{\partial \theta}\right] = \int dx \frac{\partial \mu_\theta(x)}{\partial \theta}. \quad (1.42)$$

If we can interchange the derivative and the integral we obtain

$$\text{E}\left[\frac{\partial \ln(\mu_\theta(x))}{\partial \theta}\right] = \frac{\partial}{\partial \theta} \int dx \mu_\theta(x). \quad (1.43)$$

Since  $\int dx \mu_\theta(x) = 1$  by definition, and since the derivative of 1 is zero, we retrieve the regularity condition (1.33)

$$\text{E}\left[\frac{\partial \ln(\mu_\theta(x))}{\partial \theta}\right] = 0. \quad (1.44)$$

Therefore we see that a sufficient condition for the regularity condition to hold is that we can interchange the derivative and the limits<sup>2</sup>. This turns out to be true for most of the common probability distributions. An obvious case where this fails to be true is when the parameter appears in the limit of the integral (as it is the case for the estimation of the upper bound of a uniform distribution).

We now look at the second requirement appearing in the theorem: the unbiasedness of the estimator. By differentiating with respect to  $\theta$  both sides of the unbiasedness condition  $\int dx \mu_\theta(x) \hat{g}_{\text{est}}(x) = g(\theta)$  and interchanging the derivative and integral we obtain

$$\frac{\partial g(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \int dx \mu_\theta(x) \hat{g}_{\text{est}} = \int dx \mu_\theta(x) \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \hat{g}_{\text{est}}. \quad (1.45)$$

We used the fact that the estimator *does not depend on the value of  $\theta$* . The fact that its average value *does depend* on  $\theta$  should not mislead the reader. This dependency comes from the probability

---

<sup>2</sup>We could have just proposed this condition as regularity condition. Nevertheless the stronger version presented here has the advantage to give directly the possibility to re-write the Fisher information as in Eq. (1.64).

distribution  $\mu_\theta$  and not from the estimator  $\hat{\theta}_{\text{est}}$ . From an operational point of view this translates the fact that the method used to estimate the value of  $\theta$  should not depend on  $\theta$ . Multiplying both sides of the regularity condition (1.33) by  $g(\theta)$  we obtain

$$\int dx \mu_\theta(x) \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} g(\theta) = 0, \quad (1.46)$$

where we used the fact that  $g(\theta)$  is independent of  $x$ . By subtracting Eq. (1.46) from Eq. (1.45) we finally obtain

$$\frac{\partial g(\theta)}{\partial \theta} = \int dx \mu_\theta(x) \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} (\hat{g}_{\text{est}} - g(\theta)). \quad (1.47)$$

At this stage most of the work is already done. The next and last step is to apply the Cauchy–Schwarz inequality. We first define a scalar product  $\langle \cdot, \cdot \rangle_{\mu_\theta}$  in the space of real integrable functions. For two real functions  $a(x)$  and  $b(x)$  and for any probability distribution  $\mu_\theta(x)$  we define the scalar product between  $a$  and  $b$  as

$$\langle a(x), b(x) \rangle_{\mu_\theta} := \int dx a(x) b(x) \mu_\theta(x). \quad (1.48)$$

Using this scalar product, we can write Eq. (1.47) as

$$\frac{\partial g(\theta)}{\partial \theta} = \left\langle \frac{\partial \ln(\mu_\theta(x))}{\partial \theta}, \hat{g}_{\text{est}} - g(\theta) \right\rangle_{\mu_\theta}. \quad (1.49)$$

Applying the Cauchy–Schwarz inequality to the right hand side we obtain

$$\left\langle \frac{\partial \ln(\mu_\theta(x))}{\partial \theta}, \hat{g}_{\text{est}} - g(\theta) \right\rangle_{\mu_\theta}^2 \leq \left\langle \frac{\partial \ln(\mu_\theta(x))}{\partial \theta}, \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right\rangle_{\mu_\theta} \langle \hat{g}_{\text{est}} - g(\theta), \hat{g}_{\text{est}} - g(\theta) \rangle_{\mu_\theta}. \quad (1.50)$$

Finally we are left with

$$\int dx \mu_\theta(x) (\hat{g}_{\text{est}} - g(\theta))^2 \geq \frac{\left( \frac{\partial g(\theta)}{\partial \theta} \right)^2}{\int dx \mu_\theta(x) \left( \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right)^2}. \quad (1.51)$$

By definition  $\int dx \mu_\theta(x) (\hat{g}_{\text{est}} - g(\theta))^2 = \text{Var}[\hat{g}_{\text{est}}]$ . Injecting this in the last equation we obtain the Cramér–Rao bound

$$\text{Var}[\hat{g}_{\text{est}}] \geq \frac{\left( \frac{\partial g(\theta)}{\partial \theta} \right)^2}{J(\mu_\theta; \theta)}. \quad (1.52)$$

The original Cramér–Rao bound (1.35) is then obtained by taking  $g(\theta) = \theta$ .

In this derivation the Fisher information appears as the norm of the derivative of the log-likelihood

$$J(\mu_\theta; \theta) = \left\langle \frac{\partial \ln(\mu_\theta(x))}{\partial \theta}, \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right\rangle_{\mu_\theta}. \quad (1.53)$$

The derivative of the log-likelihood is sometimes called the *score function*. When the log-likelihood varies a lot —meaning that its derivative, the score function, is high— we are sensitive to small variations of  $\theta$  and thus the FI, which is the norm of the score, is large.

### 1.3.4 Cramér-Rao theorem and biased estimators

The Cramér-Rao theorem as stated in Theorem 1.1 applies only to unbiased estimators. If we relax this assumption we can still calculate a bound for the variance of any estimator, using the result on the change of parameters.

Suppose we want to estimate  $\theta$  but that the estimator  $\hat{\theta}_{\text{est}}$  is not unbiased:  $E[\hat{\theta}_{\text{est}}] = g(\theta)$ . The bias  $b(\theta)$  is thus equal to  $b(\theta) = g(\theta) - \theta$ . We can actually still use the Cramér-Rao theorem since  $\hat{\theta}_{\text{est}}$  is an unbiased estimator of  $g(\theta)$ . Using the fact that  $g(\theta) = \theta + b(\theta)$  along with Eq. (1.40) we obtain:

$$\text{Var}[\hat{\theta}_{\text{est}}] \geq \frac{\left(1 + \frac{\partial b(\theta)}{\partial \theta}\right)^2}{J(\mu_\theta; \theta)} = \frac{\left(\frac{\partial E[\hat{\theta}_{\text{est}}]}{\partial \theta}\right)^2}{J(\mu_\theta; \theta)}. \quad (1.54)$$

The estimator  $\hat{\theta}_{\text{est}}$  being biased its variance is not equal to its MSE, and it is this latter that is a proper risk function and not the variance. Indeed for a biased estimator it does not make sense to seek for a minimal variance. We do not want to be closer to the biased value. For example a constant estimator has a zero variance but is pretty poor. Using Eq. (1.9) we get a bound on the MSE

$$\text{MSE}[\hat{\theta}_{\text{est}}] \geq \frac{\left(1 + \frac{\partial b(\theta)}{\partial \theta}\right)^2}{J(\mu_\theta; \theta)} + b(\theta)^2. \quad (1.55)$$

This bound can be lowered by certain biased estimators. It would be enough to have a bias with a slope close to minus one to lower drastically the MSE. If in addition the value of the bias vanishes at this point we can have an MSE arbitrarily close to zero. Asking for the bias to vanish at a certain point could appear to be equivalent to ask the estimator to be unbiased. This simple analysis shows — and we will explicitly work out this point later on — that not only the value of the bias matter but also its first derivative.

The problem with biased estimators is that it is difficult to assess their performance. Since the lower bound depends on the bias there is a different bound for each estimator. Furthermore the bias is usually difficult to compute. Altogether we see why we focused on unbiased estimators: it allows to have a bound which depends only on the distribution.

## 1.4 Fisher Information

An important quantity that appears in the Cramér-Rao theorem is the Fisher information  $J$ . The reciprocal of  $J$  determines the lower bound for the variance of unbiased estimators. The larger  $J$  is the lower is the bound, which means that we can extract more information from the probability distribution, giving to  $J$  this feature of measuring the amount of information about  $\theta$  contained in  $\mu_\theta$ . From this point of view  $J$  can also be seen as a measure of the sensitivity of  $\mu_\theta$  with respect to  $\theta$ . A geometric interpretation of parameter estimation will confirm this point of view.

### 1.4.1 Additivity of the Fisher information for $n$ -samples

We formulated the Cramér-Rao theorem for a general random variable  $X \sim \mu_\theta$ . But in practice we often use repetitions of the random process to estimate  $\theta$ , a situation described by an  $n$ -sample  $X^{(n)}$ . We now look at the Cramér-Rao theorem for this specific case.

Using Eq. (1.3) for the distribution of the  $n$ -sample we obtain

$$\ln(\mu_\theta^{(n)}(\{x_i\})) = \sum_{i=1}^n \ln(\mu_\theta(x_i)). \quad (1.56)$$

The Fisher information for  $\mu_\theta^{(n)}$  is thus equal to

$$J(\mu_\theta^{(n)}; \theta) = \int dx_1 \mu_\theta(x_1) \cdots \int dx_n \mu_\theta(x_n) \left( \sum_{i,j} \frac{\partial \ln(\mu_\theta(x_i))}{\partial \theta} \frac{\partial \ln(\mu_\theta(x_j))}{\partial \theta} \right). \quad (1.57)$$

Using the normalization of the individual distribution  $\int dx_i \mu_\theta(x_i) = 1$  and regrouping the terms properly we obtain

$$J(\mu_\theta^{(n)}; \theta) = n \int dx \mu_\theta(x) \left( \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right)^2 + n(n-1) \left( \int dx \mu_\theta(x) \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right)^2. \quad (1.58)$$

The regularity condition (1.33) for  $\mu_\theta^{(n)}$  reads  $E[\ln(\mu_\theta^{(n)})] = 0$  and leads to

$$n \int dx \mu_\theta(x) \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} = 0. \quad (1.59)$$

Using this result the Fisher information for an  $n$ -sample is given by

$$J(\mu_\theta^{(n)}; \theta) = n J(\mu_\theta; \theta). \quad (1.60)$$

This property is known as the additivity of the Fisher information. When translated to the Cramér-Rao theorem we obtain the following lemma:

**Lemma 1.1 — CRAMÉR-RAO THEOREM FOR  $n$ -SAMPLES (ADDITIVITY).**

*When using an  $n$ -sample of the random variable  $X$  distributed according to  $\mu_\theta$ , the variance of any unbiased estimator of  $\theta$  is bounded as*

$$\text{Var}[\hat{\theta}_{\text{est}}] \geq \frac{1}{n J(\mu_\theta; \theta)}. \quad (1.61)$$

This lemma has important consequences in quantum metrology in the sense that it defines the standard scaling obtained when repeating an experiment. A variance divided by the number  $n$  of random variables is reminiscent from the central limit theorem, where the variance of the sum of the random variable also decreases with  $1/n$ . We will see that this intuition is actually true when studying the asymptotic properties of the theorem.

## 1.4.2 Re-writing the Fisher information

We can use the regularity condition (1.33) to express the Fisher information in a different form. By differentiating the regularity condition and interchanging the order of integral and derivative we obtain

$$\int dx \left( \frac{\partial \mu_\theta(x)}{\partial \theta} \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} + \frac{\partial^2 \ln(\mu_\theta(x))}{\partial \theta^2} \right) = 0, \quad (1.62)$$

which can be re-written as

$$\int dx \mu_\theta(x) \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} = - \int dx \mu_\theta(x) \frac{\partial^2 \ln(\mu_\theta(x))}{\partial \theta^2} . \quad (1.63)$$

We recognize the definition of the Fisher information on the left hand side and we thus found a different way to express it

$$J(\mu_\theta; \theta) = - \int dx \mu_\theta(x) \frac{\partial^2 \ln(\mu_\theta(x))}{\partial \theta^2} . \quad (1.64)$$

### 1.4.3 Geometric interpretation of the Fisher Information

A common task in information theory is to check how different two sets of data are. An important measure of distance between two data sets of size  $n$   $\{p_i\}$  and  $\{q_i\}$  is the Bhattacharyya coefficient  $c_B(\{p_i\}, \{q_i\}) := \sum_{i=1}^n \sqrt{p_i q_i}$ . Notice that this coefficient is not a true distance. Indeed we will use the term distance in a pretty loose way throughout this section. Data being generated by random variables, it is also natural to try to define distances between two probability distributions  $p(X)$  and  $q(X)$ . We can extend the Bhattacharyya coefficient to two probability distributions  $p$  and  $q$  as  $c_B(p, q) := \int dx \sqrt{p(x)q(x)}$ .

Using the Bhattacharyya coefficient we can define two distances between  $p$  and  $q$ : The Bhattacharyya distance and the Hellinger distance. The Bhattacharyya distance is defined as  $d_B(p, q) := -\ln(c_B(p, q))$  and the Hellinger distance as  $d_H(p, q) := 1 - c_B(p, q)$ . Strictly speaking, the Bhattacharyya distance is only a semi-distance as it does not fulfil the triangle inequality.

The Hellinger distance is directly connected to the Fisher information. Consider a distribution  $\mu_\theta$ . The infinitesimal distance between  $\mu_\theta$  and  $\mu_{\theta+d\theta}$  is related to the Fisher information through

$$4 \lim_{d\theta \rightarrow 0} d_H(\mu_\theta, \mu_{\theta+d\theta}) = J(\mu_\theta; \theta) . \quad (1.65)$$

This result gives a nice interpretation to the Fisher information: The Fisher information measures how sensitive the probability distribution is to changes of the value of the parameter.

Not only the Hellinger distance is linked to the Fisher information. The Kullback-Leibler divergence, also known as relative entropy in information theory, is defined as  $d_{KL}(p|q) := \int dx p(x) \frac{p(x)}{q(x)}$ . Notice that this is not a distance as it is not symmetric. Again we can express the Fisher information as a limit

$$4 \lim_{d\theta \rightarrow 0} d_{KL}(\mu_\theta | \mu_{\theta+d\theta}) = J(\mu_\theta; \theta) . \quad (1.66)$$

The study of the relation between divergence (like the Kullback-Leibler one) and Fisher information is properly treated by the geometry of information developed by Amari. In this framework we use the parameter dependence of the probability distribution to construct manifolds and then use the tools of differential geometry. It was shown in [Amari and Nagaoka, 2007] that there is a class of divergences, the  $\alpha$ -divergences, that are related to the Fisher information. Indeed the Fisher information corresponds to the Hessian of these divergences and constitutes a metric<sup>3</sup> in the manifold. This approach is especially fruitful in the case of vector parameters.

---

<sup>3</sup>By metric we mean here only a quadratic form. This is sometimes referred as a metric tensor and some authors keeps the word metric for distance functions.

## 1.5 Saturation of the Cramér-Rao bound in the finite and asymptotic cases

A crucial question in relation with the Cramér-Rao is the possibility of saturating the bound. It is informative on its own to know that we cannot do better than a certain threshold, but if the bound is not tight at all its interest decreases. We will investigate the conditions under which one can find an estimator that saturates the Cramér-Rao bound. To do so we will analyse separately the case of finite  $n$ -sample and the asymptotic case.

### 1.5.1 Saturating the bound

The proof of the Cramér-Rao theorem is based on the Cauchy-Schwarz inequality for a suitable scalar product. Looking for the conditions for saturating the bound reduces to looking for the conditions of equality in the Cauchy-Schwarz inequality.

For arbitrary functions  $a(x)$  and  $b(x)$ , the Cauchy-Schwarz inequality reaches the equality for  $a(x) = \alpha b(x)$ . If we apply this to the inequality (1.50) we obtain the condition

$$\frac{\partial \ln(\mu_\theta(x))}{\partial \theta} = \alpha(\hat{g}_{\text{est}} - g(\theta)). \quad (1.67)$$

The factor of proportionality  $\alpha$  should not depend on  $x$  but can depend on  $\theta$ :  $\alpha = \alpha(\theta)$ .

In order to determine  $\alpha(\theta)$  we start by differentiating Eq. (1.67) with respect to  $\theta$

$$\frac{\partial^2 \ln(\mu_\theta(x))}{\partial \theta^2} = \frac{\partial \alpha(\theta)}{\partial \theta} (\hat{g}_{\text{est}} - g(\theta)) - \frac{\partial g(\theta)}{\partial \theta} \alpha(\theta). \quad (1.68)$$

To get rid of the dependency on  $x$  we now multiply by the probability distribution  $\mu_\theta$  and then integrate over  $x$  to obtain

$$\int dx \mu_\theta(x) \frac{\partial^2 \ln(\mu_\theta(x))}{\partial \theta^2} = \frac{\partial \alpha(\theta)}{\partial \theta} \int dx \mu_\theta(x) (\hat{g}_{\text{est}} - g(\theta)) - \alpha(\theta) \int dx \mu_\theta(x) \frac{\partial g(\theta)}{\partial \theta}. \quad (1.69)$$

We see that the left hand side is equal to the Fisher information  $J(\mu_\theta; \theta)$ . By using the fact that  $\hat{g}_{\text{est}}$  is unbiased and that  $\frac{\partial g(\theta)}{\partial \theta}$  is independent of  $x$  we can also rewrite the right hand side to finally obtain

$$J(\mu_\theta; \theta) = \alpha(\theta) \frac{\partial g(\theta)}{\partial \theta}. \quad (1.70)$$

By injecting this equation in Eq. (1.67) we can state the following lemma

**Lemma 1.2 — SATURATION OF CRAMÉR-RAO BOUND FOR FINITE  $n$ -SAMPLES.**

*We consider a family of probability distributions  $\{\mu_\theta(x)\}$  indexed by a parameter  $\theta \in \Theta$  for which the Cramér-Rao theorem (theorem 1.1) holds. An efficient estimator of the transformed parameter  $g(\theta)$  exists if and only if there exists a function  $f$  satisfying*

$$\frac{\partial \ln(\mu_\theta(x))}{\partial \theta} = \frac{J(\mu_\theta; \theta)}{\partial g(\theta)/\partial \theta} (f(x) - g(\theta)). \quad (1.71)$$

*The efficient estimator  $\hat{\theta}_{\text{eff}}$  saturating the bound is then  $\hat{\theta}_{\text{eff}} = f(x)$ .*

### 1.5.2 Exponential families

In statistics many distributions are used and studied. A particular important class among them is the exponential family. The exponential family comprises many well-known distributions: Poisson, Gamma, Bernoulli, Normal, Pareto, Chi-squared, etc. In full generality, distributions from the exponential family can be written as

$$\mu_{\theta}^{\text{exp}}(x) = \zeta(x) e^{\gamma(\theta)T(x) - \beta(\theta)}. \quad (1.72)$$

The normalization condition reads  $e^{-\beta(\theta)} \int dx \zeta(x) e^{\gamma(\theta)T(x)} = 1$ , which shows that  $\beta(\theta)$  plays the role of a normalizing factor. Eventually the effect of  $\theta$  on the probability distribution is given by the function  $\gamma(\theta)$ .

The likelihood for an  $n$ -sample is given by

$$l(\theta, \{x_i\}) = e^{-n\beta(\theta)} \left( \prod_{i=1}^n \zeta(x_i) \right) e^{\gamma(\theta) \sum_i T(x_i)}, \quad (1.73)$$

and the log-likelihood by

$$L(\theta, \{x_i\}) = -n\beta(\theta) + \sum_{i=1}^n \ln(\zeta(x_i)) + \gamma(\theta) \sum_{i=1}^n T(x_i). \quad (1.74)$$

The maximum likelihood estimator is found by setting the derivative of the log-likelihood to zero, *i.e.* it is the solution of

$$n \frac{\partial \beta(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_{\text{MLE}}} = \frac{\partial \gamma(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_{\text{MLE}}} \sum_{i=1}^n T(x_i). \quad (1.75)$$

The specific case where  $\gamma(\theta) = \theta$  is known as the natural case. Using the generator of moments, one can show that in the natural case the mean and variance of  $T(X)$  are given by  $E[T(X)] = n \frac{\partial \beta(\theta)}{\partial \theta}$  and  $\text{Var}[T(X)] = n \frac{\partial^2 \beta(\theta)}{\partial \theta^2}$ . This shows that for natural parameters the maximum likelihood estimator is the solution of

$$E[T(X)] = \sum_{i=1}^n T(x_i) = \overline{T(x_i)}, \quad (1.76)$$

which is nothing else than the equation for the first moment estimator.

More importantly it is possible to show that under certain reasonable conditions not detailed here, only members of the exponential family can have efficient estimators (although not all of them). This is a capital result as it shows that in general we should not expect to find an estimator that saturates the Cramér-Rao bound.

### 1.5.3 Maximum likelihood estimator and asymptotic study

We presented in Sec. 1.2.3 the maximum likelihood estimator (MLE) but did not investigate its statistical properties. In the same way as for the moment estimator we will work especially in the asymptotic case.

**Theorem 1.3 — STATISTICAL BEHAVIOUR OF THE MLE.**

*We consider a random variable  $X$  with distribution  $\mu_{\theta}$  along with the  $n$ -sample  $X^{(n)} = \{X_i\}$  with the distribution  $\mu_{\theta}^{(n)}$ . We assume that the first and second derivatives of  $L(\theta; X)$  exist and that the regularity condition (1.33) holds. Then, in the limit of large sample sizes, the maximum likelihood*

estimator  $\hat{\theta}_{\text{mle}}$  is normally distributed (see Sec. A.5.2) with average  $\theta$  and variance equal to the Fisher information:

$$\hat{\theta}_{\text{mle}} \sim \mathcal{N}(\theta, J(\mu_\theta; \theta)^{-1/2}). \quad (1.77)$$

This theorem implies the important following corollary:

**Corollary 1.1 — ASYMPTOTIC EFFICIENCY OF THE MLE.**

The maximum likelihood estimator is asymptotically unbiased,

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}_{\text{mle}}] = \theta, \quad (1.78)$$

and is also asymptotically efficient,

$$\lim_{n \rightarrow \infty} \left( \text{Var}[\hat{\theta}_{\text{mle}}(\{X_i\}_{i=1, \dots, n})] - \frac{1}{n J(\mu_\theta; \theta)} \right) = 0. \quad (1.79)$$

This corollary is central for the justification of the use of the Cramér-Rao theorem: *we always know that for a large enough  $n$ -sample, the bound is reachable.* But as it is often the case with asymptotic laws in probability, we do not know how big the sample should be to be sure that we saturate or almost saturate the bound. We will nevertheless consider that for most of the cases in which we may be interested, this threshold stays reasonable and thus justifies the use of the Cramér-Rao bound and the MLE.

## 1.6 Locality

In the theory of estimation presented up to here we did not put the accent on the notion of globality/locality. Still we have seen how they come into play. For example the discussion on the MSE and bias showed that when going to local properties one needs to take into account the property of the bias in the vicinity of the point we are interested in. Locality also appears to be an important question for optimality. Importantly we have seen that there does not always exist an MVU estimator for all values of the parameter, and that instead one has to look for local MVU estimators. Actually the theory of parameter estimation that we use through this thesis is a local theory.

In this context we are not anymore interested in the estimation of  $\theta$  for all values of  $\theta$  but rather for estimating  $\theta$  at a specific point  $\theta_0$ . This is somehow an arbitrary choice, but made possible by the natural local character of the Fisher information and of the Cramér-Rao theorem.

### 1.6.1 Locality of the Fisher information

The Fisher information for a probability distribution  $\mu_\theta$  has been defined as

$$J(\mu_\theta; \theta) = \int dx \mu_\theta(x) \left( \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right)^2. \quad (1.80)$$

In general this is a function of  $\theta$ . The dependency on  $\theta$  appears through  $\mu_\theta(x)$  and its first derivative  $\frac{\partial}{\partial \theta} \mu_\theta(x)$ . This shows that the Fisher information is, in the space of parameters, a local quantity. As a consequence, any probability distribution taking the same value at a point  $\theta_0$  and having the same derivative at this point will produce the same Fisher information, independently of the global form of the distribution.

The locality of the Fisher information arises naturally when adopting a geometric picture of the estimation problem. Since the Fisher information represents a metric on the manifold produced by the parametric family it is clear that in general the Fisher information will depend on the value of the parameter, *i.e.* on the point on the manifold.

### 1.6.2 Locality in the Cramér-Rao theorem

A direct consequence of the locality of the Fisher information is the local character of the Cramér-Rao bound: the lower bound depends on the value of the parameter. This suggests that we can also weaken our assumptions for the Cramér-Rao theorem to make it a theorem with a local character.

#### Local unbiased condition

In order for the bound to hold we need two conditions: The regularity condition and the unbiasedness condition. The unbiasedness condition reads

$$\int dx \mu_\theta(x) \hat{g}_{\text{est}} = g(\theta) , \quad (1.81)$$

and *a priori* this should be true for all values of  $\theta$ . Nevertheless since the bound is local it should be enough to have only a local unbiasedness condition. Naively we could think that the local version of this requirement is only the condition  $E[\hat{g}_{\text{est}}]_{\theta_0} = g(\theta_0)$ . A first hint that this is not true was given by our discussion on the MSE, where we saw that the first derivative of the bias matters also. A close look to the proof of the theorem (see Eq. (1.45)) shows that we also need

$$\frac{\partial}{\partial \theta} \int dx \mu_\theta(x) \hat{g}_{\text{est}} = \frac{\partial g(\theta)}{\partial \theta} . \quad (1.82)$$

to hold at  $\theta_0$ , giving us a second requirement for local unbiasedness. Altogether the unbiasedness condition in the local framework reads

$$E[\hat{g}_{\text{est}}]_{\theta_0} = g(\theta_0) , \quad (1.83a)$$

$$\left. \frac{\partial}{\partial \theta} E[\hat{g}_{\text{est}}] \right|_{\theta_0} = \left. \frac{\partial g(\theta)}{\partial \theta} \right|_{\theta_0} . \quad (1.83b)$$

Obviously if the estimator is *globally* unbiased, it will also be *locally* unbiased everywhere. From these two equations, it turns out that it is the second one that is really restrictive and important. The first one can always be made true by adding offsets to the estimator. Notice also that it is the second condition that prevents constant estimator to be efficient local estimators. The local unbiasedness condition can be re-written using the bias

$$b(\theta_0) = 0 , \quad (1.84)$$

$$\left. \frac{\partial b(\theta)}{\partial \theta} \right|_{\theta_0} = 0 . \quad (1.85)$$

The local regularity condition reads

$$E\left[ \left. \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right|_{\theta_0} \right] = 0 . \quad (1.86)$$

### Locality and bias

We now turn our attention to the Cramér-Rao theorem for possibly biased estimators. The local unbiasedness condition reads

$$\mathbb{E}[\hat{\theta}_{\text{est}}] \Big|_{\theta_0} = \theta_0 + b(\theta_0), \quad (1.87a)$$

$$\frac{\partial}{\partial \theta} \mathbb{E}[\hat{\theta}_{\text{est}}] \Big|_{\theta_0} = 1 + \frac{\partial b(\theta)}{\partial \theta} \Big|_{\theta_0}. \quad (1.87b)$$

This emphasizes that when using a biased estimator, what matters is the value of the bias at the point considered and the slope of the bias at this point.

#### Box 3: Local estimation theory in a nutshell

In local estimation theory we are looking for the best estimators of  $\theta$  at the specific value  $\theta_0$ . The first criteria is that the estimator  $\hat{\theta}_{\text{est}}$  should be locally unbiased,

$$\begin{aligned} \mathbb{E}[\hat{g}_{\text{est}}] \Big|_{\theta_0} &= g(\theta_0), \\ \frac{\partial}{\partial \theta} \mathbb{E}[\hat{g}_{\text{est}}] \Big|_{\theta_0} &= \frac{\partial g(\theta)}{\partial \theta} \Big|_{\theta_0}. \end{aligned}$$

We then look for the estimator which minimizes the variance, called the *local minimum variance unbiased estimator*. In this context the Cramér-Rao theorem provides a bound for the variance of unbiased estimators at  $\theta_0$  using a  $n$ -sample,

$$\text{Var}[\hat{\theta}_{\text{est}}] \Big|_{\theta_0} \geq \frac{1}{n J(\mu_\theta; \theta) \Big|_{\theta_0}}. \quad (1.88)$$

By using the maximum likelihood estimator we know that this bound is asymptotically reachable when the size  $n$  of the sample increases.

### 1.6.3 Operational justification of local estimation

We have now shown the local character of the Cramér-Rao theorem. What is the meaning of this locality in terms of estimation theory?

When we defined the concept of MVU estimators, the bias or the variance we were implicitly and naturally asking for properties that hold for every  $\theta$ . Naturally because the idea of estimation theory is that we do not know what is the value of  $\theta$ . We just know that  $\theta$  belongs to the set  $\Theta$  and were thus asking for *global* properties.

We can also propose a different approach to parameter estimation theory, namely the *local* estimation. In this case, we relax the strong requirement that the statistical properties of our estimators should hold for every  $\theta$ , and just ask them to hold at the point  $\theta_0$  and in its neighbourhood. While we show that in the global case only the distribution belonging to the exponential family allows for an MVU estimator, in the local case we can always construct a local MVU estimator:

Consider the estimator

$$\hat{\theta}_{\text{est}}(\{x_i\}) := \theta_0 + \frac{1}{N J(\mu_\theta; \theta)|_{\theta_0}} \left. \frac{\partial L(\theta, \{x_i\})}{\partial \theta} \right|_{\theta_0}. \quad (1.89)$$

By construction this estimator saturates the Cramér-Rao bound at  $\theta_0$ . But this estimator can appear as artificial as it requires the knowledge of  $\theta_0$  which we want to estimate.

Actually this is not a problem: First, we should realize that the local approach corresponds to an ideal situation, in the sense that we allow for an estimator specially designed for the value of the parameter we are interested in. Being an ideal situation, any realistic situation can only lead to a worse estimation, and therefore the local approach serves to establish a bound on any other approach. Second we made explicit that the basis of our analysis of estimation theory is done through the Cramér-Rao bound, *and the fact that this bound can be saturated asymptotically*. The MLE estimator being globally efficient in the asymptotic limit (it produces the MVU estimator in this limit) we know that we will be able to saturate the bound.

## Summary Chapter 1

- **Estimation problem:** We consider a  $\theta$ -dependent probability distribution  $\mu_\theta(X)$  from which we draw a  $n$ -sample. Using this  $n$ -sample we want to estimate the value of  $\theta$ .
- **Estimator:** Function  $\hat{\theta}_{\text{est}}$  from the  $n$ -sample to the parameter space.  $\hat{\theta}_{\text{est}}$  is locally unbiased at the point  $\theta_0$  when  $E[\hat{\theta}_{\text{est}}]_{\theta=\theta_0} = \theta_0$  and  $\frac{\partial E[\hat{\theta}_{\text{est}}]}{\partial \theta}|_{\theta=\theta_0} = 1$ .
- **Cramér-Rao Theorem:** Gives a bound on how precisely we can estimate the parameter  $\theta$ : the variance of any locally unbiased estimator is lower bounded by the inverse of the Fisher information:  $\text{Var}[\hat{\theta}_{\text{est}}] \geq (n J(\mu_\theta; \theta))^{-1}$ .
- **Fisher information:** The Fisher information is given by  $J(\mu_\theta; \theta) = E\left[\left(\frac{\partial \ln(\mu_\theta(X))}{\partial \theta}\right)^2\right]$ . It characterizes how much information on  $\theta$  is contained in  $\mu_\theta$ .
- **Saturation:** The Cramér-Rao bound is always saturated asymptotically ( $n \rightarrow \infty$ ) by the maximum likelihood estimator.
- **Maximum likelihood estimator:** The idea is to find the value of  $\theta$  that is the most likely for a specific realization of the  $n$ -sample. Formally  $\hat{\theta}_{\text{mle}} = \text{argmax}_\theta \ln(\mu_\theta(X))$ .

# Chapter 2

## Foundation of quantum theory

This chapter is devoted to the basics of quantum mechanics. With one century of history, quantum mechanics gave rise to many different sub-fields, each of them coming with its own vocabulary and formalism. We will review the fundamental tools of quantum mechanics from a point of view of quantum information. We do not seek to be exhaustive, rather we focus on what is needed to formalize and understand quantum parameter estimation theory and quantum metrology. To not make the presentation too cluttered, and because it suffices to our needs, we present only the formalism for finite dimensional systems. As it was the case in the first chapter, most of the results presented here are standard and can be found in textbooks. Over years the reference for an introduction to quantum computation and quantum information has become [Nielsen and Chuang, 2011]. A somehow more mathematical treatments of questions relative to quantum information (with a strong emphasis on geometry) can be found in [Bengtsson and Zyczkowski, 2006] .

### 2.1 States

#### **Postulate 1 — STATE OF A SYSTEM.**

*To any physical system we associate a Hilbert space  $\mathcal{H}$ . We define the space  $\mathcal{B}(\mathcal{H})$  of bounded linear operators acting on  $\mathcal{H}$  and the space  $\mathcal{S}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$  of positive bounded linear operators with trace one<sup>1</sup>. The state of the physical system is represented by a vector  $\rho \in \mathcal{S}(\mathcal{H})$ , and we refer to it as a density operator or density matrix.*

The elements of  $\mathcal{H}$  are referred to as ket vectors ( $|\rangle$ ) and the elements of its dual,  $\mathcal{H}^*$  are referred as bra vectors ( $\langle|$ ). Using this notation, the positivity<sup>2</sup> of an operator  $A$  is defined as:  $\langle\psi|A|\psi\rangle \geq 0$  for all  $|\psi\rangle \in \mathcal{H}$ .

#### 2.1.1 Pure and mixed states

The positivity of the density matrix  $\rho$  implies that  $\rho$  is a Hermitian operator (see [Bengtsson and Zyczkowski, 2006], p.194). The spectral theorem ensures us that we can always decompose  $\rho$  in

---

<sup>1</sup>We use the generic notation  $\mathcal{B}(\mathcal{A})$  to refer to the set of bounded operators acting on the generic Hilbert space  $\mathcal{A}$ . The same remark holds for  $\mathcal{S}(\mathcal{A})$ .

<sup>2</sup>An operator  $A$  acting on  $\mathcal{H}$  is said positive if and only if  $\langle\psi|A|\psi\rangle \geq 0$  for all vectors  $|\psi\rangle$  elements of  $\mathcal{H}$ . This property is also often referred as semi-positive definiteness.

an orthonormal basis as

$$\rho = \sum_{i=1}^d p_i |i\rangle\langle i| \quad (2.1)$$

with  $d = \dim \mathcal{H}$  and where the eigenvalues  $\{p_i\}$  are real and the eigenvectors  $\{|i\rangle\}$  form an orthonormal basis:  $\langle i|j\rangle = \delta_{ij}$  for all  $i, j$ . Since  $\rho$  is positive and has trace one, we also have  $\sum_i p_i = 1$  and  $0 \leq p_i \leq 1$ . We define the rank of a density matrix as the number of  $p_i$ s different from zero. A state with a rank-one density matrix is called a *pure state* and can always be written as  $\rho = |\psi\rangle\langle\psi|$ . A state whose density matrix is not pure, meaning that its rank is strictly greater than one, is called a *mixed state*. The state which is a balanced convex sum of all the basis states is called the *maximally mixed state* or simply the *identity*

$$\frac{\mathcal{I}}{d} := \frac{1}{d} \sum_{i=1}^d |i\rangle\langle i|. \quad (2.2)$$

An interesting feature of the set of density matrices  $\mathcal{S}(\mathcal{H})$  is its convexity. Consider two density matrices  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ . The convex sum  $\tau \equiv \lambda\rho + (1 - \lambda)\sigma$  with  $0 \leq \lambda \leq 1$  has trace one

$$\text{tr}[\tau] = \text{tr}[\lambda\rho + (1 - \lambda)\sigma] = \lambda \text{tr}[\rho] + (1 - \lambda) \text{tr}[\sigma] = 1. \quad (2.3)$$

By definition  $\rho$  and  $\sigma$  are positive, meaning that  $\langle x|\rho|x\rangle \geq 0$  and  $\langle x|\sigma|x\rangle \geq 0$  for all elements  $|x\rangle$  of  $\mathcal{H}$ . We see that

$$\langle x|\tau|x\rangle = \lambda\langle x|\rho|x\rangle + (1 - \lambda)\langle x|\sigma|x\rangle \geq 0, \quad (2.4)$$

showing that  $\tau$  also is a positive operator. The boundary of  $\mathcal{S}(\mathcal{H})$ , noted  $\partial\mathcal{S}(\mathcal{H})$ , corresponds to the pure states —they cannot be obtained by convex combination of other elements of the set— while the interior points are mixed states.

### 2.1.2 Composite systems and entanglement

Suppose we have two physical systems  $A$  and  $B$ . To these two systems correspond two Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$ . The two systems together are described by using the tensor product " $\otimes$ ". We associate to the total system  $AB$  the Hilbert space  $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ . The elements of this Hilbert space are called *bipartite states* and can be written as

$$\rho_{AB} = \sum_{i,j,k,l} p_{ijkl} |i\rangle\langle k| \otimes |j\rangle\langle l|, \quad (2.5)$$

or equally

$$\rho_{AB} = \sum_{ab} p_{ab} \rho_a \otimes \rho_b. \quad (2.6)$$

The two different forms of writing a bipartite state comes from the isomorphism  $\mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H}) \cong \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ . One can go from Eq. (2.5) to Eq. (2.6) — and vice versa — by taking  $p_{ijkl} = \sum_{a,b} p_{ab} p_{ij}^a p_{kl}^b$ ,  $\rho_a = \sum_{i,j} p_{ij}^a |i\rangle\langle j|$  and  $\rho_b = \sum_{k,l} p_{kl}^b |k\rangle\langle l|$ . We can iterate the process to build systems composed of many subsystems. The state of such systems are called *multi-partite states*.

### Separable and entangled states

One of the most striking features of quantum mechanics is certainly entanglement. A state  $\rho_{AB}$  is *separable* if and only if there exists a decomposition

$$\rho_{AB}^{\text{sep}} = \sum_k p_k \rho_{A,k} \otimes \rho_{B,k} , \quad p_k \geq 0 \quad \forall k . \quad (2.7)$$

The states that are not separable are called *entangled* states. If a state can be written as

$$\rho_{AB}^{\text{prod}} = \rho \otimes \sigma , \quad (2.8)$$

it is called a *product state* and is obviously separable. But not all the separable states are product states. Only states of the form (2.7) with only one  $p_k$  different from zero (and therefore equal to one) are product states.

### From two systems to one system

We just saw how to describe two systems together. But what is the reverse process? How do we go from a system composed of two subsystems to a description of only one of the two subsystems? This is done by using the partial trace. Starting from a system  $A + B$  we should trace out the system  $B$  to be left with system  $A$  only and vice versa. The state after reduction of the system is called the *reduced state* of the system. The states that describe, respectively, system  $A$  alone or system  $B$  alone are given by

$$\rho_A := \text{tr}_B[\rho_{AB}] \quad , \quad \rho_B := \text{tr}_A[\rho_{AB}] . \quad (2.9)$$

The partial traces  $\text{tr}_A$  and  $\text{tr}_B$  correspond to a trace over just one of the two systems. If  $\{|i\rangle_A\}_{1 \leq i \leq d_a}$  is a basis of  $\mathcal{H}_A$  and  $\{|i\rangle_B\}_{1 \leq i \leq d_b}$  is a basis of  $\mathcal{H}_B$ , then the reduced states are given by

$$\rho_A = \sum_{i=1}^{d_b} {}_B \langle i | \rho_{AB} | i \rangle_B \quad , \quad \rho_B = \sum_{i=1}^{d_a} {}_A \langle i | \rho_{AB} | i \rangle_A . \quad (2.10)$$

Note that when a state is a product state we have  $\rho^{\text{prod}} = \text{tr}_B[\rho^{\text{prod}}] \otimes \text{tr}_A[\rho^{\text{prod}}]$ .

## 2.2 Evolution

### Postulate 2 — EVOLUTION OF A SYSTEM.

The evolution of a physical system is described by a quantum channel  $\mathcal{E}$ . A quantum channel is a completely positive trace preserving (CPTP) linear operator acting on  $\mathcal{B}(\mathcal{H})$ . Starting with the state  $\rho$ , the state after evolution is given by  $\mathcal{E}(\rho)$ . To be a CPTP map, an operator  $\mathcal{E}$  should be:

i) *Trace preserving:* The trace of an operator  $A \in \mathcal{B}(\mathcal{H})$  should remain constant when applying the channel<sup>3</sup>

$$\text{tr}[\mathcal{E}(A)] = \text{tr}[A] . \quad (2.11)$$

ii) *Completely positive:* Let  $A \in \mathcal{B}(\mathcal{H}_1)$  be a positive operator,  $A \geq 0$ . Let us introduce an ancillary system represented by a Hilbert space  $\mathcal{H}_2$  of arbitrary dimension and denote by  $\text{Id}_2$  the identity channel on  $\mathcal{B}(\mathcal{H}_2)$  ( $\text{Id}_2(A_2) = A_2$  for any element  $A_2 \in \mathcal{B}(\mathcal{H}_2)$ ). Then  $\mathcal{E}$  is completely positive if and only if the channel  $\mathcal{E} \otimes \text{Id}_2$  is also a positive channel

$$(\mathcal{E} \otimes \text{Id}_2)(B) \geq 0 \quad \forall B \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) . \quad (2.12)$$

<sup>3</sup>Sometimes only the property of being not trace-increasing is required.

### 2.2.1 Kraus representation

The Kraus theorem gives a useful way to characterize the CPTP maps:

**Theorem 2.1 — KRAUS THEOREM.**

Every trace preserving map  $\mathcal{E}$  can be represented by a set  $\{E_k\}_{1 \leq k \leq r}$  of operators  $E_k \in \mathcal{B}(\mathcal{H})$ , called the Kraus operators. The action of  $\mathcal{E}$  on  $\rho$  is then given by

$$\mathcal{E}(\rho) = \sum_{k=1}^r E_k \rho E_k^\dagger. \quad (2.13)$$

Furthermore the map  $\mathcal{E}$  is also trace preserving if and only if

$$\sum_{k=1}^r E_k^\dagger E_k = \mathcal{I}. \quad (2.14)$$

The Kraus representation of a channel is not unique. Any quantum channel acting on  $\mathcal{B}(\mathcal{H})$  can be written with at most  $d^2$  Kraus operators with  $\dim \mathcal{H} = d$ . The minimal number of Kraus operators needed to represent  $\mathcal{E}$  defines the rank of the channel. When the rank of the channel is equal to  $d^2$  the channel is said to be *full-rank*.

Rank one channels correspond to unitary evolutions  $\mathcal{U}$ . These channels are said *noiseless* as they preserve the purity of states. Unitary evolution corresponds to a Hamiltonian  $H \in \mathcal{B}(\mathcal{H})$  which is a Hermitian operator  $H^\dagger = H$ . The single Kraus operator needed to describe such evolution is written as  $U = e^{-itH}$ . We have that  $U^\dagger = U^{-1}$ , which ensures us that this is a CPTP map as,  $UU^\dagger = \mathcal{I}$ .

Finally we note a last property of quantum channels, their convex linearity:

$$\mathcal{E}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{E}(\rho_i). \quad (2.15)$$

This reflects the fact that the quantum channels are defined as linear maps, the terms convex referring here to the fact that  $\sum_i p_i = 1$ . Using convex linearity, we can write the effect of the channel  $(\mathcal{E} \otimes \text{Id}_2)$  on  $B = \sum_{i,j} b_{ij} C_i \otimes C_j$  —  $B \in \mathcal{B}(H_1 \otimes H_2)$  — as

$$(\mathcal{E} \otimes \text{Id}_2)(B) = \sum_{i,j} b_{ij} \mathcal{E}(C_i) \otimes C_j. \quad (2.16)$$

## 2.3 Measurement

**Postulate 3 — MEASUREMENT ON A SYSTEM.**

Quantum measurements are described by Positive Operator Valued Measure (POVM). A POVM is a set  $\{M_\xi\}$  of positive operators  $M_\xi$  fulfilling the condition  $\sum_\xi M_\xi = \mathcal{I}$ . The result of a measurement is labelled by  $\xi$ . The probability  $p(\xi)$  of obtaining the outcome  $\xi$  when measuring the state  $\rho$  is given by

$$p(\xi) = \text{tr}[M_\xi \rho]. \quad (2.17)$$

Introducing the decomposition of the measurement operators  $M_\xi = E_\xi^\dagger E_\xi$ , the state  $\rho_\xi$  after the measurement with outcome  $\xi$  is given by

$$\rho_\xi = \frac{E_\xi \rho E_\xi^\dagger}{p(\xi)}. \quad (2.18)$$

In general the elements of the POVM do not need to be orthogonal. When they turn out to be orthogonal the measurement is a projective measurement. From a quantum observable  $A = \sum_{i=1}^d \alpha_i |i\rangle\langle i|$ , defined as a Hermitian operator with full rank we can always build a projective measurement  $\{|i\rangle\langle i|\}$  (defined below, cf Postulate 3').

## 2.4 Going to a larger Hilbert space

### 2.4.1 The usual postulates

The postulates of quantum mechanics introduced in the first section correspond to the description of a system that may be open. The historical approach of quantum mechanics focuses on closed system. We present now the corresponding postulates.

**Postulate 1' — STATES.**

*The state of a physical system is represented by a normalized vector  $|\psi\rangle$  of the Hilbert space<sup>4</sup>  $\mathcal{H}$ .*

**Postulate 2' — EVOLUTION.**

*The time evolution of the state of a system is described by the Schrödinger equation*

$$-\frac{i}{\hbar} \frac{\partial |\psi(t)\rangle}{\partial t} = H(t) |\psi(t)\rangle. \quad (2.19)$$

For time independent Hamiltonians  $H$ , starting with the state  $|\psi_0\rangle$  the state after a time  $t$  is given by

$$\psi(t) = U(t) |\psi_0\rangle = e^{-iHt/\hbar} |\psi_0\rangle. \quad (2.20)$$

where  $U$  is called the time evolution operator. For time dependent Hamiltonians  $H(t)$  we introduce the time ordering operator  $\mathcal{T}$ , defined as

$$\mathcal{T}[A(t_1)B(t_2)] := \begin{cases} A(t_1)B(t_2) & \text{if } t_1 > t_2, \\ B(t_2)A(t_1) & \text{if } t_2 > t_1 \end{cases}. \quad (2.21)$$

The state after a time  $t$  is given by (see Appendix B):

$$|\psi(t)\rangle = \mathcal{T} \left[ e^{-i \int_0^t H(t_1) dt_1} \right] |\psi_0\rangle. \quad (2.22)$$

Dynamics generated by time evolution operators are reversible as there always exists an inverse time evolution operator.

**Postulate 3' — MEASUREMENT.**

*A measurement is represented by a quantum observable  $A$ , defined as a full rank Hermitian operator acting on  $\mathcal{H}$  and written  $A = \sum_i \alpha_i P_i$ , where  $P_i$  is the projector onto the eigenspace with eigenvalue  $\alpha_i$ . The possible outcomes of the measurement are given by the eigenvalues  $\alpha_i$ . The probability of getting the outcome  $\alpha_i$  when measuring the state  $|\psi\rangle$  is given by  $p(\alpha_i) = \langle \psi | P_i | \psi \rangle$ . The state after getting the result  $\alpha_i$  is  $|\psi(\alpha_i)\rangle = P_i |\psi\rangle / \sqrt{p(\alpha_i)}$ .*

We actually do not need the eigenvalues of the observable to define a measurement: We define a projective measurement as a set of projectors  $P_i$  such that  $\sum_i P_i = \mathcal{I}$ .

---

<sup>4</sup>More formally states of the system are defined as rays of the Hilbert space, i.e. by an equivalence class of vectors with the equivalence relation:  $v \sim w$  when  $v = \alpha w$  for  $\alpha \in \mathbf{C}^*$ , the set of complex numbers without zero. The set of equivalence class is called the projective Hilbert space. This reflects that an overall phase cannot be observed and thus cannot play any role in the description of a state.

## 2.4.2 The Church of the larger Hilbert space

We present the different theorems that show how to go from one description of quantum mechanics to another one. All these theorems are based on the use of an additional Hilbert space.

### From density matrices to pure states

The purification theorem ensures us that every state  $\rho$  can be written as a pure state  $|\psi\rangle$  in a greater Hilbert space:

**Theorem 2.2 — PURIFICATION OF MIXED STATES.**

Let  $\rho$  be a density matrix on a Hilbert space  $\mathcal{H}_1$  ( $\rho \in \mathcal{S}(\mathcal{H}_1)$ ). There exists a Hilbert space  $\mathcal{H}_2$  and a pure state  $|\psi\rangle\langle\psi|$  on  $\mathcal{H}_1 \otimes \mathcal{H}_2$  such that  $\rho = \text{tr}_2[|\psi\rangle\langle\psi|]$ .

The purification of states is related to another important theorem known as the Schmidt decomposition:

**Theorem 2.3 — SCHMIDT DECOMPOSITION.**

Let  $|\psi\rangle$  be a pure state, element of a Hilbert space  $\mathcal{H}_1 \otimes \mathcal{H}_2$ . We can always write  $|\psi\rangle$  as

$$|\psi\rangle = \sum_{i=1}^d \sqrt{p_i} |\alpha_i\rangle \otimes |\beta_i\rangle, \quad (2.23)$$

where  $d = \min(d_1, d_2)$  and  $p_i \geq 0$ . The sets  $\{|\alpha_i\rangle\}$  and  $\{|\beta_i\rangle\}$  are orthonormal families of  $\mathcal{H}_1$  and  $\mathcal{H}_2$  (but in general they do not constitute a basis of their respective Hilbert space). The  $\{p_i\}$  and  $\{|\alpha_i\rangle\}$  (resp.  $\{|\beta_i\rangle\}$ ) are the eigenvalues and eigenvectors of  $\rho_1 = \text{tr}_2[|\psi\rangle\langle\psi|]$  (resp.  $\rho_2 = \text{tr}_1[|\psi\rangle\langle\psi|]$ ).

### From quantum channels to evolution operators

The Stinespring theorem makes the connection between channels and unitary evolutions in a larger Hilbert space: To every quantum channel corresponds a unitary evolution in a larger Hilbert space.

**Theorem 2.4 — STINESPRING.**

Let  $\mathcal{E}$  be a quantum channel:  $\mathcal{E} : \mathcal{S}(\mathcal{H}_1) \rightarrow \mathcal{S}(\mathcal{H}_1)$ . There exists a Hilbert space  $\mathcal{H}_2$ , a pure state  $|\psi\rangle_2 \in \mathcal{H}_2$ , and a unitary operator  $U$  on  $\mathcal{H}_1 \otimes \mathcal{H}_2$  such that for any density matrix  $\rho$  on  $\mathcal{H}_1$ :

$$\mathcal{E}(\rho) = \text{tr}_2[U\rho \otimes |\psi\rangle_2\langle\psi|_2 U^\dagger]. \quad (2.24)$$

### From POVM to projective measurement

Consequence of the Stinespring theorem, the Neumark's (sometime spelled Naimark's) dilation theorem shows that to every POVM corresponds a projective measurement in a larger Hilbert space:

**Corollary 2.1 — NEUMARK'S DILATION.**

Let  $\{M_\xi\}$  be a POVM on  $\mathcal{H}_1$  with  $\dim \mathcal{H}_1 = d_1$  and  $p = \sum_\xi \text{rank } M_\xi$ . There exists a larger Hilbert space  $\mathcal{H}_2$  of dimension  $d_2 := \dim \mathcal{H}_2 = p - d_1 + 1$ , a pure state  $|\psi\rangle_2 \in \mathcal{H}_2$  and a projective measurement  $\{\Pi_\xi^{12}\}$  such that

$$M_\xi = {}_2\langle\psi|\Pi_\xi^{12}|\psi\rangle_2. \quad (2.25)$$

## Box 4: The Choi–Jamiołkowski isomorphism or channel-state duality

Consider a Hilbert space  $\mathcal{H}$  and the Hilbert space  $\mathcal{B}(\mathcal{H})$  defined as the set of bounded operators acting on  $\mathcal{H}$  equipped with the Hilbert-Schmidt norm. There exists an obvious isomorphism between  $\mathcal{B}(\mathcal{H})$  and  $\mathcal{H} \otimes \mathcal{H}$ . This is best seen using the Dirac notation: To any element  $A = \sum_{i,j=1}^d p_{ij} |i\rangle\langle j|$  of  $\mathcal{B}(\mathcal{H})$  we can associate the element  $\sum_{i,j=1}^d p_{ij} |i\rangle \otimes |j\rangle$  of  $\mathcal{H} \otimes \mathcal{H}$  and vice versa, which shows that  $\mathcal{B}(\mathcal{H}) \cong \mathcal{H} \otimes \mathcal{H}$ .

We can use this isomorphism to put in correspondence  $\mathcal{B}(\mathcal{B}(\mathcal{H}))$  — the set of bounded operators acting on the bounded operators acting on the elements of  $\mathcal{H}$  — with the elements of  $\mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H})$ , itself isomorphic to  $\mathcal{B}(\mathcal{H} \otimes \mathcal{H})$  — the set of operators acting on the elements of  $\mathcal{H} \otimes \mathcal{H}$ . If we restrain our attention to the set of completely positive operators there is an important theorem [Choi, 1975] that states that  $\mathcal{E}$  is completely positive if and only if  $\rho_{\mathcal{E}}$  is a positive operator, with  $\rho_{\mathcal{E}}$  defined as

$$\rho_{\mathcal{E}} \equiv (\mathcal{E} \otimes \text{Id})(|\psi^{\text{m.e.}}\rangle\langle\psi^{\text{m.e.}}|) \quad (2.26)$$

where  $|\psi^{\text{m.e.}}\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle \otimes |i\rangle$  is a maximally entangled state. This is known as the Choi-Jamiołkowski<sup>a</sup> isomorphism between completely positive maps and bipartite states, also known as channel-state duality. If we furthermore impose the constraint on  $\mathcal{E}$  to be trace preserving, then we only obtain an injection from the CPTP maps to a subset of  $\mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ . We now turn to the inverse mapping. Starting with the state  $\rho = \sum_i \omega_i |w_i\rangle\langle w_i| \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ , the effect of the corresponding channel  $\mathcal{E}_{\rho}$  on an input pure state  $|\varphi\rangle\langle\varphi| \equiv \sum_i \varphi_i |i\rangle\langle i| \in \mathcal{B}(\mathcal{H})$  is given by

$$\mathcal{E}(|\varphi\rangle\langle\varphi|) = \sum_i \omega_i \langle\varphi^*|w_i\rangle\langle w_i|\varphi^*\rangle, \quad (2.27)$$

where  $|\varphi^*\rangle = \sum_i \varphi_i^* |i\rangle$ . Due to the linearity of the quantum channels, having the effect of the channel on an arbitrary pure input state amounts to having the effect of the channel on any state.

Notice that expressed in this way, the isomorphism depends on the basis  $\{|i\rangle\}$  taken to construct the maximally entangled state. In the other direction the freedom comes from the choice of the state preparation  $\{\omega_i, |w_i\rangle\}$ .

A detailed discussion on the Choi-Jamiołkowski isomorphism can be found in [Jiang et al., 2013].

<sup>a</sup>In this version the isomorphism is due to Choi, based on the previous results from Jamiołkowski and de Pillis [de Pillis, 1967; Jamiołkowski, 1972].

### 2.4.3 Interpretation and ambiguities

When going from the original Hilbert space to a larger Hilbert space there is a freedom that appears in the procedure. We review some of these ambiguities that will play a role in quantum parameter estimation theory.

### Ensemble preparation

Let  $\rho$  be a density matrix of rank  $r$ . We can write  $\rho$  as a sum of  $q \geq r$  pure states:

$$\rho = \sum_{i=1}^q \omega_i |w_i\rangle\langle w_i|, \quad (2.28)$$

with  $\omega_i > 0$ . The family  $\{\sqrt{\omega_i}|w_i\rangle\}_{1 \leq i \leq q}$  is called a  $(\rho, q)$ -*decomposition*, or simply a *preparation* of  $\rho$ . The term preparation stems from the physical interpretation of Eq. (2.28):  $\rho$  can be constructed by classically mixing pure states  $|w_i\rangle$  with weights  $\omega_i$ . For a given  $q$  this decomposition is not unique and we can go from one  $(\rho, q)$ -decomposition  $\{\sqrt{\omega_i}|w_i\rangle\}_{1 \leq i \leq q}$  to another  $(\rho, q)$ -decomposition  $\{\sqrt{\nu_i}|v_i\rangle\}_{1 \leq i \leq q}$  by a unitary transformation  $U$ <sup>5</sup>, which is an element of the unitary group of degree  $q$  noted  $\mathbb{U}(q)$ :

$$\{\sqrt{\nu_i}|v_i\rangle\} = \left\{ \sum_{k=1}^q u_{ki} \sqrt{\omega_k} |w_k\rangle \right\}, \quad (2.29)$$

with  $u_{ki} = (U)_{ki}$ . It turns out that *all* the  $(\rho, q)$ -decompositions can be obtained starting from a reference  $(\rho, q)$ -decomposition and using the unitary matrices. We denote the set of all  $(\rho, q)$ -decompositions as  $W_q(\theta)$  and we call it the  $(\rho, q)$ -*ensemble*.

### Freedom in the purification

We start with a  $(\rho, q)$ -decomposition  $\{\sqrt{\omega_i}|w_i\rangle\}$  of  $\rho$ . Then we can write a purification of  $\rho$  as

$$|\psi_\rho(e_i)\rangle = \sum_{i=1}^q \sqrt{\omega_i} |w_i\rangle \otimes |e_i\rangle, \quad (2.30)$$

where  $\{|e_i\rangle\}$  is an orthonormal family in  $\mathcal{H}_2$ . One checks that  $\text{tr}_2[|\psi_\rho(e_i)\rangle\langle\psi_\rho(e_i)|] = \rho$ . Since we did not have to specify the form of the vectors  $|e_i\rangle$  we can pick any orthonormal family. This freedom leads to an infinity of different purifications. This freedom is different from the one that appears in the state decomposition. In the state decomposition the freedom allows us to write the *same* state  $\rho$  in different ways, but for purifications the freedom leads to *different* states  $|\psi_\rho\rangle$ . To switch between two purifications one uses unitary matrices:

$$|\psi_\rho(f_i)\rangle = \sum_{i=1}^q \sqrt{\omega_i} |w_i\rangle \otimes U |e_i\rangle, \quad (2.31)$$

with the unitary matrix  $U$  chosen such that  $|f_i\rangle = U |e_i\rangle$ .

Although carrying different meanings, the freedom in state preparation and the freedom in state purification are related. The right action of the unitary matrix  $U$  on the set of pure states of the  $(\rho, q)$ -ensemble corresponds to the left action of  $U$  on the orthonormal basis of the extra Hilbert space introduced for state purification. To show this we start with  $\rho = \sum_{i=1}^q \omega_i |w_i\rangle\langle w_i|$ . The right action of  $U$  on  $\{\sqrt{\omega_i}|w_i\rangle\}$  can be written as  $\{\sqrt{\omega_i}|w_i\rangle\}U = \{\sum_{k=1}^q \sqrt{\omega_k}|w_k\rangle u_{ki}\}$ .

---

<sup>5</sup>Notice that the action of  $U$  on  $\{|w_i\rangle\}_{1 \leq i \leq q}$  does not correspond to the left action but to the right action:  $\{|v_i\rangle\}_{1 \leq i \leq q} \rightarrow \{|w_i\rangle\}_{1 \leq i \leq q} U$ .

Inserting this in the reference preparation gives  $\rho = \sum_{i,k,l}^q \sqrt{\omega_k \omega_l} |w_k\rangle u_{ki} u_{li}^* \langle w_l|$ . Then we can write the purification as

$$|\psi_\rho(f_i)\rangle = \sum_{i=1}^q \sum_{k=1}^q \sqrt{\omega_k} |w_k\rangle u_{ki} \otimes |f_i\rangle = \sum_{k=1}^q \sqrt{\omega_k} |w_k\rangle \otimes \sum_{i=1}^q u_{ki} |f_i\rangle = |\psi_\rho(e_i)\rangle, \quad (2.32)$$

where the  $|e_i\rangle$  are defined as  $|e_i\rangle = \sum_{k=1}^q u_{ki} |f_i\rangle$ , *i.e.* through the left action of  $U$  on the basis vector needed for the purification:  $\{|e_i\rangle\} = U\{|f_i\rangle\}$ .

### Environment model for quantum channels and freedom in the Kraus operators

The physical meaning of the Kraus representation is easily seen when working in the environment model. Consider an ancillary system (often called environment) with a Hilbert space  $\mathcal{H}_2$ . We assume without loss of generality that the environment is initially in a pure state  $|a_0\rangle$  — if the state was mixed we could always increase the dimension of  $\mathcal{H}_2$  to purify it. The total system is closed and therefore evolves under a unitary dynamics  $V: \rho \otimes |a_0\rangle\langle a_0| \rightarrow V(\rho \otimes |a_0\rangle\langle a_0|)V^\dagger$ . By tracing out the environment we obtain the effect of the quantum channel:

$$\mathcal{E}(\rho) := \text{tr}_2[V(\rho \otimes |a_0\rangle\langle a_0|)V^\dagger] \quad (2.33)$$

$$= \sum_{k=1}^{d_2} \langle a_k | V(\rho \otimes |a_0\rangle\langle a_0|) V^\dagger | a_k \rangle \quad (2.34)$$

$$= \sum_{k=1}^{d_2} A_k \rho A_k^\dagger, \quad (2.35)$$

with the Kraus operators  $A_k := \langle a_k | V | a_0 \rangle$  fulfilling  $\sum_{k=1}^{d_2} A_k^\dagger A_k = \mathcal{I}_2$  where  $d_2 := \dim \mathcal{H}_2$ . A set of  $q$  Kraus operators  $\{A_k\}_{1 \leq k \leq q}$  for the channel  $\mathcal{E}$  is called a  $q$ -Kraus decomposition. The Stinespring theorem states that for any  $q$ -Kraus decomposition it is always possible to construct an environment model.

Given a reference  $q$ -Kraus decomposition  $\mathcal{A} = \{A_j\}_{1 \leq j \leq q}$  we can construct all the other  $q$ -Kraus decompositions through the unitary matrices  $U$  of size  $q$ :

$$B_j = \sum_k u_{jk} A_k. \quad (2.36)$$

The set of all  $q$ -Kraus decompositions is called the  $q$ -Kraus ensemble and is denoted  $\mathcal{A}_q$ . This freedom comes from the freedom of the choice of the basis of the environment: we can always pick up a unitary transformation  $U$  acting only on the environment before tracing it out. This is equivalent to performing the partial trace in a different basis. As a result the new Kraus operators are defined as

$$B_j := \langle a_j | (\mathcal{I}_1 \otimes U) V | a_0 \rangle. \quad (2.37)$$

Inserting the identity  $\sum_{k=1}^{d_2} |a_k\rangle\langle a_k| = \mathcal{I}_2$ , we obtain  $B_j = \sum_k \langle a_j | \mathcal{I}_1 \otimes U | a_k \rangle \langle a_k | V | a_0 \rangle = \sum_k u_{jk} A_k$ . The situation is very similar to the one encountered with states: the freedom in the  $q$ -Kraus decomposition is equivalent to the freedom in the Stinespring dilation in the same way as the freedom in the  $(\rho, q)$ -ensemble is equivalent to the freedom in the purification.

We can make this analogy more formal by using the Choi-Jamiołkowski isomorphism (see box 4). To the quantum channel  $\mathcal{E}_\rho$  corresponds a state  $\rho_\mathcal{E}$  in  $\mathcal{S}(\mathcal{H} \otimes \mathcal{H})$ . The Stinespring dilation

theorem can then be obtained through the purification of the state  $\rho_{\mathcal{E}}$ , or the other way around, the purification theorem can be obtained through the application of the Stinespring theorem to  $\mathcal{E}_{\rho}$ .

## 2.5 Distance between states

The set  $\mathcal{S}(\mathcal{H})$  of density matrices is a convex set embedded in the set of bounded linear operators  $\mathcal{B}(\mathcal{H})$ . The latter has the structure of a Hilbert space when equipped with the Hilbert-Schmidt scalar product  $\langle A, B \rangle$  defined as

$$\langle A, B \rangle = \text{tr}[A^{\dagger}B] . \quad (2.38)$$

From this scalar product we can express easily the distance between two elements  $A$  and  $B$  of  $\mathcal{B}(\mathcal{H})$  as  $d_{\text{HS}}(A, B) = \sqrt{\langle A - B, A - B \rangle} = \sqrt{\text{tr}[(A - B)(A^{\dagger} - B^{\dagger})]}$ . We can also use the Hilbert-Schmidt distance to calculate how far apart two states are from each other.

Apart from the Hilbert-Schmidt norm there are many norms in  $\mathcal{B}(\mathcal{H})$  and therefore as many distances between states. There is a connection between the geometry of quantum states and quantum estimation theory, as there is such a connection between geometry of probability distributions and classical estimation theory. In quantum mechanics the Bures distance makes this connection.

### 2.5.1 Fidelity and Bures distance

Before going to the Bures distance we present the fidelity. The fidelity  $\mathcal{F}(\rho, \sigma)$  between two states  $\rho$  and  $\sigma$  is defined as

$$\mathcal{F}(\rho, \sigma) = \left( \text{tr} \left[ (\rho^{1/2} \sigma \rho^{1/2})^{1/2} \right] \right)^2 , \quad (2.39)$$

where the square root  $A^{1/2}$  of an operator  $A$  is defined such that  $A^{1/2}A^{1/2} = A$ . It is uniquely defined (up to a minus sign) when the operator  $A$  is positive. For pure states the fidelity reduces to the overlap between the states

$$\mathcal{F}(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = |\langle\psi|\phi\rangle|^2 . \quad (2.40)$$

As its name indicates the fidelity describes how similar two states are. When states are identical the fidelity is equal to one  $\mathcal{F}(\rho, \rho) = 1$ , while states with an orthogonal support have a vanishing fidelity:  $\mathcal{F}(\rho, \sigma) = 0$  for  $\text{supp}(\rho) \perp \text{supp}(\sigma)$ . We see that this is already close to the idea of a distinguishability distance.

Starting from the fidelity we can construct a function which fulfils the condition of a distance, the Bures distance  $d_{\text{B}}$ . The Bures distance between  $\rho$  and  $\sigma$  is given by

$$d_{\text{B}}(\rho, \sigma) = (2 - 2\sqrt{\mathcal{F}(\rho, \sigma)})^{1/2} . \quad (2.41)$$

### 2.5.2 Properties of the Bures distance

The fidelity being a function from  $\mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H})$  to  $[0, 1]$  the Bures distance is a function from  $\mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H})$  to  $[0, \sqrt{2}]$ . The boundaries are reached for

- Identical states:  $\rho = \sigma$  which implies  $d_{\text{B}}(\rho, \sigma) = 0$ .
- Orthogonal states:  $\text{supp}(\rho) \perp \text{supp}(\sigma)$  which implies  $d_{\text{B}}(\rho, \sigma) = \sqrt{2}$ .

A property which will have important physical consequences when going back to metrology is the monotonicity of the Bures distance over quantum channels. This property is also known as contractivity. It states that when applying a quantum channel to two states, the distance between the new states can only be smaller or equal to the distance between the original states:

$$d_{\text{B}}(\rho, \sigma) \leq d_{\text{B}}(\mathcal{E}(\rho), \mathcal{E}(\sigma)). \quad (2.42)$$

Equality is reached for unitary channels

$$d_{\text{B}}(\rho, \sigma) = d_{\text{B}}(U\rho U^\dagger, U\sigma U^\dagger), \quad (2.43)$$

with  $U$  a unitary matrix.

## 2.6 Two-level system: the model of the qubit

Throughout this thesis we focus on quantum systems with finite dimension. The smallest (in terms of dimension of the associated Hilbert space) non-trivial system that one could think of is a system with only two levels. In the field of quantum information such systems are usually called qubits. This name makes the parallel with the basic unit of information in classical information theory, the bit. From a physical perspective any system that has only two levels, or from which we consider only two levels, can be treated as a qubit.

### 2.6.1 State of a qubit and Bloch vector picture

#### Basis of the Hilbert space

The Hilbert space associated to a qubit is  $\mathcal{H} = \mathbf{C}^2$ , where  $\mathbf{C}$  is the set of the complex numbers. The three Pauli matrices  $X$ ,  $Y$  and  $Z$  — along with the identity operator — form a basis of the elements of  $\mathcal{B}(\mathcal{H})$  and are defined as

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.44)$$

In quantum information, it is rewarding to introduce a specific notation called the computational notation for the states of a qubit:  $\{|0\rangle, |1\rangle\}$ . Especially the effect of  $Z$  on the basis vectors is given by

$$Z|0\rangle = |0\rangle, \quad Z|1\rangle = -|1\rangle, \quad (2.45)$$

meaning that when looking at a Hamiltonian given by  $Z$ , the state  $|0\rangle$  has the higher energy (excited state, or state up) and the state  $|1\rangle$  has the lower energy (ground state, or state down). The advantage of the computational basis (and notation) becomes evident when we deal with multiple qubits: we can then use the string of "0" and "1" of a state of multiple qubits to go from a binary label to a decimal one.

The eigenvectors of  $X$  are usually denoted as  $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$  and  $|-\rangle := (|0\rangle - |1\rangle)/\sqrt{2}$  with respective eigenvalues 1 and  $-1$ .

### Mixed state and Bloch ball

The simplicity of the qubit allows us to parametrize the density matrix of a qubit in a simple and elegant way. To do so we use that the three Pauli matrices plus the identity matrix constitute a basis of the  $2 \times 2$  Hermitian matrices. We can write any density matrix  $\rho$  for a qubit as

$$\rho = \frac{\mathcal{I} + r_x X + r_y Y + r_z Z}{2}. \quad (2.46)$$

Since the Pauli matrices have a vanishing trace, this state has always trace one. To be a proper density matrix we should also be sure that it is a positive operator. This is the case when  $r_x^2 + r_y^2 + r_z^2 \leq 1$ . This parametrization provides a nice geometrical picture. We define the Bloch vector  $\mathbf{r} := (r_x, r_y, r_z)$ . As we said this vector has to have a norm smaller than one, and therefore any state of a qubit can be directly mapped through the Bloch vector to a point of the ball. The surface of the ball, the Bloch sphere, represents the pure states. Following this representation, it is convenient to introduce the ket  $|\theta, \phi\rangle = \cos(\theta/2)|0\rangle + e^{i\varphi} \sin(\theta/2)|1\rangle$ , where  $\theta$  is called the polar angle and  $\varphi$  the azimuthal angle (see Fig. 2.1). If we also allow mixed states, we parametrize the Bloch vector as  $\mathbf{r} = (r \cos(\theta) \cos(\phi), r \cos(\theta) \sin(\phi), r \sin(\theta))$ . The purity of the qubit is given by  $(1 + r^2)/2$ , with  $r = \sqrt{\mathbf{r} \cdot \mathbf{r}}$ , which is equal to one for pure states and to one-half for the maximally mixed state  $\mathcal{I}/2$ .

In order to shorten our notation we can also write the Pauli matrices as  $\sigma_1 = X$ ,  $\sigma_2 = Y$  and  $\sigma_3 = Z$ . Then we can introduce the Pauli vector  $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ . Using Einstein notation the density matrix is written as  $\rho = (\mathcal{I} + r_i \sigma_i)/2$  with  $i = 1, 2, 3$ . We can shorten even more our notation by using four vectors and define  $\tilde{\mathbf{r}} := (r_0 = 1, r_1, r_2, r_3)$  and  $\tilde{\boldsymbol{\sigma}} := (\sigma_0 = \mathcal{I}, \sigma_1, \sigma_2, \sigma_3)$ . Then the state is simply written  $\rho = \tilde{r}_\nu \tilde{\sigma}_\nu$  with  $\nu = 0, 1, 2, 3$ .

### 2.6.2 Assembly of qubits and spins

It is quite common when studying systems with qubits to consider an assembly of qubits. Interestingly such an assembly can be represented by introducing effective spin systems. We will now briefly review the notion of spin and then see how we can go from multiple qubits to spin systems.

#### Spin and spin coupling

Formally we introduce the spin through the spin operators  $S_x$ ,  $S_y$  and  $S_z$ , which, being Hermitian, are also observables. These operators follow the commutation relation  $[S_i, S_j] = i \varepsilon_{ijk} S_k$  for all  $i, j, k \in \{x, y, z\}$  (we take  $\hbar = 1$ ). It turns out that up to a factor  $i$  the spin operators are the generators of  $\mathfrak{su}(2)$ , the Lie algebra of  $SU(2)$ , the special unitary group of dimension two. The different irreducible representations (irreps) of  $SU(2)$  are represented by integer and half integer values which correspond to the different values of the spin. The fundamental irrep corresponds to a spin  $\frac{1}{2}$ , *i.e.* to a qubit. We can further introduce the spin states  $|j, m\rangle$  as the common eigenvectors —elements of the space on which the representation acts— of  $J^2 := J_x^2 + J_y^2 + J_z^2$  and, say,  $J_z$ :

$$J^2 |j, m\rangle = j(j+1) |j, m\rangle, \quad (2.47)$$

$$J_z |j, m\rangle = m |j, m\rangle. \quad (2.48)$$

In general we have seen that in order to describe multiple systems we should use the tensor product. When we have two spins, *i.e.* two irreps of  $SU(2)$ , the total system is also given by the tensor product of the irreps, giving rise to a reducible representation. By definition this reducible

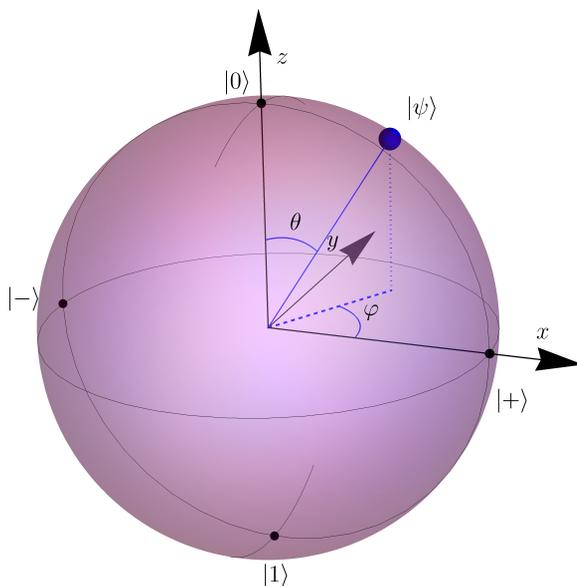


Figure 2.1: Bloch sphere. We represented the states  $|0\rangle$ ,  $|1\rangle$ ,  $|+\rangle$  and  $|-\rangle$  as well as a general pure state  $|\psi\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)e^{i\varphi}|1\rangle$ . The angle  $\theta$  is the polar angle and the angle  $\varphi$  is the azimuthal angle.

representation can be written as a direct sum of irreps. In general if we add two spins  $j_1$  and  $j_2$  we can construct all the spins  $j$  between  $j_1 + j_2$  and  $|j_1 - j_2|$ . Denoting the Hilbert space of a spin  $j$  as  $\mathcal{H}_j$  we have  $\mathcal{H}_{j_1} \otimes \mathcal{H}_{j_2} = \bigoplus_{j=|j_1-j_2|}^{j_1+j_2} \bigoplus_{i_j=1}^{d_j} \mathcal{H}_j^{(i_j)}$  with  $d_j$  the degeneracy of the irrep  $j$ . The extremal spins  $j = j_1 + j_2$  and  $j = |j_1 - j_2|$  are not degenerate ( $d_{j_1+j_2} = d_{|j_1-j_2|} = 1$ ).

### Symmetric subspace of $N$ qubits and coherent state

We now turn our attention to the coupling of  $N$  qubits. Following the rules of addition for the addition of spins when coupling two qubits we obtain a spin 0 and a spin 1. Adding another qubit we obtain a spin  $\frac{3}{2}$ , a spin  $\frac{1}{2}$ , and another spin  $\frac{1}{2}$ . A qubit more and we are left with a spin 2, three spins 1 and two spins 0. In general, for  $N$  qubits we obtain one spin  $\frac{N}{2}$  and all the other spins from  $\frac{N}{2}$  to 0 for  $N$  even, or to spin  $\frac{1}{2}$  for  $N$  odd.

The case with the total spin  $\frac{N}{2}$  corresponds to completely symmetric states of the  $N$  qubits. Suppose all the qubits are in the state  $|\psi\rangle = \cos(\alpha)|0\rangle + \sin(\alpha)e^{i\varphi}|1\rangle$ . The state of the total system is written as

$$|\psi^{(N)}(\alpha, \varphi)\rangle := \bigotimes_{i=1}^N (\cos(\alpha)|0\rangle_i + \sin(\alpha)e^{i\varphi}|1\rangle_i). \quad (2.49)$$

We introduce the Dicke state  $|\frac{N}{2}, m\rangle$  which is obtained by taking all the permutations of states

with  $N/2 + m$  qubits in the state “1” and all the other in state “0”:

$$|\frac{N}{2}, m\rangle := \frac{1}{\sqrt{\binom{N}{m+\frac{N}{2}}}} \sum_{\sigma \in S_N} |0\rangle_{\sigma(1)} \cdots |0\rangle_{\sigma(\frac{N}{2}-m)} |1\rangle_{\sigma(\frac{N}{2}-m+1)} \cdots |1\rangle_{\sigma(N)}, \quad (2.50)$$

where  $S_N$  is the group of permutation over  $N$  elements [Dicke, 1954]. It turns out that the Dicke states form an orthonormal basis of the space of the effective spin  $j = \frac{N}{2}$ . By developing the right hand side of Eq. (2.49) we can re-write  $|\psi^{(N)}(\alpha, \varphi)\rangle$  as a spin  $j = \frac{N}{2}$ ,

$$|\psi^{(N)}(\alpha, \varphi)\rangle = \sum_{m=-N/2}^{N/2} \sqrt{\binom{N}{m+\frac{N}{2}}} \cos(\alpha)^{\frac{N}{2}+m} (\sin(\alpha) e^{i\varphi})^{\frac{N}{2}-m} |\frac{N}{2}, m\rangle. \quad (2.51)$$

The states  $|\psi^{(N)}(\alpha, \varphi)\rangle$  are known as coherent spin states [Radcliffe, 1971; Agarwal, 1981].

### Two qubits and Bell state

When dealing with two qubits there are four states, known as the *Bell states* that are often used in quantum information and therefore worth to present:

$$|\phi_{\pm}\rangle = \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}, \quad |\varphi_{\pm}\rangle = \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}. \quad (2.52)$$

All these four states are maximally entangled, which means that if we trace out one or the other qubit we end up with the maximally mixed state.

## 2.6.3 Quantum channels for qubits

### Arbitrary channels acting on a qubit

As we have seen, the quantum channels are linear operators acting on density matrices. We can use the representation of a qubit as a four vector  $\tilde{\mathbf{r}}$  to describe the channel. With this notation a channel can be represented by a  $4 \times 4$  matrix  $D$ . If  $\tilde{\mathbf{r}}$  is the four-vector associated to the initial state  $\rho$ , then the four-vector after the action of  $\mathcal{E}$  is given by  $D\tilde{\mathbf{r}}$ . Trace preservation and complete positivity impose constraints on  $D$ . Among them, trace preservation is the easiest to implement. Indeed the condition that  $\text{tr}[\rho] = 1$  amounts in forcing the first element of the four-vector representing the state to be equal to one. This enforces the first row of the matrix  $D$  to be equal to  $(1, 0, 0, 0)$ . By writing this matrix as

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ t_1 & M_{11} & M_{12} & M_{13} \\ t_2 & M_{21} & M_{22} & M_{23} \\ t_3 & M_{31} & M_{32} & M_{33} \end{pmatrix},$$

we obtain the transformation of the Bloch three-vector  $\mathbf{r} \rightarrow \mathbf{r}' = \mathbf{t} + M\mathbf{r}$ , *i.e.* the linear transformation of the state corresponds to an affine transformation of the Bloch vector (but to a linear transformation of the four-vector). We can diagonalize the matrix  $M$  as  $M = O_1 \Lambda O_2^t$ , with  $O_1$  and  $O_2$  two orthogonal matrices, and  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ . The geometrical interpretation is that

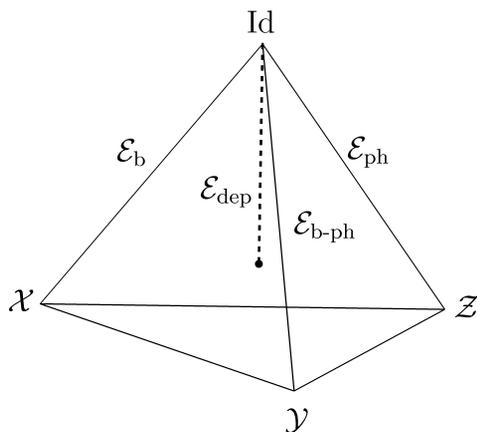


Figure 2.2: Tetrahedron representing the unital channels. The vertices represent:  $\mathcal{X}$  the deterministic bit flip,  $\mathcal{Z}$  the deterministic phase flip,  $\mathcal{Y}$  the deterministic bit-phase flip and  $\text{Id}$  the identity channel. We also labeled the edge representing the three qubit error channels ( $\mathcal{E}_b$ ,  $\mathcal{E}_{\text{ph}}$  and  $\mathcal{E}_{b\text{-ph}}$ ). The dashed line represents the depolarizing channel  $\mathcal{E}_{\text{dep}}$ .

$\Lambda$  transforms the Bloch ball to an ellipsoid, while  $\mathbf{t}$  is a shift of the center of mass. The two orthogonal transformations  $O_1$  and  $O_2$  correspond only to a change of orientation, and can often be omitted.

We should have a look at the complete positivity of this channel. The simple positivity amounts to a clear geometrical condition: the ellipsoid should lie inside the original Bloch sphere. But complete positivity is more restricting, and it turns out that the ellipsoid being inside the Bloch sphere is not a sufficient condition. The exact conditions are far from being trivial [Beth Ruskai et al., 2002; Braun et al., 2014a], but there is one case where this conditions simplifya lot and give rise to a very nice geometric representation, namely when dealing with unital channels.

## Unital and Pauli channels

**Fujiwara-Algoet conditions** Unital channels are defined as channel leaving the identity matrix invariant (or equivalently the completely mixed state):  $\mathcal{E}(\mathcal{I}) = \mathcal{I}$ . They are represented by a matrix  $D$  with  $\mathbf{t} = \mathbf{0}$ , which means that there is only a deformation of the Bloch sphere and not a translation of the ellipsoid. The conditions for complete positivity, known as the Fujiwara-Algoet conditions [Fujiwara and Algoet, 1999], take the simple form

$$(1 \pm \lambda_1)^2 \geq (\lambda_2 \pm \lambda_3)^2. \quad (2.53)$$

If we represent a channel by a vector  $\boldsymbol{\lambda} := (\lambda_1, \lambda_2, \lambda_3)$ , then the Fujiwara-Algoet conditions describe a tetrahedron, contained in a cube which represents the simply positive channels. The vertices of the tetrahedron correspond to the channels defined by  $\boldsymbol{\lambda} \equiv (\lambda_1, \lambda_2, \lambda_3) = (1, 1, 1)$ ,  $(1, -1, -1)$ ,  $(-1, 1, -1)$  and  $(-1, -1, 1)$ .

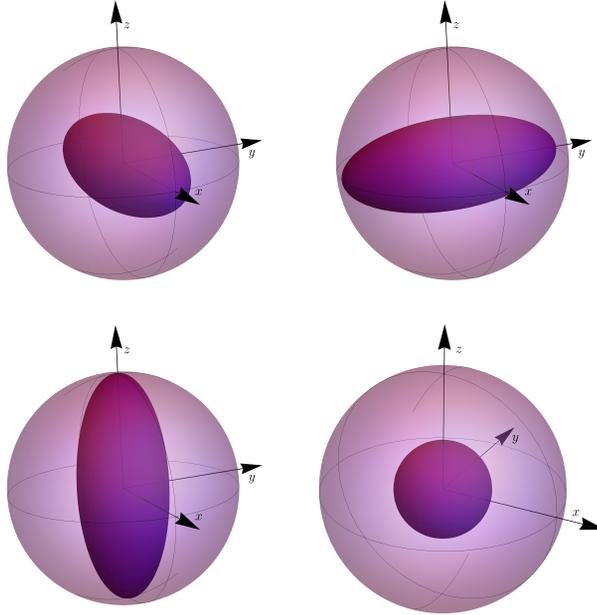


Figure 2.3: From left to right and top to bottom: Effect of the bit-flip, bit-phase flip, phase-flip and depolarizing channel on the Bloch sphere.

**Pauli representation** Pauli channels are channels built by using the Pauli matrices  $\sigma_i$  as Kraus operators:

$$\mathcal{E}(\rho) = \sum_{i=0}^3 d_i \sigma_i \rho \sigma_i, \quad (2.54)$$

with  $\sum_{i=0}^3 d_i = 1$ . Using that  $\sigma_i^2 = \mathcal{I}$  one checks easily that the Pauli channels are unital. We can represent a Pauli channel by the four vector  $\mathbf{d} := (d_0, d_1, d_2, d_3)$ . In this representation the edges of the channel correspond to the vectors  $(1, 0, 0, 0)$ ,  $(0, 1, 0, 0)$ ,  $(0, 0, 1, 0)$  and  $(0, 0, 0, 1)$ . The connection between the representation in terms of distortion of the Bloch sphere given by  $\boldsymbol{\lambda}$  and the representation in terms of Pauli operators given by  $\mathbf{d}$  is given by

$$\begin{cases} d_1 = (1 + \lambda_1 - \lambda_2 - \lambda_3)/4 \\ d_2 = (1 - \lambda_1 + \lambda_2 - \lambda_3)/4 \\ d_3 = (1 - \lambda_1 - \lambda_2 + \lambda_3)/4 \end{cases} \Leftrightarrow \begin{cases} \lambda_1 = 1 - 2d_2 - 2d_3 \\ \lambda_2 = 1 - 2d_1 - 2d_3 \\ \lambda_3 = 1 - 2d_1 - 2d_2 \end{cases}, \quad (2.55)$$

and by the normalization condition we find  $d_0 = (1 + \lambda_1 + \lambda_2 + \lambda_3)/4$ . Using this equation we can get the correspondence between the edges of the tetrahedron. For example the edge  $\boldsymbol{\lambda} = (1, 1, 1)$  is equivalent to  $\mathbf{d} = (1, 0, 0, 0)$ . This was expected, as  $\boldsymbol{\lambda} = (1, 1, 1)$  means that the Bloch sphere does not suffer any distortion, which in terms of Kraus operators will be reached by acting only with the identity.

### Some qubit channels

**Error channels** Some Pauli channels are widely used in quantum computation and information. First let us mention the three "qubit error channels". Here the term error comes from the field of computation, where those three channels are considered as the basic errors that can affect a qubit.

1. The bit-flip channel  $\mathcal{E}_b$ : Flips the basis states of the qubit with a probability  $p$ :  $\mathcal{E}_b(\rho) = pX\rho X + (1-p)\rho$  with  $p \in [0, 1]$ . This channel correspond to the convex combination of the identity channel and the deterministic bit flip  $\mathcal{X}$  defined as  $\mathcal{X}(\rho) = X\rho X$ . We see that this channel is represented by the edge connecting the vertex corresponding to the identity to the vertex corresponding to  $\mathcal{X}$ . See top-left figure in Fig. 2.3.
2. The phase-flip channel  $\mathcal{E}_{ph}$ : Flips the phase of the qubit with a probability  $p$ :  $\mathcal{E}_{ph}(\rho) = pZ\rho Z + (1-p)\rho = p\mathcal{Z}(\rho) + (1-p)\text{Id}(\rho)$  with  $p \in [0, 1]$ . The corresponding edge is the one connecting the identity to  $\mathcal{Z}$ . See bottom-left figure in Fig. 2.3.
3. The bit-phase flip channel  $\mathcal{E}_{b-ph}$ : Combination of the phase-flip error and the bit-flip error:  $\mathcal{E}_{b-ph}(\rho) = pY\rho Y + (1-p)\rho = p\mathcal{Y}(\rho) + (1-p)\text{Id}(\rho)$  with  $p \in [0, 1]$ . This channel corresponds to the edge connecting the identity to the vertex corresponding to  $\mathcal{Y}$ . See top-right figure in Fig. 2.3.

The importance of these three errors is that they can serve as a "basis" for errors in a qubit-based quantum computer. If one can fight these three errors one will be able to fight any kind of error affecting a qubit [Nielsen and Chuang, 2011].

**Depolarizing channel** Finally we introduce a last important Pauli channel, the depolarizing channel  $\mathcal{E}_{dep}$ . From a geometric perspective this channel corresponds to the line between the center of mass of the tetrahedron and the vertex corresponding to the identity. Its effect is to contract uniformly the Bloch sphere as depicted in the bottom-right figure in Fig. 2.3. The effect of the channel is  $\mathcal{E}_{dep} = p\frac{\mathcal{I}}{2} + (1-p)\rho$  with  $p \in [0, 1]$ , which has the interpretation of replacing with a probability  $p$  the original state by the maximally mixed state. The Kraus representation corresponds to  $\mathbf{d}_{dep} = (1 - 3\frac{p}{4}, \frac{p}{4}, \frac{p}{4}, \frac{p}{4})$ . This leads to another interpretation of the depolarizing channel: Each basic error, bit-flip, phase-flip and bit-phase flip, happens with a probability  $\frac{p}{4}$ .

## Summary Chapter 2

- **States of a system:** Represented by positive trace one operators  $\rho$  acting on  $\mathcal{H}$ :  $\rho \in \mathcal{S}(\mathcal{H})$ .
- **Dynamic:** Evolution is done through quantum channels  $\mathcal{E}$ . Quantum channels are Completely Positive Trace Preserving maps. Every quantum channel can be written in Kraus decomposition  $\mathcal{E}(\rho) = \sum_{k=1}^r E_k \rho E_k^\dagger$  with  $\sum_{k=1}^r E_k^\dagger E_k = 1$ .
- **Measurement:** Corresponds to a Positive Operator Valued Measure (POVM), a set of operators  $\{M_\xi\}$  with  $\sum_\xi M_\xi = \mathcal{I}$  and  $M_\xi \geq 0$ . The probability of obtaining the result  $\xi$  when measuring  $\rho$  is given by  $p(\xi) = \text{tr}[M_\xi \rho]$ .
- **Bures distance:** Distance between quantum states. In terms of fidelity  $\mathcal{F}(\rho, \sigma) = \text{tr}[(\rho^{1/2} \sigma \rho^{1/2})^{1/2}]^2$ , the Bures distance is written  $d_B(\rho, \sigma) = (2 - 2\sqrt{\mathcal{F}(\rho, \sigma)})^{1/2}$ .
- **Qubit:** Two-level system with state  $\rho = \frac{\mathcal{I} + r_1 \sigma_1 + r_2 \sigma_2 + r_3 \sigma_3}{2}$ . The Bloch vector  $\mathbf{r} = (r_1, r_2, r_3)$  represents the states of a qubit in a unit sphere of dimension three.
- **Pauli channels:** Act on qubits and admit the totally mixed state as stationary state. Their Kraus representation is given by  $\mathcal{E}(\rho) = \sum_{i=0}^3 d_i \sigma_i \rho \sigma_i$  with  $\sum_{i=0}^3 d_i = 1$ .

## Chapter 3

# Quantum parameter estimation theory

In this chapter we present the theory which lies at the core of this thesis, *quantum parameter estimation theory*, abbreviated as q-pet. In q-pet we usually do not interrogate the concept of measurement itself, rather we use the formalism of measurement in its modern form, through POVM's, to investigate how precisely we can estimate a parameter encoded in a quantum system. Obviously the question of measurement is also a cornerstone of quantum mechanics. Collapse of wave functions and models to explain it, von Neumann chains, decoherence and pointer basis, or weak measurements raise fundamental questions and are still the topic of intense investigations. In quantum parameter estimation we emphasize neither how to design protocols to increase the sensitivity of a specific measurement, which is the topic of quantum metrology and will be treated in the next chapter.

Quantum parameter estimation theory has been raised in the seventies and became a wide and rich field. We can, as it is also done in classical estimation theory, divide the field in two main approaches, the frequentist approach and the Bayesian approach. In this chapter we present the frequentist theory, putting the emphasis on the Quantum Cramér-Rao theorem and the quantum Fisher information.

### 3.1 Introduction to q-pet

#### 3.1.1 Estimation theory in physics

In our presentation of parameter estimation theory in Chapter 1 we have seen that the theory is built on the concept of probabilities and random processes. To apply this mathematical field to the realm of physics we need physical processes which have a random character. Notice that in a certain sense this is a fair return, since, as we briefly discuss it in the appendix, the mathematical conceptualization of probability is historically based on physical questions.

Quantum theory being the only fundamental physical theory which carries an intrinsic random character, it is natural to apply parameter estimation theory to it. In quantum physics random variables appear directly with the pair "State + POVM". Moreover the concept of measurement is well defined in quantum mechanics, with a simple and elegant formulation. When dealing with classical physics the concept of randomness is not as fundamental. Newtonian mechanics or

classical electromagnetism are intrinsically deterministic theories. Moreover they do not provide a unique and mathematically well defined way to describe measurements. To use parameter estimation theory in this context we have to rely on specific models of measurement and noise.

As an important consequence, it turns out that quantum parameter estimation theory carries a fundamental meaning. It allows one to calculate fundamental bounds on the precision with which we can estimate a *physical* parameter. These bounds are said fundamental in the sense that they are not dictated by a lack of knowledge of our system but by the law of physics themselves.

### 3.1.2 Formalization of quantum parameter estimation theory

We want to estimate a parameter encoded in a quantum system. What is the natural way to formalize this problem? Working with a physical system the only way to obtain information is to make some kind of measurement. We know that measurements in quantum mechanics correspond to POVMs, and that the object which is measured is a quantum state represented by a density matrix.

We state the fundamental question in quantum parameter estimation theory: *given a family of states  $\{\rho_\theta\}$  what can we say about the value of  $\theta$ ?*

This question is the transposition to quantum mechanics, a physical theory, of the question which lies at the core of parameter estimation theory, a mathematical theory. We will see that answering this question will lead us to the quantum version of the Cramér-Rao theorem. There are two different ways to derive the quantum version of the theorem. The first approach, which we will call the physical approach, is directly based on the parameter estimation theory developed in Chapter 1 and offers a nice operational meaning. The second approach, historically first developed, is more abstract and we refer to it as the mathematical approach. It consists in redefining some quantities used in parameter estimation theory and to follow the same line of the proof to obtain directly the quantum Cramér-Rao theorem.

### 3.1.3 Classical *versus* Quantum

We now discuss a semantic problem raised by our presentation of quantum parameter estimation theory. It concerns the use of the adjectives *classical* and *quantum*. Indeed the field of parameter estimation theory as presented in Chapter 1 is often named, by quantum physicists, *classical* parameter estimation theory. Usually, in modern physics the term *classical* is used to describe the physical theories that appeared before quantum mechanics and general relativity. But parameter estimation theory as presented in Chapter 1 is a mathematical theory which does not seem to be based on any physical assumptions.

First it is important to recall that probability theory did not pop-up as a pure abstract mathematical theory but rather to formalize random processes observed at that time (games of chance). And at that time whatever was observed was classical. It actually turns out that we can also construct a non-classical theory of probabilities. The crucial point to distinguish them is the commutativity. In classical theory, the fundamental object, apart of the random variable, is the probability distribution, while in quantum theory the fundamental object is the state. And we know that in general states do not commute. In this framework the opposition classical/quantum is indeed an opposition based on the dichotomy commutative/noncommutative.

## 3.2 Quantum Cramér-Rao Bound

The application of results and ideas from classical statistics and from information theory to quantum mechanics started in the sixties. Carl W. Helstrom (1925-2013), one of the founding fathers of quantum information theory, was at that time working on signal detection for radars and communication technologies [Helstrom, 1960]. During the sixties his work on electrical engineering led him to start to develop what would be later known as q-pet, with a focus on the detection of optical signals.

In 1967, Helstrom published what is probably the primal paper of q-pet [Helstrom, 1967], entitled “Minimum mean-squared error of estimates in quantum statistics”. In slightly less than two pages he introduced the problem of estimation in a quantum setting, formulated the quantum Cramér-Rao theorem and proved it! In 1968, he extended his work to the multi-parameter case [Helstrom, 1968a]. He gave a first review of this work in [Helstrom, 1968b, 1969] and wrote a reference text book on these topics in 1976 [Helstrom, 1976]. Thereafter, authors from the mathematical community started to work on similar problems [Holevo, 1974, 1973a,b], providing a more abstract and mathematically rigorous way to deal with estimation theory in quantum mechanics.

Interestingly we see that by the end of the seventies a large part of the fundamental ideas and results of q-pet were already discovered, but it would still take about twenty years before the field became very popular and widely studied and applied.

### 3.2.1 Quantum version of the theorem

The situation is the following: We are given a quantum state  $\rho_\theta$  of known form and want to estimate the value of  $\theta$ . Once more we will focus on scalar parameter estimation:  $\theta \in \Theta \subset \mathbf{R}$ . To estimate  $\theta$  we will work directly in the framework of local estimation theory. Our goal is to find a bound on the variance of an unbiased estimator.

The answer is given by the following theorem:

**Theorem 3.1 — QUANTUM CRAMER RAO BOUND (QCRB).**

Consider a family of states  $\{\rho_\theta\}$  depending on a scalar parameter  $\theta \in \Theta$ . The variance  $\text{Var}[\hat{\theta}_{\text{est}}]$  of any locally unbiased estimator  $\hat{\theta}_{\text{est}}$  at  $\theta = \theta_0$ ,

$$\mathbb{E}[\hat{\theta}_{\text{est}}] \Big|_{\theta_0} = \theta_0, \quad \frac{\partial}{\partial \theta} \mathbb{E}[\hat{\theta}_{\text{est}}] \Big|_{\theta_0} = 1,$$

is bounded by the inverse of the Quantum Fisher Information

$$\text{Var}[\hat{\theta}_{\text{est}}] \geq \frac{1}{I(\rho_\theta; \theta)}, \quad (3.1)$$

with  $I(\rho_\theta; \theta)$  the Quantum Fisher Information (QFI).

As there are two equivalent ways to prove this theorem, there are two alternative but equivalent ways of defining the QFI. As a reference we will take the definition coming from the mathematical approach. The following definition is thus due to Helstrom.

**Definition 3.1 — QUANTUM FISHER INFORMATION.**

The QFI  $I(\rho_\theta; \theta)$  for the family of states  $\{\rho_\theta\}$  is defined as<sup>1</sup>:

$$I(\rho_\theta; \theta) := \text{tr}[L_{\rho_\theta}^2 \rho_\theta], \quad (3.2)$$

---

<sup>1</sup>For the sake of concision, when calculating the QFI of a pure state  $|\psi_\theta\rangle$  we use the notation  $I(|\psi_\theta\rangle\langle\psi_\theta|; \theta) \equiv I(|\psi_\theta\rangle; \theta)$ .

with  $L_{\rho_\theta}$  the symmetric logarithmic derivative (SLD):

$$\frac{\partial \rho_\theta}{\partial \theta} = \frac{L_{\rho_\theta} \rho_\theta + \rho_\theta L_{\rho_\theta}}{2}. \quad (3.3)$$

On the other hand the physical approach is based on the application of the Cramér-Rao theorem to a family of distributions popping up in quantum mechanics. We introduce a quantum statistical model  $\{p_\theta(\xi)\}$  by using a density matrix  $\rho_\theta$  and a POVM  $\{M_\xi\}$ :  $\{p_\theta(\xi) = \text{tr}[\rho_\theta M_\xi]\}$ . Then the QFI is shown to be equal to the maximal FI when considering all the possible POVMs:

**Theorem 3.2 — QFI AS THE MAXIMAL FI.**

The QFI (3.2) is equal to

$$I(\rho_\theta; \theta) = \max_{\{M_\xi\} \in \text{POVM}} J(p_\theta(\xi); \theta) \quad (3.4)$$

with

$$J(p_\theta; \theta) = \int d\xi p_\theta(\xi) \left( \frac{\partial \ln(p_\theta(\xi))}{\partial \theta} \right)^2 = \int d\xi \frac{1}{\text{tr}[\rho_\theta M_\xi]} \text{tr} \left[ \frac{\partial \rho_\theta}{\partial \theta} M_\xi \right]^2. \quad (3.5)$$

### 3.2.2 Proof of the QCRB I: Mathematical approach

We start with the mathematical approach based on Helstrom’s work as formalized by Holevo. The idea of the proof is to define a quantum analogue of the classical statistics (non-commutative statistics) and then to derive the analogue of the Cramér-Rao theorem following the proof of the classical Cramér-Rao theorem. This proof is inspired by Holevo’s work [Holevo, 2001, 2011] and by [Hayashi, 2006], although with less mathematical rigour.

To transpose the proof of the Cramér-Rao theorem to the quantum world we need to define quantum analogues to our tools from statistics. While the basic object from statistics was the probability distribution, our basic object is the state. To extract information from the state we need a POVM, which will actually lead to non-commutativity. To prove the theorem we need the Cauchy–Schwarz inequality, which means that we will have to design a proper inner product.

In the classical version of the theorem the FI appeared to be the norm with respect to a suitable scalar product of the score function  $\frac{\partial \ln(p_\theta)}{\partial \theta}$ . We will need an equivalent of the score function in the quantum case, with the difficulty of taking the derivative of the logarithm of the state. This problem can be solved using the SLD. Let  $X$  be a bounded operator on  $\mathcal{H}$ ,  $X \in \mathcal{B}(\mathcal{H})$ . We define the linear functional  $\mathcal{L}_{\rho_\theta}$  by

$$\mathcal{L}_{\rho_\theta}(X) := \text{tr} \left[ \frac{\partial \rho_\theta}{\partial \theta} X \right]. \quad (3.6)$$

We now define the scalar product fitting our purpose (there are several possible choices<sup>2</sup>). We want this scalar product to depend on  $\rho_\theta$ . We choose the completely symmetric inner product  $\langle Y, X \rangle_{\rho_\theta} := \frac{1}{2} \text{tr}[\rho_\theta(XY + YX)]$ , which leads to the tightest bound in the scalar case. By the cyclic permutation invariance of the trace we have  $\langle Y, X \rangle_{\rho_\theta} = \frac{1}{2} \text{tr}[(\rho_\theta X + X \rho_\theta)Y] = \frac{1}{2} \text{tr}[(\rho_\theta Y + Y \rho_\theta)X]$ .

Equipped with this scalar product we can use the Riesz-Frechet lemma that ensures that there exists an operator  $L_{\rho_\theta}$  such that

$$\mathcal{L}_{\rho_\theta}(X) = \langle L_{\rho_\theta}, X \rangle_{\rho_\theta}. \quad (3.7)$$

---

<sup>2</sup>A general form for  $\langle A, B \rangle_\sigma$  which respects the symmetry between  $A$  and  $B$  is given by  $\text{tr}[A(\int \sigma^\lambda B \sigma^{1-\lambda} p(d\lambda))]$  with  $p$  a probability distribution on  $[0, 1]$ . Different choices of the scalar product lead to different quantum generalizations of the Cramér-Rao theorem. This result is actually quite important and can prove useful in specific contexts, with the multi-parameter case as a primary example. One of the first examples for the use of a different scalar product than the symmetric one was investigated in [Yuen and Lax, 1973].

Using this inner product we can write the effect of the functional  $\mathcal{L}$  as

$$\mathcal{L}_{\rho_\theta}(X) = \frac{1}{2} \operatorname{tr}[(\rho_\theta L_{\rho_\theta} + L_{\rho_\theta} \rho_\theta)X] , \quad (3.8)$$

and by comparing with Eq. (3.6) we obtain:

$$\frac{\partial \rho_\theta}{\partial \theta} = \frac{\rho_\theta L_{\rho_\theta} + L_{\rho_\theta} \rho_\theta}{2} . \quad (3.9)$$

The operator  $L_{\rho_\theta}$  is called the symmetric logarithmic derivative (SLD) of  $\rho_\theta$ , or the quantum score function in reference to the score function defined as the derivative of the log-likelihood. Using the SLD we can write the derivative of the expectation value of any bounded operator  $X$  as

$$\frac{\partial}{\partial \theta} \operatorname{tr}[\rho_\theta X] = \operatorname{tr} \left[ \frac{\partial \rho_\theta}{\partial \theta} X \right] = \langle L_{\rho_\theta}, X \rangle_{\rho_\theta} . \quad (3.10)$$

where the subscript  $\rho_\theta$  in the expectation value expresses the dependence of the scalar product on  $\rho_\theta$ .

Now that we have a scalar product and a way to define score functions we have a look at the POVM, the estimator and the different conditions needed to carry out the proof. As we did in the classical case we will derive the bound for the most general case where one seeks to estimate a function  $g(\theta)$  of  $\theta$ . Overall, the estimation process is composed of (i) a POVM  $\{M_\xi\}$  with real outcomes  $\xi$  and (ii) a classical estimator  $\hat{g}_{\text{est}}$ . The POVM along with the state gives rise to the probability distribution  $p_\theta(\xi) := \operatorname{tr}[\rho_\theta M_\xi]$ . Furthermore we assume the regularity condition to hold:

$$\int d\xi \frac{\partial}{\partial \theta} p_\theta(\xi) = \frac{\partial}{\partial \theta} \int d\xi p_\theta(\xi) . \quad (3.11)$$

The unbiasedness condition reads  $E[\hat{g}_{\text{est}}] = g(\theta)$  with  $E[\hat{g}_{\text{est}}] = \int d\xi p_\theta(\xi) \hat{g}_{\text{est}}$ . Taking the derivative with respect to  $\theta$  we obtain

$$\frac{\partial}{\partial \theta} E[\hat{g}_{\text{est}}] = \frac{\partial}{\partial \theta} \int d\xi p_\theta(\xi) \hat{g}_{\text{est}} = \int d\xi \frac{\partial}{\partial \theta} p_\theta(\xi) \hat{g}_{\text{est}} , \quad (3.12)$$

where, in order to interchange the integral and the derivative, we used the regularity condition. Using the SLD, the definition of the probability distribution and the unbiasedness condition, we can rewrite this as

$$\frac{\partial g(\theta)}{\partial \theta} = \langle L_{\rho_\theta}, X_M \rangle_{\rho_\theta} , \quad (3.13)$$

with  $X_M := \int d\xi M_\xi \hat{g}_{\text{est}}$ . The regularity condition expressed as  $E[\frac{\partial \ln(p_\theta(\xi))}{\partial \theta}] = 0$  can also be written as

$$\int d\xi \frac{\partial \operatorname{tr}[\rho_\theta M_\xi]}{\partial \theta} = \operatorname{tr} \left[ \int d\xi \frac{\partial \rho_\theta M_\xi}{\partial \theta} \right] = 0 . \quad (3.14)$$

Using Eq. (3.10) the regularity condition takes eventually the form

$$0 = \left\langle L_{\rho_\theta}, \int d\xi M_\xi g(\theta) \right\rangle_{\rho_\theta} . \quad (3.15)$$

Subtracting Eq. (3.15) from Eq. (3.13) and taking the square of the resulting equation we obtain

$$\left| \frac{\partial g(\theta)}{\partial \theta} \right|^2 = \left| \langle L_{\rho_\theta}, \tilde{X}_M \rangle_{\rho_\theta} \right|^2 , \quad (3.16)$$

with  $\tilde{X}_M := \int d\xi M_\xi (\hat{g}_{\text{est}} - g(\theta))$ . We apply the Cauchy–Schwarz inequality to obtain

$$\langle L_{\rho_\theta}, L_{\rho_\theta} \rangle_{\rho_\theta} \langle \tilde{X}_M, \tilde{X}_M \rangle_{\rho_\theta} \geq \left| \frac{\partial g(\theta)}{\partial \theta} \right|^2. \quad (3.17)$$

This equation is pretty similar to the Cramér-Rao bound but with the difference that the variance of the estimator is replaced by  $\langle \tilde{X}_M, \tilde{X}_M \rangle_{\rho_\theta}$ . Using that the estimator is unbiased, we can write its variance as

$$\text{Var}[\hat{g}_{\text{est}}] = \text{tr} \left[ \rho_\theta \int d\xi M_\xi \hat{g}_{\text{est}}^2 \right] - g(\theta)^2. \quad (3.18)$$

We have to find a way know to connect  $\langle \tilde{X}_M, \tilde{X}_M \rangle_{\rho_\theta}$  to  $\text{Var}[\hat{g}_{\text{est}}]$ . To do so we use the following inequality<sup>3</sup>:

$$\left( \int d\xi M_\xi f(\xi) \right)^2 \geq \int d\xi M_\xi f(\xi)^2. \quad (3.19)$$

Replacing  $f(\xi)$  by  $\hat{g}_{\text{est}}$  we obtain  $\text{tr}[\rho_\theta \int d\xi M_\xi \hat{g}_{\text{est}}^2] \geq \text{tr}[\rho_\theta (\int d\xi M_\xi \hat{g}_{\text{est}})^2]$ , which leads to

$$\text{Var}[\hat{g}_{\text{est}}] \geq \langle \tilde{X}_M, \tilde{X}_M \rangle_{\rho_\theta}. \quad (3.20)$$

Combining Eq. (3.17) and Eq. (3.20) we eventually obtain the quantum Cramér-Rao bound

$$\text{Var}[\hat{g}_{\text{est}}] \geq \frac{\left| \frac{\partial}{\partial \theta} \text{E}[\hat{g}_{\text{est}}] \right|^2}{I(\rho_\theta; \theta)}, \quad (3.21)$$

with  $I(\rho_\theta; \theta) := \langle L_{\rho_\theta}, L_{\rho_\theta} \rangle_{\rho_\theta} = \text{tr}[\rho_\theta L_{\rho_\theta}^2]$  the QFI, which is the norm of the quantum score function.

### 3.2.3 Physical proof

In the mathematical derivation of the quantum Cramér-Rao bound what we gain in rigour we lose in intuition, especially due to the introduction of the inner product  $\langle \cdot, \cdot \rangle_{\rho_\theta}$ . By using a two-step procedure one can go back to the usual Hilbert-Schmidt inner product. This approach was first studied in [Braunstein and Caves, 1994]. Here the authors were interested in generalizing the concept of statistical distance (also called distinguishability metric) between states originally proposed in [Wootters, 1981]. The idea of Wootters was to propose a distance between states based on a physical operational meaning. He defines a statistical distance by counting the number of distinguishable states lying between two states when measured with the same measurement. Interestingly he showed that for pure states the usual distance in  $\mathcal{H}$  defined as  $\arccos(|\langle \phi | \psi \rangle|^2)$ , sometimes called quantum angle, is equal to the statistical distance.

One of the biggest advantages of the two-step derivation is that it allows to identify in a simple fashion the condition for saturating the bound. It also leads to the definition of the QFI as the maximal FI when optimizing over the set of all POVMs. Due to its importance we will have a look at how the proof works.

<sup>3</sup>We can show this inequality as follows: Define  $F(\xi) := f(\xi) - \int d\xi M_\xi$ . Using the positivity of the element of the POVM we have  $\int d\xi F(\xi) M_\xi F(\xi) = \int d\xi F(\xi) \sqrt{M_\xi} \sqrt{M_\xi} F(\xi)$ . By rearranging the terms and using the Hermiticity of the element of the POVM we obtain  $\int d\xi F(\xi) M_\xi F(\xi) = \int d\xi (F(\xi) \sqrt{M_\xi}) (F(\xi) \sqrt{M_\xi})^\dagger$ , which is clearly positive:  $\int d\xi F(\xi) M_\xi F(\xi) \geq 0$ . Inserting the definition of  $F(\xi)$  into  $\int d\xi F(\xi) M_\xi F(\xi)$  we obtain Eq. (3.19).

## Box 5: A different figure of merit

In [Braunstein and Caves, 1994] the authors do neither use directly the variance of the estimator as a figure of merit nor the MSE. Instead they define a measure of statistical deviation (their cost function) as

$$\delta\theta := \frac{\hat{\theta}_{\text{est}}}{\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta} - \theta, \quad (3.22)$$

where the division of  $\hat{\theta}_{\text{est}}$  by  $\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta$  is there to rescale locally the estimator. The corresponding figure of merit (risk function) is defined as  $\mathbb{E}[(\delta\theta)^2]$  (up to a prefactor  $n$  because they work with  $n$ -samples and are not interested in the natural factor  $n$  due to the additivity of the FI).

The expectation value of  $(\delta\theta)^2$  is equal to

$$\mathbb{E}[(\delta\theta)^2] = \theta^2 + \frac{\mathbb{E}[\hat{\theta}_{\text{est}}^2]}{(\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta)^2} - 2\theta \frac{\mathbb{E}[\hat{\theta}_{\text{est}}]}{\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta}. \quad (3.23)$$

Using the Cramér-Rao bound in the form of Eq. (1.54) we obtain

$$\frac{\mathbb{E}[\hat{\theta}_{\text{est}}^2]}{(\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta)^2} \geq \frac{1}{J(p_\theta; \theta)} + \frac{\mathbb{E}[\hat{\theta}_{\text{est}}]^2}{(\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta)^2}. \quad (3.24)$$

Combining the last two equations leads to

$$\mathbb{E}[(\delta\theta)^2] \geq \frac{1}{J(p_\theta; \theta)} + \theta^2 + \left( \frac{\mathbb{E}[\hat{\theta}_{\text{est}}]}{\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta} \right)^2 - 2\theta \frac{\mathbb{E}[\hat{\theta}_{\text{est}}]}{\partial \mathbb{E}[\hat{\theta}_{\text{est}}]/\partial\theta}. \quad (3.25)$$

The right hand side is equal to  $\frac{1}{J(p_\theta; \theta)} + \mathbb{E}[(\delta\theta)]^2$ , which shows *in fine* that  $\mathbb{E}[(\delta\theta)^2]$  is bounded by the inverse of the FI:

$$\mathbb{E}[(\delta\theta)^2] \geq \frac{1}{J(p_\theta; \theta)} + \mathbb{E}[(\delta\theta)]^2 \geq \frac{1}{J(p_\theta; \theta)}. \quad (3.26)$$

The starting point for this derivation is to generate a family of probability distributions  $p_\theta$  using the state  $\rho_\theta$  and a POVM  $\{M_\xi\}$ :  $p_\theta := \text{tr}[\rho_\theta M_\xi]$ . Having a parametrized distribution we can calculate the corresponding FI,

$$J(p_\theta; \theta) = \int d\xi \frac{1}{\text{tr}[\rho_\theta M_\xi]} \text{tr} \left[ \frac{\partial \rho_\theta}{\partial \theta} M_\xi \right]^2. \quad (3.27)$$

Thanks to the Cramér-Rao theorem we know that the variance of unbiased estimators of  $\theta$  is bounded by the inverse of  $J(p_\theta; \theta)$ . But for different POVMs we have different values of the FI, which leads to different bounds. To find the most informative bound (*i.e.* the most tight) we need

to maximize the FI over the POVMs.

We introduce again the symmetric logarithmic derivative  $L_{\rho_\theta}$  defined implicitly through the relation

$$\frac{\partial \rho_\theta}{\partial \theta} = \frac{L_{\rho_\theta} \rho_\theta + \rho_\theta L_{\rho_\theta}}{2}. \quad (3.28)$$

Notice that since  $\rho_\theta$  is Hermitian, as well as its derivative, we have  $\rho_\theta L_{\rho_\theta}^\dagger + L_{\rho_\theta}^\dagger \rho_\theta = L_{\rho_\theta} \rho_\theta + \rho_\theta L_{\rho_\theta}$ , which shows that the SLD is also Hermitian:  $L_{\rho_\theta}^\dagger = L_{\rho_\theta}$ . Using the implicit definition of the SLD (3.3) we can rewrite the FI as

$$I(p_\theta; \theta) = \int d\xi \frac{1}{\text{tr}[\rho_\theta M_\xi]} \text{tr}[(L_{\rho_\theta} \rho_\theta + \rho_\theta L_{\rho_\theta}) M_\xi / 2]^2. \quad (3.29)$$

We start a chain of inequalities to bound the Fisher information:

1. The first inequality uses  $\Re\{x\}^2 \leq |x|^2$  where  $\Re$  denotes the real part. Using the cyclic property of the trace we have  $\text{tr}[(L_{\rho_\theta} \rho_\theta + \rho_\theta L_{\rho_\theta}) M_\xi / 2] = \Re\{\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi]\}$ , which leads to the inequality

$$\text{tr}[(L_{\rho_\theta} \rho_\theta + \rho_\theta L_{\rho_\theta}) M_\xi / 2] \leq |\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi]|. \quad (3.30)$$

Equality is reached for  $\Im\{\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi]\} = 0$ , for further reference labelled (i), where  $\Im$  denotes the imaginary part. Notice that this is not one equation but a set of equations, that should be true for all  $\xi$ .

2. The second inequality is obtained through the Cauchy–Schwarz inequality for the Hilbert–Schmidt scalar product:

$$|\text{tr}[A^\dagger B]|^2 \leq \text{tr}[A^\dagger A] \text{tr}[B^\dagger B]. \quad (3.31)$$

Using the positivity of the density matrix and of the POVM elements we decompose the density matrix as  $\rho_\theta = \rho_\theta^{1/2} \rho_\theta^{1/2}$  with  $(\rho_\theta^{1/2})^\dagger = \rho_\theta^{1/2}$ . In the same way we decompose the elements of the POVM as  $M_\xi = M_\xi^{1/2} M_\xi^{1/2}$  with  $(M_\xi^{1/2})^\dagger = M_\xi^{1/2}$ . Using the cyclic permutation property of the trace we obtain

$$\frac{|\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi]|^2}{\text{tr}[\rho_\theta M_\xi]} = \left| \text{tr} \left[ \frac{\rho_\theta^{1/2} M_\xi^{1/2}}{\sqrt{\text{tr}[\rho_\theta M_\xi]}} M_\xi^{1/2} L_{\rho_\theta} \rho_\theta^{1/2} \right] \right|^2. \quad (3.32)$$

Applying the Cauchy–Schwarz inequality to the right hand side with  $A = \frac{M_\xi^{1/2} \rho_\theta^{1/2}}{\sqrt{\text{tr}[\rho_\theta M_\xi]}}$  and  $B = M_\xi^{1/2} L_{\rho_\theta} \rho_\theta^{1/2}$  we find

$$\frac{|\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi]|^2}{\text{tr}[\rho_\theta M_\xi]} \leq \frac{\text{tr}[\rho_\theta M_\xi]}{\text{tr}[\rho_\theta M_\xi]} \text{tr}[M_\xi L_{\rho_\theta} \rho_\theta L_{\rho_\theta}] = \text{tr}[M_\xi L_{\rho_\theta} \rho_\theta L_{\rho_\theta}]. \quad (3.33)$$

The inequality is saturated for  $M_\xi^{1/2} \rho_\theta^{1/2} = \alpha_\xi M_\xi^{1/2} L_{\rho_\theta} \rho_\theta^{1/2}$  with  $\alpha_\xi \in \mathbf{C}$  (we can absorb the term  $\sqrt{\text{tr}[\rho_\theta M_\xi]}$  in  $\alpha_\xi$ ), labelled for further reference (ii). Here it is important to note that the factor of proportionality can depend on  $\xi$  (again here we have a set of equations and not just one equation).

Combining inequalities (3.30) and (3.33) we arrive at

$$J(p_\theta; \theta) \leq \int d\xi \operatorname{tr}[M_\xi L_{\rho_\theta} \rho_\theta L_{\rho_\theta}] . \quad (3.34)$$

By changing the order of the trace and the integral and using  $\int d\xi M_\xi = \mathcal{I}$  we have

$$\int d\xi \operatorname{tr}[M_\xi L_{\rho_\theta} \rho_\theta L_{\rho_\theta}] = \operatorname{tr}\left[\int d\xi M_\xi L_{\rho_\theta} \rho_\theta L_{\rho_\theta}\right] = \operatorname{tr}[\rho_\theta L_{\rho_\theta}^2] . \quad (3.35)$$

Eventually we found an upper bound to the FI

$$J(p_\theta; \theta) \leq I(\rho_\theta; \theta) , \quad (3.36)$$

with the QFI  $I(\rho_\theta; \theta) = \operatorname{tr}[\rho_\theta L_{\rho_\theta}^2]$ . To arrive at Theorem 3.2 we should show that the inequality is actually tight and can be saturated. This will be done in Sec. 3.4.1..

### 3.3 Quantum Fisher Information

The QFI plays a central role in q-pet theory since it provides a tight bound on the estimation of parameters. We will now derive an explicit formula for it, review its properties and study briefly the link to the geometry of quantum states.

#### 3.3.1 Closed form for the QFI

Both forms of the QFI that we have presented here are still abstract. One is based on the SLD operator which we defined only implicitly, and the other is defined through a maximization procedure. For practical reasons it is important to design a more explicit formula for the QFI.

##### General case

In this section we present a more explicit form for the QFI. A proper treatment of this question should take into consideration the possibility that the density matrix  $\rho$  does not have full rank. We follow the derivation of [Jing et al., 2014].

Let  $\{\rho_\theta\}$  be a family of density matrices on a Hilbert space  $\mathcal{H}$  of dimension  $d$ . The spectral decomposition of  $\rho_\theta$  gives

$$\rho_\theta = \sum_{i=1}^r p_i |\psi_i\rangle \langle \psi_i| , \quad (3.37)$$

where  $r$  is the rank of  $\rho_\theta$  and  $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ . We assume furthermore that the rank of  $\rho_\theta$  does not change with  $\theta$ . We extend the set  $\{|\psi_i\rangle\}_{0 \leq i \leq r}$  to  $\{|\psi_i\rangle\}_{0 \leq i \leq d}$  by a Gram-Schmidt procedure in order to get a basis of  $\mathcal{H}$ , and define the support of  $\rho_\theta$  as  $\operatorname{supp}(\rho_\theta) = \{|\psi_i\rangle \langle \psi_i|\}_{0 \leq i \leq r}$ .

To get to an explicit form of the QFI we mainly need to work out the SLD. In the eigenbasis of  $\rho_\theta$  we have

$$\langle \psi_i | \frac{\partial \rho_\theta}{\partial \theta} | \psi_j \rangle = \frac{1}{2} (p_i + p_j) L_{\theta, ij} , \quad (3.38)$$

with  $L_{\theta, ij} = \langle \psi_i | L_{\rho_\theta} | \psi_j \rangle$ . By differentiating  $\langle \psi_i | \psi_j \rangle = \delta_{ij}$  with respect to  $\theta$  we obtain  $\frac{\partial}{\partial \theta} \langle \psi_i | \psi_j \rangle = \langle \partial_\theta \psi_i | \psi_j \rangle + \langle \psi_i | \partial_\theta \psi_j \rangle = 0$ , leading to

$$\langle \partial_\theta \psi_i | \psi_j \rangle = -\langle \psi_i | \partial_\theta \psi_j \rangle . \quad (3.39)$$

Using this result the left hand side of Eq. (3.38) reads:

$$\langle \psi_i | \frac{\partial \rho_\theta}{\partial \theta} | \psi_j \rangle = \delta_{ij} \frac{\partial p_i}{\partial \theta} + (p_j - p_i) \langle \psi_i | \partial_\theta \psi_j \rangle. \quad (3.40)$$

Finally the elements of the SLD in the eigenbasis of  $\rho_\theta$  are given by

$$\frac{L_{\theta,ij}}{2} = \frac{\delta_{ij}}{p_i + p_j} \frac{\partial p_i}{\partial \theta} + \frac{p_j - p_i}{p_i + p_j} \langle \psi_i | \partial_\theta \psi_j \rangle, \quad (3.41)$$

and we set for convenience  $L_{\theta,ij} = 0$  for  $\{i, j\}$  with both  $i > r$  and  $j > r$ .

Using the definition of the QFI which reads  $I(\rho_\theta; \theta) = \sum_{i=1}^r \sum_{k=1}^d p_i L_{\theta,ik} L_{\theta,ki}$ , we obtain

$$I(\rho_\theta; \theta) = 4 \sum_{i=1}^r \sum_{k=1}^d \frac{p_i}{p_i + p_k} \left| \langle \psi_i | \frac{\partial \rho_\theta}{\partial \theta} | \psi_j \rangle \right|^2. \quad (3.42)$$

Inserting the explicit form of the elements of the SLD into it and using Eq. (3.39) we obtain

$$I(\rho_\theta; \theta) = \sum_{i=1}^r \frac{1}{p_i} \left( \frac{\partial p_i}{\partial \theta} \right)^2 + 4 \sum_{i=1}^r \sum_{k=1}^d p_i \frac{(p_i - p_k)^2}{(p_i + p_k)^2} |\langle \psi_k | \partial_\theta \psi_i \rangle|^2. \quad (3.43)$$

We could stop here as this expression of the QFI depends in a rather simple way on the eigenvalues and eigenvectors of  $\rho_\theta$ . But a closer look to the expression reveals that we actually use vectors out of the support of  $\rho_\theta$ . To avoid this we split the sum over  $k$  into two parts  $\sum_{k=1}^d \rightarrow \sum_{k=1}^r + \sum_{k=r+1}^d$ :

$$I(\rho_\theta; \theta) = \sum_{i=1}^r \frac{1}{p_i} \left( \frac{\partial p_i}{\partial \theta} \right)^2 + 4 \sum_{i,k=1}^r p_i \frac{(p_i - p_k)^2}{(p_i + p_k)^2} |\langle \psi_k | \partial_\theta \psi_i \rangle|^2 + 4 \sum_{i=1}^r \sum_{k=r+1}^d p_i |\langle \psi_k | \partial_\theta \psi_i \rangle|^2. \quad (3.44)$$

Using the decomposition of the identity  $\mathcal{I} = \sum_{i=1}^d |\psi_i\rangle\langle\psi_i| = \sum_{i=1}^r |\psi_i\rangle\langle\psi_i| + \sum_{i=r+1}^d |\psi_i\rangle\langle\psi_i|$  we get  $\sum_{i=r+1}^d |\psi_i\rangle\langle\psi_i| = \mathcal{I} - \sum_{i=1}^r |\psi_i\rangle\langle\psi_i|$ . This leads to

$$4 \sum_{i=1}^r \sum_{k=r+1}^d p_i |\langle \psi_k | \partial_\theta \psi_i \rangle|^2 = 4 \sum_{i=1}^r p_i \langle \partial_\theta \psi_i | \partial_\theta \psi_i \rangle - 4 \sum_{i,k=1}^r p_i |\langle \psi_k | \partial_\theta \psi_i \rangle|^2. \quad (3.45)$$

Using again Eq. (3.39) in the form  $|\langle \psi_k | \partial_\theta \psi_i \rangle|^2 = |\langle \psi_i | \partial_\theta \psi_k \rangle|^2$ , we obtain finally for the QFI expressed in the eigenbasis of the density matrix

$$I(\rho_\theta; \theta) = \sum_{i=1}^r \frac{1}{p_i} \left( \frac{\partial p_i}{\partial \theta} \right)^2 + 4 \sum_{i=1}^r p_i \langle \partial_\theta \psi_i | \partial_\theta \psi_i \rangle - 8 \sum_{i,k=1}^r \frac{p_i p_k}{p_i + p_k} |\langle \psi_k | \partial_\theta \psi_i \rangle|^2. \quad (3.46)$$

### Pure states

For pure states the QFI simplifies drastically. To find the QFI of a state  $|\psi_\theta\rangle$  we can directly use the formula (3.46). By definition a pure state has rank one, say  $r = 1$  and  $p_1 = 1$ . The first term in the QFI vanishes. This is to be expected: this term corresponds to the classical contribution arising from the classical mixing in a mixed state. The sum of the second and last term, which corresponds to the quantum contribution, gives

$$I(|\psi_\theta\rangle; \theta) = 4 \left( \langle \partial_\theta \psi_\theta | \partial_\theta \psi_\theta \rangle - |\langle \psi_\theta | \partial_\theta \psi_\theta \rangle|^2 \right). \quad (3.47)$$

### Integral form for the QFI [Paris, 2009]

Eq. (3.3) which defines implicitly the SLD is a Lyapunov equation. Its formal solution in a basis-independent form is

$$L_{\rho_\theta} = 2 \int_0^\infty dx e^{-\rho_\theta x} \frac{\partial \rho_\theta}{\partial \theta} e^{-\rho_\theta x} . \quad (3.48)$$

Inserting this equation in  $\text{tr} \left[ \frac{\partial \rho_\theta}{\partial \theta} L_{\rho_\theta} \right]$  we obtain the QFI in a basis independent form

$$I(\rho_\theta ; \theta) = 2 \int_0^\infty dx \text{tr} \left[ \frac{\partial \rho_\theta}{\partial \theta} e^{-\rho_\theta x} \frac{\partial \rho_\theta}{\partial \theta} e^{-\rho_\theta x} \right] . \quad (3.49)$$

### 3.3.2 Geometrical considerations - QFI and the Bures distance

We have seen for probability distributions how we can design many distances which lead to the Fisher metric. The physical derivation of the quantum Cramér-Rao bound was investigated in the context of statistical distances. We now turn to the derivation of the link between the QFI and the Bures distance [Hübner, 1992, 1993] .

#### Closed expression for the Bures metric

Let us consider a density matrix  $\rho$  on  $\mathcal{H}$ . The Bures metric is defined as the square of the infinitesimal distance between  $\rho$  and  $\rho + d\rho$  *i.e.*  $d_B(\rho, \rho + d\rho)^2$ . We introduce the Hermitian operator  $A(t) = (\rho^{1/2}(\rho + t d\rho)\rho^{1/2})^{1/2}$ . Our goal is to obtain the value of  $d_B(\rho, \rho + t d\rho)^2$  up to second order in  $t$ . With our notation we have

$$d_B(\rho, \rho + t d\rho)^2 = 2 - 2 \text{tr}[A(t)] , \quad (3.50)$$

and especially the Bures metric is equal to  $2 - 2 \text{tr}[A(1)]$ . We make the ansatz

$$d_B(\rho, \rho + t d\rho)^2 = t^2 g_{ij}(\rho) d\rho^i d\rho^j , \quad (3.51)$$

where  $\rho^i$  serves as a coordinate on the manifold of density matrices. By differentiating two times with respect to  $t$  and dividing by two in the previous two equations we get

$$g_{ij}(\rho) d\rho^i d\rho^j = - \frac{d^2}{dt^2} \text{tr}[A(t)] . \quad (3.52)$$

Since the left hand side of Eq. (3.52) is independent of  $t$ , we can set any value for  $t$  in the right hand side, especially  $t = 0$ . We would now like to find an expression for the second derivative of  $A(t)$ . To do so we differentiate two times  $A(t)^2$ : The first derivative gives (we use dots for derivative with respect to  $t$  to shorten the notation)

$$\left. \frac{dA(t)^2}{dt} \right|_{t=0} = \dot{A}(0)A(0) + A(0)\dot{A}(0) = \rho^{1/2} d\rho \rho^{1/2} , \quad (3.53)$$

where the last equation is due to  $A(t)^2 = \rho^2 + t\rho^{1/2} d\rho \rho^{1/2}$ . The second derivative gives

$$\left. \frac{d^2 A(t)^2}{dt^2} \right|_{t=0} = \ddot{A}(0)A(0) + A(0)\ddot{A}(0) + 2\dot{A}(0)\dot{A}(0) = 0 . \quad (3.54)$$

### Chapter 3. Quantum parameter estimation theory

Assuming that  $\rho$ , which is equal to  $A(0)$ , is invertible<sup>4</sup>, we can write

$$\ddot{A}(0) + A(0)\ddot{A}(0)A^{-1}(0) = -2\dot{A}(0)\dot{A}(0)A^{-1}(0). \quad (3.55)$$

Taking the trace and using the invariance under permutation we obtain

$$\text{tr}[\ddot{A}(0)] = -\text{tr}[\dot{A}(0)\dot{A}(0)A^{-1}(0)]. \quad (3.56)$$

We decompose the state in its eigenbasis  $\rho = \sum_{i=1}^r p_i |\psi_i\rangle\langle\psi_i|$ . By taking a matrix element on both sides of Eq. (3.53) we can express  $\dot{A}(0)$  as

$$\dot{A}_{ij}(0) = \frac{\sqrt{p_i}\sqrt{p_j}}{(p_i + p_j)} d\rho_{ij}, \quad (3.57)$$

with  $\dot{A}_{ij}(0) = \langle\psi_i|\dot{A}(0)|\psi_j\rangle$  and  $d\rho_{ij} = \langle\psi_i|d\rho|\psi_j\rangle$ . Using Eq. (3.56) and Eq. (3.57) we obtain

$$g_{ij}(\rho) d\rho^i d\rho^j = \text{tr}[\dot{A}(0)\dot{A}(0)A^{-1}(0)] \quad (3.58)$$

$$= \sum_{i,j,k} (\rho^{-1})_{ij} \dot{A}_{jk}(0) \dot{A}_{ki}(0) \quad (3.59)$$

$$= \sum_{i,k} \frac{1}{p_i} \dot{A}_{ik}(0) \dot{A}_{ki}(0) \quad (3.60)$$

$$= \sum_{i,k} \frac{p_k}{(p_i + p_k)^2} |d\rho_{ik}|^2. \quad (3.61)$$

We can further develop this quantity as

$$\sum_{i,k} \frac{p_k}{(p_i + p_k)^2} |d\rho_{ik}|^2 = \sum_i \frac{1}{4p_i} |d\rho_{ii}|^2 + \sum_{i \neq k} \frac{p_k}{(p_i + p_k)^2} |d\rho_{ik}|^2 \quad (3.62)$$

$$= \sum_i \frac{1}{4p_i} |d\rho_{ii}|^2 + \frac{1}{2} \left( \sum_{i \neq k} \frac{p_k}{(p_i + p_k)^2} |d\rho_{ik}|^2 + \sum_{i \neq k} \frac{p_i}{(p_i + p_k)^2} |d\rho_{ik}|^2 \right) \quad (3.63)$$

$$= \frac{1}{2} \left( \sum_i \frac{1}{2p_i} |d\rho_{ii}|^2 + \sum_{i \neq k} \frac{1}{p_i + p_k} |d\rho_{ik}|^2 \right) \quad (3.64)$$

$$= \frac{1}{2} \sum_{i,k} \frac{|d\rho_{ik}|^2}{p_i + p_k}. \quad (3.65)$$

Setting  $t = 1$  in Eq. (3.51) we finally obtain

$$d_B(\rho, \rho + d\rho)^2 = \frac{1}{2} \sum_{i,k} \frac{|d\rho_{ik}|^2}{p_i + p_k}. \quad (3.66)$$

---

<sup>4</sup>If  $\rho$  has some zero eigenvalues then it is not invertible in  $\mathcal{S}(\mathcal{H})$ . In this case we restrain the analysis to the support of  $\rho$ .

### Link between Bures distance and QFI

Consider the case where the state is explicitly parametrized by  $\theta$ . The tangent vector  $d\rho$  is then equal to  $d\theta \frac{d\rho_\theta}{d\theta}$  and we can write:

$$d_{\text{B}}(\rho, \rho + d\rho)^2 = d_{\text{B}}(\rho_\theta, \rho_{\theta+d\theta})^2. \quad (3.67)$$

We have  $d\rho_{ik} = d\theta \langle \psi_i | \frac{d\rho_\theta}{d\theta} | \psi_k \rangle$ , which by using Eq. (3.61) gives

$$d_{\text{B}}(\rho_\theta, \rho_{\theta+d\theta})^2 = \sum_{i,k} \frac{p_k}{(p_i + p_k)^2} \left| d\theta \langle \psi_i | \frac{d\rho_\theta}{d\theta} | \psi_k \rangle \right|^2, \quad (3.68)$$

which is the same expression up to a prefactor as the right hand side of Eq. (3.42) and we thus have

$$d_{\text{B}}(\rho_\theta, \rho_{\theta+d\theta})^2 = \frac{d\theta^2}{4} I(\rho_\theta; \theta). \quad (3.69)$$

Finally we found the exact link between the QFI and the Bures distance

$$I(\rho_\theta; \theta) = 4 \frac{d_{\text{B}}(\rho_\theta, \rho_{\theta+d\theta})^2}{d\theta^2}. \quad (3.70)$$

### 3.3.3 Properties of the QFI

#### Theorem 3.3 — ADDITIVITY OF THE QFI.

Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be two Hilbert spaces and  $\{\rho_{1,\theta}\}$  and  $\{\rho_{2,\theta}\}$  two families of density matrices with  $\rho_{1,\theta} \in \mathcal{S}(\mathcal{H}_1)$  and  $\rho_{2,\theta} \in \mathcal{S}(\mathcal{H}_2)$ . Then we have

$$I(\rho_{1,\theta} \otimes \rho_{2,\theta}; \theta) = I(\rho_{1,\theta}; \theta) + I(\rho_{2,\theta}; \theta). \quad (3.71)$$

#### Theorem 3.4 — MONOTONICITY OF QFI.

Let  $\mathcal{E}$  be a quantum channel independent of the parameter  $\theta$  acting on  $\mathcal{S}(\mathcal{H})$ , and consider a family of density matrices  $\{\rho_\theta\}$  with  $\rho_\theta \in \mathcal{S}(\mathcal{H})$ . Then we have

$$I(\mathcal{E}(\rho_\theta); \theta) \leq I(\rho_\theta; \theta). \quad (3.72)$$

For the extremal quantum channels  $\mathcal{U}$  (unitary channels) we have

$$I(\mathcal{U}(\rho_\theta); \theta) = I(U\rho_\theta U^\dagger; \theta) = I(\rho_\theta; \theta), \quad (3.73)$$

#### Theorem 3.5 — CONVEXITY OF THE QFI.

Let  $\{\rho_\theta\}$  and  $\{\sigma_\theta\}$  be two families of quantum states parametrized by  $\theta$ , where  $\rho_\theta, \sigma_\theta$  are elements of  $\mathcal{S}(\mathcal{H})$ . Then, for  $0 \leq \lambda \leq 1$  we have

$$I(\lambda\rho_\theta + (1-\lambda)\sigma_\theta; \theta) \leq \lambda I(\rho_\theta; \theta) + (1-\lambda) I(\sigma_\theta; \theta). \quad (3.74)$$

A proof of the additivity of the QFI can be found in [Ji et al., 2008]. The monotonicity of the metric was discussed by Petz [1996] and more recently in the book on geometry of quantum states by [Bengtsson and Zyczkowski, 2006]. A proof of the convexity of the QFI can be found in [Fujiwara, 2001b].

## 3.4 Locality and saturation of the bound

### 3.4.1 Saturation

We analyse the conditions for reaching the quantum Cramér-Rao bound [Barndorff-Nielsen and Gill, 2000]. A first note, which is from the mathematical side, is that we used again the condition of interchanging the derivation and the integration (regularity condition). While this appears explicitly in the mathematical proof it is hidden in the physical proof, but still it is necessary there because we use the Cramér-Rao theorem which holds only if the regularity condition also holds.

Let us now have a look on the two conditions (i) and (ii) arising in the physical proof. The first one reads  $\Im \{\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi]\} = 0$ . The second one reads  $M_\xi^{1/2} \rho_\theta^{1/2} = \alpha_\xi M_\xi^{1/2} L_{\rho_\theta} \rho_\theta^{1/2}$  with  $\alpha_\xi \in \mathbf{C}$ . In order to find a sufficient condition we will work on the eigenbasis of the SLD  $\{|l_{\theta,i}\rangle\}$  defined by the spectral decomposition of  $L_{\rho_\theta}$

$$L_{\rho_\theta} = \sum_{i=1}^l \lambda_i |l_{\theta,i}\rangle \langle l_{\theta,i}|. \quad (3.75)$$

Notice that while  $\rho_\theta$  has rank  $r$ ,  $L_{\rho_\theta}$  has a rank  $l$  with  $r \leq l \leq d$ . Thus we can express the eigenvectors of  $\rho_\theta$  in the basis  $\{|l_{\theta,i}\rangle\}$  as

$$|\psi_i\rangle = \sum_{a=1}^l \psi_{i,a} |l_{\theta,a}\rangle. \quad (3.76)$$

We write  $\rho_\theta$  and  $\rho_\theta^{1/2}$  in the new basis as

$$\rho_\theta = \sum_i^r \sum_{a,b}^l p_i \psi_{i,a} \psi_{i,b}^* |l_{\theta,a}\rangle \langle l_{\theta,b}|, \quad (3.77)$$

$$\rho_\theta^{1/2} = \sum_i^r \sum_{a,b}^l \sqrt{p_i} \psi_{i,a} \psi_{i,b}^* |l_{\theta,a}\rangle \langle l_{\theta,b}|. \quad (3.78)$$

The condition (i) is thus written as (henceforth the matrix elements are taken in the basis  $\{|l_{\theta,i}\rangle\}$ )

$$\sum_{i,j,k}^l \Im \{L_{\theta,ij} \rho_{\theta,jk} M_{\xi,ki}\} = 0, \quad (3.79)$$

and the condition (ii) reads

$$\sum_k^l M_{\xi,ik}^{1/2} \rho_{\theta,kj}^{1/2} = \sum_{k,n}^l \alpha_\xi M_{\xi,ik}^{1/2} L_{\theta,kn} \rho_{\theta,nj}^{1/2} \quad \forall 1 \leq i, j \leq l. \quad (3.80)$$

Inserting Eq. (3.75) we obtain

$$\sum_k^l M_{\xi,ik}^{1/2} \rho_{\theta,kj}^{1/2} = \sum_k^l \alpha_\xi M_{\xi,ik}^{1/2} \lambda_k \rho_{\theta,kj}^{1/2}. \quad (3.81)$$

Now we could think to set  $\alpha_\xi = \lambda_k^{-1}$  (more precisely to set  $\alpha_\xi = \delta_{\xi k} \lambda_k$ ) to solve the equation, but this will not be valid since  $\alpha_\xi$  should be a constant depending only on the value of  $\xi$ . A proper solution is found by enforcing the POVM elements to incorporate  $\delta_{\xi k}$ . We thus take

$$M_{\xi,ij} = m_\xi^2 \delta_{\xi i} \delta_{\xi j} \quad (3.82)$$

leading to

$$m_\xi \delta_{\xi i} \rho_{\theta,\xi j}^{1/2} = m_\xi \delta_{\xi i} \alpha_\xi \lambda_\xi \rho_{\theta,\xi j}^{1/2}, \quad (3.83)$$

which is satisfied by  $\alpha_\xi = 1/\lambda_\xi$ . We should check if this choice of POVM elements satisfies also the condition (i). By inserting this POVM into  $\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi]$  we obtain

$$\text{tr}[L_{\rho_\theta} \rho_\theta M_\xi] = L_{\rho_\theta,ij} \rho_{\theta,ik} M_{\xi,ki} \quad (3.84)$$

$$= \lambda_i \rho_{\theta,ik} \delta_{\xi k} \delta_{\xi i} m_\xi^2 \quad (3.85)$$

$$= \lambda_\xi \rho_{\theta,\xi\xi} m_\xi^2 \quad (3.86)$$

$$= \lambda_\xi m_\xi^2 \sum_i^r p_i |\psi_{i\xi}|^2, \quad (3.87)$$

which is a real quantity as required by condition (i). We are thus left with a sufficient condition on the POVM elements in order to saturate the quantum Cramer-Rao bound, namely we need a POVM realizing a von Neumann measurement of the eigenbasis of the SLD, which we can write as

$$\{M_{\theta,\xi}\} = \{|l_{\theta,\xi}\rangle \langle l_{\theta,\xi}|\}_{1 \leq \xi \leq l}. \quad (3.88)$$

The fact that there exists always a POVM saturating the inequality  $I(\rho_\theta; \theta) \geq J(\text{tr}[\rho_\theta M_\xi]; \theta)$  shows the validity of the theorem 3.2.

### 3.4.2 Locality

We just saw that the POVM which allows us to reach the QCRB may depend in general on the value of  $\theta$ . When this is the case it means that one should know the value of the parameter in order to be able to estimate it. This comes as an echo of the question of locality studied in Chapter 1. Having two proofs of the theorem in its quantum version, we will choose the most useful depending on the situation.

#### Locality of the quantum Fisher information

The locality of the QFI as defined in Eq. (3.4) arises as a consequence of the locality of the FI. If we have a look at Eq. (3.5) we can further specify in which sense the QFI is a local quantity. While in the classical case we were interested in the space of probability distributions, and locality was to be understood in this way, here locality arises in the space of density matrices. The definition of the QFI at a point  $\theta_0$  depends only on two quantities, the density matrix at this point  $\rho_{\theta_0}$  and its first derivative  $\frac{\partial \rho_\theta}{\partial \theta} |_{\theta_0}$ . Then the QFI is completely specified by the pair  $(\rho_{\theta_0}, \frac{\partial \rho_\theta}{\partial \theta} |_{\theta_0})$ .

#### Locality in the QCRB

We now study locality for the QCRB. We will focus on the physical proof. As already for the classical case, the most obvious argument for locality is that the inequality in the theorem is

## Box 6: Quantum vector parameter estimation

We have seen how in classical parameter estimation theory we can easily generalize results for scalar parameters to vector parameters. In q-pet the situation gets more complicated. We can still generalize the quantum Cramér-Rao theorem by providing a bound on the covariance matrix of any unbiased estimators of the vector parameter  $\boldsymbol{\theta} := \{\theta_i\}$ :

$$\text{Cov}[\hat{\boldsymbol{\theta}}_{\text{est}}] - \mathbf{I}(\boldsymbol{\mu}_{\boldsymbol{\theta}}; \boldsymbol{\theta})^{-1} \geq 0 . \quad (3.89)$$

The elements of the quantum Fisher information matrix  $\mathbf{I}(\boldsymbol{\mu}_{\boldsymbol{\theta}}; \boldsymbol{\theta})$  are defined as

$$(\mathbf{I}(\boldsymbol{\mu}_{\boldsymbol{\theta}}; \boldsymbol{\theta}))_{ij} := \left\langle L_{\rho_{\theta_j}}, L_{\rho_{\theta_i}} \right\rangle_{\rho_{\boldsymbol{\theta}}} , \quad (3.90)$$

with the SLD  $L_{\rho_{\theta_i}}$  defined implicitly as a solution of  $\frac{\partial \rho_{\boldsymbol{\theta}}}{\partial \theta_i} = \frac{L_{\rho_{\boldsymbol{\theta}}} \rho_{\boldsymbol{\theta}} + \rho_{\boldsymbol{\theta}} L_{\rho_{\boldsymbol{\theta}}}}{2}$ . The crucial difference with classical parameter estimation theory lies in the saturation of the bound. Let us work in the simple case where the QFI matrix is diagonal. The Cramér-Rao bound becomes thus a list of inequalities involving the diagonal elements of the QFI matrix,  $\left\langle L_{\rho_{\theta_i}}, L_{\rho_{\theta_i}} \right\rangle_{\rho_{\boldsymbol{\theta}}}$ .

Each inequality is saturated by using a POVM constructed with the eigenvectors of  $L_{\rho_{\theta_i}}$ . But to saturate *all* the inequalities at once we need the POVMs to commute, which is in general not the case! A solution, although not optimal, is to pick-up one of the POVM that saturates one of the inequality, and then calculate the FI matrix for this POVM.

Finally let us mention that in the vector parameter case the RLD is sometimes more informative (more tight) than the SLD.

based on a local quantity, the QFI. The discussion about the unbiasedness of the estimators in the QCRB is the same as the one for the CRB. We do not need to ask for global unbiasedness, it will be enough to have it locally. The condition for local unbiasedness reads

$$\mathbb{E}[\hat{\boldsymbol{\theta}}_{\text{est}}] \Big|_{\theta_0} = \int d\xi \text{tr}[\rho_{\theta_0} M_{\xi}] \hat{\boldsymbol{\theta}}_{\text{est}} = \boldsymbol{\theta}_0 , \quad (3.91a)$$

$$\frac{\partial \mathbb{E}[\hat{\boldsymbol{\theta}}_{\text{est}}]}{\partial \theta} \Big|_{\theta_0} = \int d\xi \text{tr} \left[ \frac{\partial \rho_{\boldsymbol{\theta}}}{\partial \theta} \Big|_{\theta_0} M_{\xi} \right] \hat{\boldsymbol{\theta}}_{\text{est}} = \mathbf{1} . \quad (3.91b)$$

Up to here nothing is very new in terms of locality. When previously the discussion focused on probability distributions, now we focus on states. Nevertheless there is a new ingredient, the POVM. In all our derivations we considered implicitly that the POVM was independent of the value of  $\theta$  such that when differentiating the probability distribution only the state was differentiated. Though, we ended the section of saturation of the bound by saying that a POVM which allows to reach the bound is based on the SLD (Eq. (3.88)), which in general depends on  $\theta$ . If this was the case, we should actually write the derivative of the probability distribution as

$$\frac{\partial p_{\boldsymbol{\theta}}(\xi)}{\partial \theta} \rightarrow \text{tr} \left[ \frac{\partial \rho_{\boldsymbol{\theta}}}{\partial \theta} M_{\xi} + \frac{\partial M_{\xi}}{\partial \theta} \rho_{\boldsymbol{\theta}} \right] .$$

Pursuing the calculation with this new equation will end with a modified version of the quantum

Cramér-Rao bound (see Box 7). So how do we have to interpret our result?

The answer emerges when having a closer look at the dependency of the POVM and is directly related to the problem of locality. Indeed when looking at Sec. 3.4.1 we see that all we need to saturate the inequality at the point  $\theta_0$  is that the POVM projects into the spectrum of the SLD at the point  $\theta_0$ ; we do not need any property defined in the vicinity of the point  $\theta_0$ . From an operational point of view, this means that in order to saturate the quantum Cramér-Rao bound at a point  $\theta_0$  we should use the fixed POVM

$$\{M_{\theta_0, \xi}\} = \{|l_{\theta_0, \xi}\rangle\langle l_{\theta_0, \xi}|\}_{1 \leq \xi \leq l}. \quad (3.92)$$

The analogue situation for the state would be to start the experiment not with  $\rho_\theta$  but directly with  $\rho_{\theta_0}$ . If one does this the probability distribution does not depend on the parameter anymore and there is nothing to estimate.

In practice, to saturate the bound when we have no prior knowledge on  $\theta$  we can use an adaptive procedure [Nagaoka, 2005; Fujiwara, 2006; Barndorff-Nielsen and Gill, 2000], which in its simplest form amounts to a two-step procedure. There one starts by using a small fraction of the available states to obtain a first estimate of  $\theta$ . Then one uses this estimate to construct the quasi-optimal POVM which *in fine* allows to saturate the bound. Like in classical parameter estimation theory, the problems of locality can be overcome because we work in the asymptotic regime.

## 3.5 Channel estimation

Up to here the fundamental physical object which contains the information has been the state of the system. As indicated by its name, in channel estimation it is not directly the state that depends on the parameter but rather a quantum channel. We will see in the next chapter that starting with a parameter dependent channel instead of starting with a parameter dependent state corresponds to a common situation in quantum metrology. Since this chapter is devoted to pure parameter estimation theory, and not to actual metrology, we will emphasize here the formal and mathematical aspects of channel estimation, deferring the corresponding metrological task to the next chapter.

Formally, channel estimation aims to answer the following question: Given a CPTP map  $\mathcal{E}_\theta$  which depends smoothly on a parameter  $\theta \in \Theta$ , how precisely can we estimate the value of  $\theta$ , *i.e.* what is the lower bound on the variance of unbiased estimators of  $\theta$ ? The resolution of this question does not require to develop a new formalism from scratch. In the same way as we mapped the problem of quantum state estimation to classical parameter estimation, we can map channel estimation to quantum state estimation. Indeed taking an initial state  $\rho$ , we can feed the quantum channel and obtain a parameter-dependent state  $\rho_\theta = \mathcal{E}_\theta(\rho)$ . We can then compute the QFI for this state. By proceeding in this way we end up with a new degree of freedom, the initial state. If we want to avoid this freedom, we can repeat our procedure, and maximize the QFI over the set of initial states to obtain a new figure of merit, the *channel QFI*, as we maximized the FI over the set of POVMs to obtain the QFI:

### Definition 3.2 — CHANNEL QFI.

For a given family  $\{\mathcal{E}_\theta\}$  of channels  $\mathcal{E}_\theta$  acting on  $\mathcal{S}(\mathcal{H})$  and parametrized by  $\theta \in \Theta$ , the channel QFI is defined as the maximal QFI over the set of input states:

$$C(\mathcal{E}_\theta; \theta) := \max_{\rho \in \mathcal{S}(\mathcal{H})} I(\mathcal{E}_\theta(\rho); \theta). \quad (3.95)$$

## Box 7: Parameter dependent POVM and sample space [Seveso et al., 2017]

We claimed that the QCRB gives the ultimate bound on the precision for the estimation of a parameter  $\theta$  by measuring a state  $\rho_\theta$ . We already know that there are some non-trivial assumptions to ensure this claim to be true. We emphasized the fact that the theorem holds only for unbiased estimators and that it is reachable in general in the asymptotic limit. Nevertheless there are some other assumptions that we did not stress.

An important hidden condition is the fact that all the information on  $\theta$  comes from  $\rho_\theta$ , which is not imposed by physical law but a matter of choice. We could equally consider that the POVM itself depends on the parameter:  $\{M_\xi\} \rightarrow \{M_\xi(\theta)\}$  [Seveso et al., 2017]. With such  $\theta$ -dependent POVM the FI is given by:

$$J(\text{tr}[\rho_\theta M_\xi(\theta)] ; \theta) = \int d\xi \left( \text{tr}[(\partial_\theta \rho_\theta) M_\xi(\theta)]^2 + \text{tr}[\rho_\theta (\partial_\theta M_\xi(\theta))]^2 \right. \\ \left. + 2 \text{tr}[(\partial_\theta \rho_\theta) M_\xi(\theta)] \text{tr}[\rho_\theta (\partial_\theta M_\xi(\theta))] \right) / \text{tr}[\rho_\theta M_\xi(\theta)] . \quad (3.93)$$

The first term in the integral gives back the usual FI which maximized over POVM would give the QFI. The two others terms are extra and come from the dependence of the POVM on  $\theta$ . Importantly this FI is not maximized over the POVM and is not an equivalent of the QFI.

A second way to introduce an extra dependence on the parameter is to consider a parameter dependent sample space:  $d\xi \rightarrow d\xi m_\theta(\xi)$ . Notice that this is at the level of classical parameter estimation, we do not need quantum mechanics to do that. With such parameter dependent measure the FI for  $p_\theta$  becomes

$$J^{m_\theta}(p_\theta; \theta) = \int d\xi m_\theta \frac{(\partial_\theta p_\theta(\xi))^2}{p_\theta} + \int d\xi p_\theta \frac{(\partial_\theta m_\theta(\xi))^2}{m_\theta} \\ + \int d\xi (\partial_\theta p_\theta(\xi)) (\partial_\theta m_\theta(\xi)) . \quad (3.94)$$

The result looks similar as if we started with a standard measure  $dx$  and considered the probability distribution  $p_\theta(\xi)m_\theta(\xi)$ . There is still a difference which becomes important when going to the quantum case: with parameter dependent sample space the resolution of the identity for the POVM is written  $\int d\xi m_\theta(\xi) M_\xi = \mathcal{I}$ , which results in a FI with a last term in Eq. (3.94) that vanishes.

It was shown in [Fujiwara, 2001a] that it is enough to maximize over pure input states:

$$C(\mathcal{E}_\theta; \theta) = \max_{|\psi\rangle \in \mathcal{H}} I(\mathcal{E}_\theta(|\psi\rangle\langle\psi|); \theta) . \quad (3.96)$$

In Fig. 3.1 we represented schematically the link between the different parameter estimation problems (from channel to state to probability distributions) and their respective figure of merit (from FI to QFI to channel QFI).

### 3.5.1 A first upper bound and QFI as a minimization procedure

While the estimation of some specific quantum channels (and among them especially unitary channels) has been carried out since the beginnings of the field, the formal problem of channel estimation is more recent, and can be traced back to [Sarovar and Milburn, 2006]. In this paper, the authors tried to find a figure of merit for the estimation of quantum channels. They based their work on the Kraus decomposition of CPTP maps. To a quantum channel  $\mathcal{E}_\theta$  we can associate a  $q$ -Kraus decomposition  $\{A_k(\theta)\}_{1 \leq k \leq q}$  such that the state  $\rho$  after the evolution becomes  $\mathcal{E}_\theta(\rho) = \sum_{k=1}^q A_k(\theta)\rho A_k(\theta)^\dagger$ . In general even if the initial state  $\rho$  is pure the output state will be mixed. If we want to use the formula (3.46) to compute the QFI we need first to diagonalize the state  $\mathcal{E}_\theta(\rho)$ , which is a task that cannot be performed in a general fashion, prohibiting us to write the QFI in a simple way.

To circumvent this issue the authors went back to the FI for the probability distribution  $p_\theta(\xi) = \text{tr}_1[\rho_\theta E_\xi]$ . We introduce the subscript 1 in the partial trace as the trick here is to use the Stinespring dilation theorem to write the state after evolution as  $\rho_\theta = \text{tr}_2[U(\theta)(\rho \otimes |a_0\rangle\langle a_0|)U(\theta)^\dagger]$ . By using the Cauchy–Schwarz inequality they showed that

$$J(\mu_\theta; \theta) \leq \Gamma_{\text{MS}} := 4 \text{tr} \left[ \sum_j^q \dot{A}_j^\dagger(\theta) \dot{A}_j(\theta) \rho \right]. \quad (3.97)$$

where the dot stands for the derivation with respect to  $\theta$ :  $\dot{A}_j(\theta) = \partial A_j(\theta)/\partial \theta$ . We emphasize that this is a bound for the FI and not for the QFI. The bound  $\Gamma_{\text{MS}}$  raises a problem related to the non-uniqueness of the Kraus representation. Indeed we know that we can go from one Kraus representation to another through unitary matrices. And these unitary matrices can depend on  $\theta$ . As a result we can construct a Kraus decomposition that makes the corresponding bound  $\Gamma_{\text{MS}}$  infinite!

It is then clear that one should not try to maximize the bound to get more insights. Going to the other direction it is natural to ask whether the bound is tight or not. It appears that one can identify the condition under which the upper bound is tight, *i.e.* the condition for which the upper bound equals the QFI. Still, in general the upper bound is not tight. We will see below that in order to clarify (but not solve) this problem one has to rely on a trick: channel extension.

### 3.5.2 Channel QFI and channel extensions [Fujiwara and Imai, 2008]

We now build a proper theory of channel estimation. To do so we will need different ingredients. The first of them, channel extension, appears in this section as a tool introduced to solve what seems a more mathematical than physical problem. The next chapter will show that this is not true: Extension plays an important role in ancillary assisted metrology.

#### Channel extension

Quantum channels are CPTP maps. Behind this tautology lies the basic idea for channel extension: being *completely* positive, we can naturally extend a channel as

$$\mathcal{E}_\theta^{(\mathcal{A})} := \mathcal{E}_\theta \otimes \mathcal{A}, \quad (3.98)$$

with  $\mathcal{A}$  an arbitrary  $\theta$ -independent channel acting on the operators of a second Hilbert space, called ancillary Hilbert space. This  $\mathcal{A}$ -extended channel can be split into

$$\mathcal{E}_\theta \otimes \mathcal{A} = (\mathcal{E}_\theta \otimes \text{Id}) \circ (\text{Id} \otimes \mathcal{A}) = (\text{Id} \otimes \mathcal{A}) \circ (\mathcal{E}_\theta \otimes \text{Id}). \quad (3.99)$$

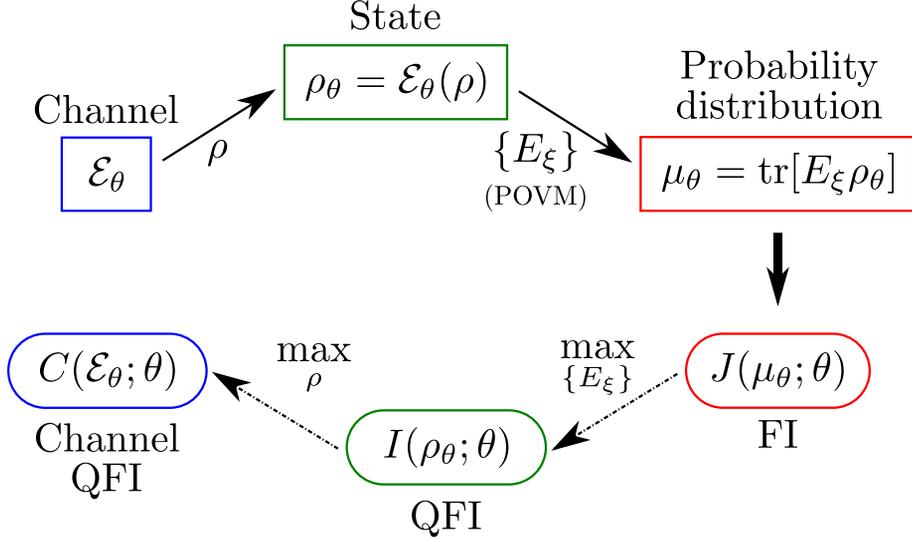


Figure 3.1: Diagram for quantum parameter estimation theory.

We can naturally ask what the impact of extension with regard to parameter estimation theory is. Importantly if we are interested in the QFI of the  $\mathcal{A}$ -extended channel, monotonicity of the QFI leads to

$$I((\mathcal{E}_\theta \otimes \mathcal{A})(\rho); \theta) = I((\text{Id} \otimes \mathcal{A}) \circ (\mathcal{E}_\theta \otimes \text{Id})(\rho); \theta) \leq I((\mathcal{E}_\theta \otimes \text{Id})(\rho); \theta), \quad (3.100)$$

since  $[\text{Id} \otimes \mathcal{A}]$  is  $\theta$ -independent. The equality is achieved when  $[\text{Id} \otimes \mathcal{A}]$  is a unitary channel, *i.e.* when  $\mathcal{A}$  itself is a unitary channel. It is therefore enough to consider the simplest of unitary channels, the identity channel. Since we are interested in optimizing the QFI, we consider  $\text{Id}$ -extended channels, *i.e.* with

$$\mathcal{E}_\theta^{\text{ext}} := \mathcal{E}_\theta \otimes \text{Id}, \quad (3.101)$$

and we call such extensions simply channel extensions. As we saw that it was difficult to make progress with the estimation of the original channel, we will work with the extended channel. We will also stop working with a specific fixed input state but rather look directly at the optimal case, meaning that we will focus on the channel QFI instead of the QFI. While this might look like a harder task, it can also result in a simplified problem because there is one degree of freedom less. In order to carry out this maximization procedure of the QFI we first need to express the QFI in a different way than the one presented in Sec. 3.3.1, since we saw that the need of diagonalizing the state is crippling there.

### QFI as minimization over $(\rho, q)$ -ensembles

We already saw how the freedom in the Kraus operator made the use of  $\Gamma_{\text{MS}}$  a delicate task. Ironically we will see how this freedom helps to solve the problem. But for the moment, it is not the freedom on the Kraus operators that we will use, but rather the freedom in the state preparation (see Sec. 3.3.1):

**Theorem 3.6 — QUANTUM FISHER INFORMATION AS A MINIMIZATION.**

The QFI  $I(\rho_\theta; \theta)$  can be expressed as a minimization over any of the  $(\rho_\theta, q)$ -decomposition, elements of the  $(\rho, q)$ -ensemble  $W_q(\theta)$  (with  $q \geq r$ ):

$$I(\rho_\theta; \theta) = 4 \min_{\{w_i\} \in W_q(\theta)} \operatorname{tr} \left[ \sum_{w_\theta} |\dot{w}_\theta\rangle \langle \dot{w}_\theta| \right]. \quad (3.102)$$

Notice that we could write this theorem with a minimization over the  $q \times q$  unitary matrices instead of minimizing over  $(\rho, q)$ -ensembles. By providing a reference  $(\rho_\theta, q)$ -decomposition we can generate all the  $(\rho_\theta, q)$ -ensembles with the right action of  $U$  on the  $(\rho_\theta, q)$ -decomposition.

**Channel extended QFI**

We have now all the ingredients to build a proper theory of channel parameter estimation. Our goal will be to find a closed expression for the extended channel QFI. To achieve this goal we will start from the QFI, in its minimization-based version, and then maximize over the input states  $\tilde{\rho} \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H})$ :

$$C(\mathcal{E}_\theta \otimes \operatorname{Id}; \theta) = \max_{\tilde{\rho} \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H})} I((\mathcal{E}_\theta \otimes \operatorname{Id})(\tilde{\rho}); \theta) = \max_{|\tilde{\psi}\rangle \in \mathcal{H} \otimes \mathcal{H}} I((\mathcal{E}_\theta \otimes \operatorname{Id})(|\tilde{\psi}\rangle \langle \tilde{\psi}|); \theta). \quad (3.103)$$

Notice that the same statement can be derived by looking at the upper bound  $\Gamma_{\text{MS}}$ . This is done using the linearity of the trace and the positivity of  $\sum_j^q \dot{A}_j^\dagger(\theta) \dot{A}_j(\theta)$ . As a result the  $\Gamma_{\text{MS}}$  is a concave function of the state  $\rho$ , *in fine* showing that it is a pure state — *i.e.* an element of the boundary  $\partial\mathcal{S}(\mathcal{H})$  of  $\mathcal{S}(\mathcal{H})$  — that maximizes the function. Our goal is thus to compute  $\max_{|\tilde{\psi}\rangle \in \mathcal{H} \otimes \mathcal{H}} I((\mathcal{E}_\theta \otimes \operatorname{Id})(|\tilde{\psi}\rangle \langle \tilde{\psi}|); \theta)$ .

The use of a  $q$ -Kraus decomposition  $\{A_j(\theta)\}_{1 \leq j \leq q}$  leads to a  $(\tilde{\sigma}_\theta, q)$ -decomposition of the term we seek to maximize:

$$\tilde{\sigma}_\theta = (\mathcal{E}_\theta \otimes \operatorname{Id})(|\tilde{\psi}\rangle \langle \tilde{\psi}|) = \sum_{j=1}^q (A_j(\theta) \otimes \mathcal{I}) |\tilde{\psi}\rangle \langle \tilde{\psi}| (A_j^\dagger(\theta) \otimes \mathcal{I}),$$

where  $\{(A_j(\theta) \otimes \mathcal{I})|\tilde{\psi}\rangle\}_{0 \leq j \leq 1}$  is a  $(\tilde{\sigma}_\theta, q)$ -decomposition (for simplicity we write the decomposition with non-normalized vectors instead of using the doublet "eigenvalue + eigenvector"). With the unitary transformation, we can go from a minimization over the  $(\tilde{\sigma}_\theta, q)$ -decomposition (given in theorem 3.6) to a minimization over the  $q$ -Kraus decomposition:

$$\min_{w_{\rho_\theta, q} \in \mathcal{W}_q(\theta)} \leftrightarrow \min_{U(\theta) \in U(q)} \leftrightarrow \min_{\mathcal{A}(\theta) \in \mathcal{A}_q(\theta)}.$$

Having a  $(\tilde{\sigma}_\theta, q)$ -decomposition we can use our minimization-based QFI (3.102). By splitting the trace as  $\operatorname{tr} \leftrightarrow \operatorname{tr}_1 \operatorname{tr}_2$  we obtain

$$I(\tilde{\sigma}_\theta; \theta) = 4 \min_{\mathcal{A}(\theta)} \operatorname{tr}_{\mathcal{H}} \left[ \sigma \sum_{j=1}^q (\dot{A}_j^\dagger(\theta) \dot{A}_j(\theta)) \right], \quad (3.104)$$

with  $\sigma = \operatorname{tr}_{\mathcal{H}} [|\tilde{\psi}\rangle \langle \tilde{\psi}|]$ . The mapping  $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H} \mapsto \sigma = \operatorname{tr}_{\mathcal{H}} [|\psi\rangle \langle \psi|] \in \mathcal{S}(\mathcal{H})$  being surjective, the maximization over the pure states in  $\mathcal{H} \otimes \mathcal{H}$  leads to a maximization over all the states in  $\mathcal{S}(\mathcal{H})$ . This is an essential point, as it allows to use a minimax theorem to invert the order of the maximization and minimization which allows to express the extended channel QFI as:

**Theorem 3.7 — ON CHANNEL QFI OF EXTENDED CHANNELS.**

For a one parameter family of quantum channels  $\{\mathcal{E}_\theta\}$  and for any natural number  $q$  such that  $q \geq \text{rank}(\mathcal{E}_\theta)$ , we have

$$C(\mathcal{E}_\theta \otimes \text{Id}; \theta) = 4 \min_{\mathcal{A}(\theta) \in \mathcal{A}_q} \left\| \sum_{j=1}^q \dot{A}_j^\dagger(\theta) \dot{A}_j(\theta) \right\|_\infty, \quad (3.105)$$

with  $\mathcal{A}(\theta) = \{A_j(\theta)\}_{1 \leq j \leq q}$  and where  $\|\bullet\|_\infty$  is the infinity norm<sup>5</sup> of  $\mathcal{H}$ .

### 3.5.3 Upper bound and purification based QFI

#### Upper bound as a corollary

Because we started with the extended channel we could use the minimax theorem and therefore prove the theorem. If we start with the original channel  $\mathcal{E}_\theta$  instead of the extended channel  $\mathcal{E}_\theta \otimes \text{Id}$  we cannot carry out the same procedure. The problem would be that starting with a maximization over the elements of  $\mathcal{H}$  (pure states), we would end up with

$$4 \max_{|\psi\rangle \in \mathcal{H}} \min_{\mathcal{A}(\theta)} \text{tr} \left[ |\psi\rangle\langle\psi| \sum_{j=1}^q \dot{A}_j^\dagger(\theta) \dot{A}_j(\theta) \right].$$

From there we cannot go further as the minimax theorem does not apply.

Still, by noticing that taking the maximum only over the pure states instead of taking it over all states can only lead to a lower value, the extended channel QFI is an upper bound for the original channel QFI:

$$C(\mathcal{E}_\theta(\rho); \theta) \leq 4 \min_{\mathcal{A}(\theta) \in \mathcal{A}_q} \left\| \sum_{j=1}^q \dot{A}_j^\dagger(\theta) \dot{A}_j(\theta) \right\|_\infty, \quad (3.107)$$

with  $\mathcal{A}(\theta) = \{A_j(\theta)\}_{1 \leq j \leq q}$ , and where  $\|\bullet\|_\infty$  is the infinity norm of  $\mathcal{H}$ . This bound is not tight in general. Then we see why channel extension answers the question raised in [Sarovar and Milburn, 2006]: To find an optimal figure of merit we need to consider extended channels, otherwise we can only get an upper bound for the channel QFI. This result has a deep meaning. It tells us that the actual figure of merit attached to a quantum channel is not the channel QFI, but rather the extended channel QFI.

#### Alternative method

A different upper bound was derived later in [Escher et al., 2011]. The method to derive it is also based on the Stinespring dilation theorem and leads to

$$I(\mathcal{E}_\theta(\rho); \theta) \leq 4(\text{tr}[\rho H_1] - \text{tr}[\rho H_2]^2), \quad (3.108)$$

with  $H_1 = \sum_i \dot{A}_i^\dagger \dot{A}_i$  and  $H_2 = i \sum_i \dot{A}_i^\dagger A_i$ .

<sup>5</sup>Consider the Schatten  $p$ -norm of the operator  $A$ :  $\|A\|_p = \text{tr}[(A^\dagger A)^{p/2}]^{1/p}$ . Then the infinity norm is equal to

$$\|A\|_\infty = \lim_{p \rightarrow \infty} \|A\|_p. \quad (3.106)$$

The authors showed that there always exists a Kraus representation such that the bound is tight, providing an alternative version of the QFI as a minimization

$$I(\mathcal{E}_\theta(\rho); \theta) = 4 \min_{\mathcal{A}(\theta) \in \mathcal{A}_q} (\text{tr}[\rho H_1] - \text{tr}[\rho H_2])^2. \quad (3.109)$$

### Purification based QFI

For numerical calculations it can be useful to write the two minimization-based forms of the QFI as a minimization over purifications [Escher et al., 2012; Kolodynski, 2014]:

$$I(\rho_\theta; \theta) = 4 \min_{|\tilde{\psi}_\theta\rangle} \left( \langle \dot{\psi}_\theta | \dot{\psi}_\theta \rangle - \left| \langle \dot{\psi}_\theta | \tilde{\psi}_\theta \rangle \right|^2 \right), \quad (3.110)$$

$$I(\rho_\theta; \theta) = 4 \min_{|\tilde{\psi}_\theta\rangle} \langle \dot{\psi}_\theta | \dot{\psi}_\theta \rangle. \quad (3.111)$$

Due to the local character of the QFI, all the relevant purifications can be constructed by starting with a reference purification and constructing the others by acting with a Hermitian generator  $h_e$  on the environment.

## 3.6 Hamiltonian parameter estimation

In Sec. 3.3.1 we have written the QFI as a function of the eigenvalues and eigenvectors of the density matrix. In general it is a difficult task to diagonalize a matrix and it is hard to go further without more specification on the state. We can get some nice expressions when the channel used to imprint the parameter is a unitary channel.

### 3.6.1 Hamiltonian which commutes with its own derivative

We consider the Hamiltonian  $H(\theta)$  and the associated evolution operator  $U_H = e^{-iH(\theta)}$ . Furthermore we assume that  $H(\theta)$  commutes with its own derivative  $\partial_\theta H(\theta)$ . This will be the case when we deal with a phase-shift Hamiltonian of the form  $H(\theta) = \theta G$ , with  $G$  the generator. This example is particularly important as it appears often in physical systems, like in interferometry where  $\theta$  corresponds to a phase.

**Pure states** We start by looking at a pure initial state  $|\psi_0\rangle$ . Since the evolution is unitary, the purity does not change and the state after evolution is also pure:  $|\psi_\theta\rangle = U_H |\psi_0\rangle$ . In order to calculate the QFI we need to take the derivative of the state, which reads  $\partial_\theta |\psi_\theta\rangle = \partial_\theta U_H |\psi_0\rangle$ . In general it is not obvious how to differentiate the evolution operator. However, phase shift obeys the relation  $[H(\theta), \partial_\theta H(\theta)] = 0$  which allows to express the derivative of the evolution in a simple form:  $\partial_\theta U_H = (-i \partial_\theta H(\theta)) U_H$ . Using this result the derivative of the state is given by

$$\partial_\theta |\psi_\theta\rangle = -i \partial_\theta H(\theta) |\psi_\theta\rangle. \quad (3.114)$$

The state being pure we can use directly Eq. (3.47) to obtain the QFI

$$I(|\psi_\theta\rangle; \theta) = 4 \text{Var}[\partial_\theta H(\theta), |\psi_0\rangle] = 4(\langle \psi_0 | (\partial_\theta H(\theta))^2 | \psi_0 \rangle - \langle \psi_0 | \partial_\theta H(\theta) | \psi_0 \rangle^2). \quad (3.115)$$

## Box 8: First moment estimator and Signal-to-Noise ratio

As shown in Fig. 3.1 by optimizing over the different degrees of freedom (estimator, POVM, state) we can, starting from a probability distribution, design different figures of merit: FI (optimization over estimators), QFI (optimization over POVMs) and channel QFI (optimization over states). This process of optimizing leads to fundamental bounds, but their operational meaning—although justified by the theory—is sometimes questionable. In such cases it is preferable to go down in the hierarchy of optimization. We can thus restore the freedom of the initial state, of the POVM, and also pick a different estimator than the maximum likelihood estimator.

A common choice is to study the first moment estimator. Given a state  $\rho_\theta$  and an observable  $A$  Eq. (1.28) gives the approximate variance of the first moment estimator :

$$\text{Var}[\hat{\theta}_{m_1}] \simeq \frac{\langle A^2 \rangle - \langle A \rangle^2}{m(\partial_\theta \langle A \rangle)^2} . \quad (3.112)$$

From the right hand side of this equation we can derive a new figure of merit, often called the signal-to-noise ratio, noted  $S$ , and defined as:

$$S(\rho_\theta, A; \theta) := m \frac{(\partial_\theta \langle A \rangle)^2}{\langle A^2 \rangle - \langle A \rangle^2} , \quad (3.113)$$

with  $\langle A \rangle := \text{tr}[\rho_\theta A]$ . This figure of merit has been especially used in the early days of the field when the FI or the QFI were not yet popular [Hillery and Mlodinow, 1993; Huelga et al., 1997; Zawisky et al., 1998]. The term signal-to-noise ratio is usually defined in statistics as the power of the signal over the power of the noise. Here the signal corresponds to the variation of the expectation value of the observable  $A$  and the noise to the variance of the observable.

**Mixed states** Let us now look at the case of a mixed initial state  $\rho_0$ . We can write it in its eigenbasis to obtain  $\rho_0 = \sum_{i=1}^r p_i |\psi_i\rangle\langle\psi_i|$ , where  $r$  is the rank of the density matrix. The state  $\rho_\theta$  after evolution is given by

$$\rho_\theta = \sum_{i=1}^r p_i U_H |\psi_i\rangle\langle\psi_i| U_H^\dagger = \sum_{i=1}^r p_i |\tilde{\psi}_i(\theta)\rangle\langle\tilde{\psi}_i(\theta)| , \quad (3.116)$$

with  $|\tilde{\psi}_i(\theta)\rangle = U_H |\psi_i\rangle$ . Since  $U_H$  is a unitary operator we have  $\langle\tilde{\psi}_i(\theta)|\tilde{\psi}_j(\theta)\rangle = \langle\psi_i|\psi_j\rangle = \delta_{ij}$ , showing that the set  $\{p_i\}$  corresponds also to the eigenvalues of  $\rho_\theta$  and the set  $\{|\tilde{\psi}_i(\theta)\rangle\}$  is the new set of eigenvectors. We can then use Eq. (3.46) to express the QFI as

$$I(\rho_\theta; \theta) = 4 \sum_{i=1}^r p_i \langle\psi_i|(\partial_\theta H(\theta))^2|\psi_i\rangle - 8 \sum_{i,k=1}^r \frac{p_i p_k}{p_i + p_k} |\langle\psi_k|\partial_\theta H(\theta)|\psi_i\rangle|^2 . \quad (3.117)$$

We notice that the classical contribution to the QFI has completely vanished, and we are left only with the quantum part, even though the density matrix has some classical mixing. Indeed the

classical term in the QFI takes into account the information contained in the classical mixing: the eigenvalues should depend on the parameter. Here the mixing comes completely from the preparation of the initial state which does not depend on the parameter, and all the information about the parameter is contained in the eigenvectors. By recombining the terms in the last equation we can make clear the presence of the variance of the generator in the QFI:

$$I(\rho_\theta; \theta) = 4 \text{Var}[\partial_\theta H(\theta), \rho_0] - 8 \sum_{\substack{i,k=1 \\ i \neq k}}^r \frac{p_i p_k}{p_i + p_k} |\langle \psi_k | \partial_\theta H(\theta) | \psi_i \rangle|^2 . \quad (3.118)$$

### 3.6.2 Arbitrary Hamiltonian

**Pure states** When the Hamiltonian does not commute with its derivative, we can no longer derive the QFI as we did in the previous section. For pure states we obtain

$$I(|\psi_\theta\rangle; \theta) = 4(\langle \psi_0 | \dot{U}_H^\dagger \dot{U}_H | \psi_0 \rangle - \left| \langle \psi_0 | U_H^\dagger \dot{U}_H | \psi_0 \rangle \right|^2) . \quad (3.119)$$

We would be tempted to identify the operator  $U_H^\dagger \dot{U}_H$  with a generator in analogy with Eq. (3.115). The problem is that this operator is not Hermitian; instead it is anti-Hermitian (skew-Hermitian). Indeed if we differentiate both sides of the equation  $U_H^\dagger U_H = \mathcal{I}$  we obtain

$$U_H^\dagger \dot{U}_H = -\dot{U}_H^\dagger U_H . \quad (3.120)$$

To make the operator Hermitian we multiply it by  $i$ . We define the Hermitian operator

$$\mathcal{H} := i U_H^\dagger \dot{U}_H . \quad (3.121)$$

The square of this operator reads

$$\mathcal{H}^2 = -U_H^\dagger \dot{U}_H U_H^\dagger \dot{U}_H = \dot{U}_H^\dagger U_H U_H^\dagger \dot{U}_H = \dot{U}_H^\dagger \dot{U}_H . \quad (3.122)$$

As a result, we can again write the QFI as a variance [Giovannetti et al., 2006; Boixo et al., 2007; Pang and Brun, 2014, 2016],

$$I(|\psi_\theta\rangle; \theta) = 4 \text{Var}[\mathcal{H}, |\psi_0\rangle] = 4(\langle \psi_0 | \mathcal{H}^2 | \psi_0 \rangle - \langle \psi_0 | \mathcal{H} | \psi_0 \rangle^2) . \quad (3.123)$$

As we know the QFI is directly related to the Bures distance. In this picture, the QFI is seen as the distance between  $\rho_\theta$  and  $\rho_{\theta+d\theta}$ . By expanding  $\rho_{\theta+d\theta}$  to first order in  $d\theta$  we obtain

$$\begin{aligned} \rho_{\theta+d\theta} &\simeq (U_H + d\theta \dot{U}_H) \rho_0 (U_H^\dagger + d\theta \dot{U}_H^\dagger) \\ &\simeq (\mathcal{I} + d\theta \dot{U}_H U_H^\dagger) U_H \rho_0 U_H^\dagger (\mathcal{I} + d\theta U_H \dot{U}_H^\dagger) \\ &\simeq (\mathcal{I} - i d\theta (i \dot{U}_H U_H^\dagger)) \rho_\theta (\mathcal{I} + i d\theta (-i U_H \dot{U}_H^\dagger)) \\ &\simeq e^{-i d\theta (i \dot{U}_H U_H^\dagger)} \rho_\theta e^{i d\theta (-i U_H \dot{U}_H^\dagger)} . \end{aligned} \quad (3.124)$$

The last expression for  $\rho_{\theta+d\theta}$  is close to an evolution through a unitary operator. By taking the derivative of  $U_H U_H^\dagger = \mathcal{I}$  we obtain  $\dot{U}_H U_H^\dagger = -U_H \dot{U}_H^\dagger$ . Using this expression  $\rho_{\theta+d\theta}$  is written

$$\rho_{\theta+d\theta} \simeq e^{-i d\theta \mathcal{H}} \rho_\theta e^{i d\theta \mathcal{H}} , \quad (3.125)$$

where we defined the operator  $\mathcal{K}$  as

$$\mathcal{K} := i\dot{U}_H U_H^\dagger. \quad (3.126)$$

This operator  $\mathcal{K}$  has the interpretation of a local generator which takes the state  $\rho_\theta$  to the state  $\rho_{\theta+d\theta}$ . By working out the QFI (3.119) we can express it again as a variance, but this time involving the operator  $\mathcal{K}$  and the state  $|\psi_\theta\rangle$

$$I(|\psi_\theta\rangle; \theta) = 4 \text{Var}[\mathcal{K}, |\psi_\theta\rangle] = 4(\langle\psi_\theta|\mathcal{K}^2|\psi_\theta\rangle - \langle\psi_\theta|\mathcal{K}|\psi_\theta\rangle^2). \quad (3.127)$$

To conclude with this question of arbitrary unitary evolution and pure states we can try to get the geometrical interpretation for  $\mathcal{K}$ . In the same way as we derived equation (3.125) we obtain

$$\rho_{\theta+d\theta} \simeq U_H e^{-i d\theta \mathcal{K}} \rho_\theta e^{i d\theta \mathcal{K}} U_H^\dagger. \quad (3.128)$$

The interpretation of  $\mathcal{K}$  is less obvious, as  $\rho_{\theta+d\theta}$  is obtained by the translation due to  $U_H$ , but acting on the state  $e^{-i d\theta \mathcal{K}} \rho_\theta e^{i d\theta \mathcal{K}}$ . The invariance of the Bures distance under unitary evolution<sup>6</sup>

$$d_B(U_H e^{-i d\theta \mathcal{K}} \rho_\theta e^{i d\theta \mathcal{K}} U_H^\dagger, U_H \rho_\theta U_H^\dagger) = d_B(e^{-i d\theta \mathcal{K}} \rho_\theta e^{i d\theta \mathcal{K}}, \rho_\theta), \quad (3.129)$$

leads directly back to Eq. (3.123).

Finally, let us emphasize the connection between the spectral decompositions of  $\mathcal{H}$  and  $\mathcal{K}$ . Being Hermitian we can write  $\mathcal{K}$  as  $\mathcal{K} = \sum_{i=1}^d h_i |h_i\rangle\langle h_i|$ . Then, noticing that  $\mathcal{K} = U_H \mathcal{H} U_H^\dagger$ , we obtain the spectral decomposition of  $\mathcal{H}$  as  $\mathcal{H} = \sum_{i=1}^d h_i |\tilde{h}_i\rangle\langle \tilde{h}_i|$  with  $|\tilde{h}_i\rangle = U_H |h_i\rangle$ . These two operators have the same eigenvalues.

**Mixed states** In the same way as we did before, we can calculate the QFI for mixed states. Not very surprisingly, we obtain the same form as Eq. (3.118), but with the new generator  $\mathcal{K}$  instead of  $\partial_\theta U_H$

$$I(\rho_\theta; \theta) = 4 \text{Var}[\mathcal{K}, \rho_\theta] - 8 \sum_{\substack{i,k=1 \\ i \neq k}}^r \frac{p_i p_k}{p_i + p_k} |\langle \psi_k | \mathcal{K} | \psi_i \rangle|^2. \quad (3.130)$$

If we want to use the local generator  $\mathcal{K}$  we obtain for the QFI

$$I(\rho_\theta; \theta) = 4 \text{Var}[\mathcal{K}, \rho_\theta] - 8 \sum_{\substack{i,k=1 \\ i \neq k}}^r \frac{p_i p_k}{p_i + p_k} \left| \langle \tilde{\psi}_k | \mathcal{K} | \tilde{\psi}_i \rangle \right|^2. \quad (3.131)$$

### SNR for pure states

For a pure state  $|\psi_\theta\rangle = U_H |\psi_0\rangle$  the SNR is given by

$$S(|\psi_\theta\rangle, A; \theta) = \frac{(\partial_\theta (\langle \psi_\theta | A | \psi_\theta \rangle))^2}{\text{Var}[A, |\psi_\theta\rangle]}. \quad (3.132)$$

---

<sup>6</sup>Importantly the Bures distance is invariant under any unitary evolution, regardless of the dependence on  $\theta$ , contrarily to the Bures metric, *i.e.* the QFI, which is invariant under  $\theta$ -independent unitary evolution.

### 3.7 Perturbative Hamiltonian parameter estimation

By writing  $\langle \psi_\theta | A | \psi_\theta \rangle$  as  $\langle \psi_0 | U_H^\dagger A U_H | \psi_0 \rangle$  we can calculate easily its derivative:

$$\begin{aligned} \frac{\partial \langle \psi_\theta | A | \psi_\theta \rangle}{\partial \theta} &= \langle \psi_0 | \dot{U}_H^\dagger U_H U_H^\dagger A U_H | \psi_0 \rangle + \langle \psi_0 | U_H^\dagger A U_H U_H^\dagger \dot{U}_H | \psi_0 \rangle \\ &= i \langle \psi_0 | [\mathcal{H}, A_H] | \psi_0 \rangle . \end{aligned}$$

Using the same method as for the QFI we can express the SNR as a function of the local generator  $\mathcal{H}$ :

$$S(|\psi_\theta\rangle, A; \theta) = -\frac{\langle \psi_0 | [\mathcal{H}, A_H] | \psi_0 \rangle^2}{\text{Var}[A_H, |\psi_0\rangle]} , \quad (3.133)$$

where  $A_H := U_H^\dagger A U_H$  is the generator in the Heisenberg picture. With a similar algebra as used for the QFI we can also write the SNR as

$$S(|\psi_\theta\rangle, A; \theta) = -\frac{\langle \psi_\theta | [\mathcal{H}, A] | \psi_\theta \rangle^2}{\text{Var}[A, |\psi_\theta\rangle]} . \quad (3.134)$$

These results help us to apprehend  $\mathcal{H}$  and  $\mathcal{K}$  in a new fashion. In some way we can consider  $\mathcal{H}$  to be the local generator in the Schrödinger picture while  $\mathcal{K}$  corresponds to the local generator in the Heisenberg picture. One can also verify that  $\mathcal{K} = U_H^\dagger \mathcal{H} U_H$ , which is indeed the way of expressing an operator in the Heisenberg picture.

#### 3.6.3 Explicit form of the local generator

When we are able to write the Hamiltonian in its eigenbasis we can find an explicit form for the local generator [Pang and Brun, 2014]. Taking into account the possibility of degeneracies, the Hamiltonian is written as

$$H(\theta) = \sum_{i=1}^d \lambda_i \sum_{j_i=1}^{d_i} |\psi_i^{(j_i)}\rangle \langle \psi_i^{(j_i)}| . \quad (3.135)$$

Inserting this expression into Eq. (3.126), Pang and Brun [2014] showed that:

$$\begin{aligned} \mathcal{H} &= t \sum_{i=1}^d \frac{\partial \lambda_i}{\partial \theta} P_i + 2 \sum_{\substack{k,l=1 \\ k \neq l}}^d \sum_{i=1}^{d_k} \sum_{j=1}^{d_l} e^{-it \frac{\lambda_k - \lambda_l}{2}} \sin\left(t \frac{\lambda_k - \lambda_l}{2}\right) \\ &\quad \times \langle \psi_l^{(j)} | \partial_\theta \psi_k^{(i)} \rangle | \psi_k^{(i)} \rangle \langle \psi_l^{(j)} | , \end{aligned} \quad (3.136)$$

where  $P_i$  is the projector on the eigenspace corresponding to the eigenvalue  $\lambda_i$ . This expression shows that the generator has a term scaling linearly with time and a term periodic with time. In particular, Hamiltonians whose eigenvalues are parameter independent lead to a local generator with a purely periodic time scaling. We will see in Chapter 9 how we can restore the linear scaling by using Hamiltonian extensions.

## 3.7 Perturbative Hamiltonian parameter estimation

Calculating the QFI is in general difficult, since one needs to calculate the local generator and thus the derivative of the evolution operator. To circumvent this difficulty, we can start with a known

situation and add a perturbation. This method was introduced by Braun and Martin [2011] in the context of coherent averaging.

Consider the family of Hamiltonians  $\{H(\lambda, \theta)\}$  composed of the families  $\{H_0(\lambda)\}$  and  $\{V(\theta)\}$  in the following way:

$$H(\lambda, \theta) = H_0(\lambda) + V(\theta), \quad (3.137)$$

where  $H_0(\lambda)$  is the unperturbed Hamiltonian and  $V(\theta)$  is the perturbation. We are interested in the estimation of either the parameter  $\lambda$  or the parameter  $\theta$ , but not both at the same time. In the original publication only the estimation of  $\theta$  was considered, but it turns out that the estimation of  $\lambda$  is physically very important for coherent averaging. To deal with this situation we extended the perturbative approach to the estimation of the parameter of the unperturbed Hamiltonian.

In the interaction picture the Hamiltonian and the state are respectively given by (see Appendix B)

$$H_I(t, \lambda, \theta) := V_I = e^{iH_0(\lambda)t} V(\theta) e^{-iH_0(\lambda)t}, \quad (3.138)$$

$$|\psi_I(t, \lambda, \theta)\rangle = e^{iH_0(\lambda)t} |\psi_S(t, \theta)\rangle, \quad (3.139)$$

where  $|\psi_S(t, \theta)\rangle$  is the state in the Schrödinger picture. The perturbative solution, up to the second order in perturbation, of the Schwinger-Tomonaga equation reads

$$|\psi_I(t, \lambda, \theta)\rangle = \left( \mathcal{I} - i \int_0^t dt_1 H_I(t_1, \lambda, \theta) - \int_0^t dt_1 \int_0^{t_1} dt_2 H_I(t_1, \lambda, \theta) H_I(t_2, \lambda, \theta) \right) |\psi_0\rangle. \quad (3.140)$$

### 3.7.1 Parameter in the perturbation: PT I

We start with the estimation of  $\theta$ , the parameter encoded in the perturbation. We refer to this situation as PT I. For the sake of concision we do not write in this section the dependence on  $\lambda$  (not in the states nor in the Hamiltonians). In the Schrödinger picture the QFI when starting with the pure state  $|\psi_0\rangle$  is given by:

$$I_\theta := I(|\psi_S(t, \theta)\rangle; \theta) = 4 \left( \langle \partial_\theta \psi_S(t, \lambda, \theta) | \partial_\theta \psi_S(t, \lambda, \theta) \rangle - |\langle \psi_S(t, \lambda, \theta) | \partial_\theta \psi_S(t, \lambda, \theta) \rangle|^2 \right). \quad (3.141)$$

For the sake of concision we do not make the state dependence in the QFI explicit anymore, and we write the parameter as a subscript. Using the relation between the state in the Schrödinger picture and the state in the interaction picture we directly obtain

$$I_\theta = 4 \left( \langle \partial_\theta \psi_I(t, \theta) | \partial_\theta \psi_I(t, \theta) \rangle - |\langle \psi_I(t, \theta) | \partial_\theta \psi_I(t, \theta) \rangle|^2 \right). \quad (3.142)$$

This result is still exact.

Now we develop  $I_\theta$  up to second order in the perturbation, which will be denoted as  $I_\theta^{V_I}$ , where the superscript  $V_I$  is there to remind us that we use  $V$  as a perturbation. To calculate the perturbed QFI we need  $|\partial_\theta \psi_I(t, \theta)\rangle$  at the order two in perturbation. The derivative of the perturbed state (3.140) is found to be

$$|\partial_\theta \psi_I(t, \theta)\rangle = \left( i \int_0^t \partial_\theta H_I(t_1, \theta) dt_1 - \int_0^t dt_1 \int_0^{t_1} dt_2 \{ (\partial_\theta H_I(t_1, \theta)) H_I(t_2, \theta) \right. \quad (3.143)$$

$$\left. + H_I(t_1, \theta) (\partial_\theta H_I(t_2, \theta)) \} + \mathcal{O}(H_I^3) \right) |\psi_0\rangle. \quad (3.144)$$

to reach this result we consider that the term  $\partial_\theta H_I(t_1, \theta)$  was of the same order as the term  $H_I(t_1, \theta)$ , meaning that we consider both the Hamiltonian and its first derivative to be small. This is important to keep in mind as it for example excludes the use of a perturbation of the form  $\theta V$  with only  $\theta$  small (as then the derivative will not be small anymore). We furthermore assumed that  $\partial_\theta H_I(t, \theta)$  is Hermitian.

Up to second order in the perturbation the first term of the right hand side of Eq. (3.142) reads

$$\langle \partial_\theta \psi_I(t, \theta) | \partial_\theta \psi_I(t, \theta) \rangle = \int_0^t dt_1 \int_0^{t_1} dt_2 \langle \psi_0 | \partial_\theta H_I(t_1, \theta) \partial_\theta H_I(t_2, \theta) | \psi_0 \rangle . \quad (3.145)$$

The calculation of the second term leads to

$$|\langle \partial_\theta \psi_I(t, \theta) | \psi_I(t, \theta) \rangle|^2 = \left( \int_0^t dt_1 \langle \psi_0 | \partial_\theta H_I(t_1, \theta) | \psi_0 \rangle \right)^2 . \quad (3.146)$$

Combining Eq. (3.145) and (3.146) we get the QFI for  $\theta$  up to the order  $\mathcal{O}(H_I^2)$  in perturbation [Braun and Martin, 2011]

$$I_\theta^{Vi} = 4 \int_0^t dt_1 \int_0^{t_1} dt_2 K_{|\psi_0\rangle}(\partial_\theta H_I(t_1, \theta), \partial_\theta H_I(t_2, \theta)) + \mathcal{O}(H_I^3) , \quad (3.147)$$

where  $K_{|\psi\rangle}(A, B) = \langle \psi | AB | \psi \rangle - \langle \psi | A | \psi \rangle \langle \psi | B | \psi \rangle$ .

### 3.7.2 Parameter in the unperturbed Hamiltonian: PT II

We turn to the situation where the parameter to be estimated is encoded in the unperturbed Hamiltonian. We refer to this situation as perturbation theory II. For the sake of concision we do not write in this section the dependence on  $\theta$  (not in the states nor in the Hamiltonians). With this notation the Hamiltonian and the state in the interaction picture are written as

$$\begin{aligned} H_I(t, \lambda) &= e^{i H_0(\lambda)t} V e^{-i H_0(\lambda)t} , \\ |\psi_I(t, \lambda)\rangle &= e^{i H_0(\lambda)t} |\psi_S(t, \lambda)\rangle . \end{aligned}$$

This time we have to be more careful to express the QFI for pure states in the interaction picture. Indeed the interaction picture depends on the unperturbed Hamiltonian, which itself depends on the parameter to be estimated. Using the local generator for the free Hamiltonian  $\mathcal{H}_0 := i e^{i H_0(\lambda)t} \frac{\partial e^{-i H_0(\lambda)t}}{\partial \lambda}$ , the QFI expressed through the state in the interaction picture is given by

$$\begin{aligned} I_\lambda/4 &= \langle \partial_\lambda \psi_I(t, \lambda) | \partial_\lambda \psi_I(t, \lambda) \rangle - |\langle \psi_I(t, \lambda) | \partial_\lambda \psi_I(t, \lambda) \rangle|^2 - 2 \langle \psi_I(t, \lambda) | \mathcal{H}_0 | \psi_I(t, \lambda) \rangle \\ &\quad \times \Im \{ \langle \psi_I(t, \lambda) | \psi_I(t, \lambda) \rangle \} - 2 \Im \{ \langle \psi_I(t, \lambda) | \mathcal{H}_0 | \partial_\lambda \psi_I(t, \lambda) \rangle \} \\ &\quad + \langle \psi_I(t, \lambda) | \mathcal{H}_0^2 | \psi_I(t, \lambda) \rangle - (\langle \psi_I(t, \lambda) | \mathcal{H}_0 | \psi_I(t, \lambda) \rangle)^2 . \end{aligned} \quad (3.148)$$

Under the condition that the unperturbed Hamiltonian commutes with its own derivative with respect to  $\lambda$ , meaning that  $[H_0(\lambda), \partial_\lambda H_0(\lambda)] = 0$ , we can simplify the QFI to obtain

$$I_\lambda = 4 (I_{\lambda,1} + t I_{\lambda,1} + t^2 I_{\lambda,t^2}) , \quad (3.149)$$

with

$$I_{\lambda,1} = \langle \partial_\lambda \psi_I(t, \lambda) | \partial_\lambda \psi_I(t, \lambda) \rangle - |\langle \psi_I(t, \lambda) | \partial_\lambda \psi_I(t, \lambda) \rangle|^2, \quad (3.150)$$

$$I_{\lambda,t} = 2\Im \{ \langle \partial_\lambda \psi_I(t, \lambda) | \partial_\lambda H_0(\lambda) | \partial_\lambda \psi_I(t, \lambda) \rangle \} - 2i \langle \psi_I(t, \lambda) | \partial_\lambda H_0(\lambda) | \psi_I(t, \lambda) \rangle \\ \times \Im \{ \langle \psi_I(t, \lambda) | \partial_\lambda \psi_I(t, \lambda) \rangle \}, \quad (3.151)$$

$$I_{\lambda,t^2} = \langle \psi_I(t, \lambda) | (\partial_\lambda H_0(\lambda))^2 | \psi_I(t, \lambda) \rangle - \langle \psi_I(t, \lambda) | \partial_\lambda H_0(\lambda) | \psi_I(t, \lambda) \rangle^2. \quad (3.152)$$

Up to here the result is still exact. The point in expressing the QFI in the interaction picture is to use perturbation theory. We thus need to plug the perturbed state and its derivative into the QFI. The result is cumbersome and will not be reported here (see Appendix C.1) but deserves an important comment. Indeed, other than the QFI obtained in the previous section, the QFI includes terms of the zeroth and first order of the perturbation. Especially, considering the term of order zero we have

$$I_\lambda^{V_1} = 4 (\langle \psi_0 | (\partial_\lambda H_0(\lambda))^2 | \psi_0 \rangle - \langle \psi_0 | \partial_\lambda H_0(\lambda) | \psi_0 \rangle^2) + \mathcal{O}(H_I). \quad (3.153)$$

This is nothing else than the QFI obtained by considering only the unperturbed Hamiltonian. The other order terms will only introduce corrections, which means that in this situation the QFI is not strongly affected by the perturbation.

## 3.8 SNR in perturbation theory

Finally we look at the SNR for the measurement of an observable  $A$  in both PT I and PT II.

### 3.8.1 SNR in PT I

In PT I we want to estimate the parameter  $\theta$  and the exact SNR reads

$$S_\theta^A := S(\langle \psi_S(t, \lambda, \theta) \rangle, A; \theta) = \frac{|\frac{\partial}{\partial \theta} \langle \psi_I(t, \lambda, \theta) | A_I(t, \lambda) | \psi_I(t, \lambda, \theta) \rangle|^2}{\text{Var}[A_I(t, \lambda), | \psi_I(t, \lambda, \theta) \rangle]} \quad (3.154)$$

where  $A_I(t, \lambda) := e^{i t H_0(\lambda)} A e^{-i t H_0(\lambda)}$  and where we introduce a short notation for the SNR. The expectation value  $\langle \psi_I(t, \lambda, \theta) | A_I(t, \lambda) | \psi_I(t, \lambda, \theta) \rangle$  is given by

$$\langle \psi_I(t, \lambda, \theta) | A_I(t, \lambda) | \psi_I(t, \lambda, \theta) \rangle = \langle \psi_0 | A_I(t, \lambda) | \psi_0 \rangle + i \int_0^t dt_1 \langle \psi_0 | [H_I(t_1, \theta), A_I(t, \lambda)] | \psi_0 \rangle \\ - \int_0^t dt_1 \int_0^{t_1} dt_2 \langle \psi_0 | A_I(t, \lambda) H_I(t_1, \theta) H_I(t_2, \theta) + H_I(t_2, \theta) H_I(t_1, \theta) A_I(t, \lambda) | \psi_0 \rangle \\ + \int_0^t dt_1 \int_0^{t_1} dt_2 \langle \psi_0 | H_I(t_1, \theta) A_I(t, \lambda) H_I(t_2, \theta) | \psi_0 \rangle. \quad (3.155)$$

The average value of  $A_I(t, \lambda)^2$  is obtained in the same way.

To obtain the SNR to second order in perturbation theory we need to pay attention to the fact that we have to develop both the numerator and the denominator. Notice that since  $\text{Var}[A_I(t, \lambda), | \psi_0 \rangle]$  is independent of  $\theta$  the first order term is of order two. For the leading order we obtain

$$S_{\theta, V_1}^A = \frac{\left( \int_0^t dt_1 \langle \psi_0 | \left[ \frac{\partial H_I(t_1, \theta)}{\partial \theta}, A_I(t, \lambda) \right] | \psi_0 \rangle \right)^2}{\text{Var}[A_I(t, \lambda), | \psi_0 \rangle]} + \mathcal{O}(H_I^3), \quad (3.156)$$

where the subscript  $V_1$  is there to remind us that the result is a perturbative result.

### 3.8.2 SNR in PT II

In the case of PT II we are interested in the SNR for  $\lambda$ . Only the derivative of the expectation value changes, and we obtain

$$S_{\lambda}^A = \frac{\left| \frac{\partial}{\partial \lambda} \langle \psi_I(t, \lambda, \theta) | A_I(t, \lambda) | \psi_I(t, \lambda, \theta) \rangle \right|^2}{\text{Var}[A_I(t, \lambda), | \psi_I(t, \lambda, \theta) \rangle]} . \quad (3.157)$$

The leading order term reads

$$S_{\lambda, V_I}^A = \frac{\langle \psi_0(t, \lambda) | [\mathcal{K}_0, A] | \psi_0(t, \lambda) \rangle^2}{\text{Var}[A, | \psi_0(t, \lambda) \rangle]} + \mathcal{O}(H_I) , \quad (3.158)$$

where  $\mathcal{K}_0 := -\frac{\partial e^{-i t H_0(\lambda)}}{\partial \lambda} e^{i t H_0(\lambda)}$  and  $|\psi_0(t, \lambda)\rangle := e^{-i t H_0(\lambda)} |\psi_0\rangle$ . This is nothing else than the unperturbed SNR which can also be written as

$$S_{\lambda, V_I}^A = \frac{\langle \psi_0 | [\mathcal{K}_0, A_I(t, \lambda)] | \psi_0 \rangle^2}{\text{Var}[A_I(t, \lambda), | \psi_0 \rangle]} + \mathcal{O}(H_I) . \quad (3.159)$$

## Summary Chapter 3

- **Quantum estimation problem:** We consider a known  $\theta$ -dependent quantum state  $\rho_\theta$  and want to estimate the value of  $\theta$  by performing measurements on  $\rho_\theta$ . Equivalent to the classical problem with the freedom of choosing the POVM that produces a probability distribution.
- **Quantum Cramér-Rao bound:** To obtain a bound on  $\text{Var}[\hat{\theta}_{\text{est}}]$  we maximize the classical FI over the set of all POVM's. We thus have  $\text{Var}[\hat{\theta}_{\text{est}}] \geq (n I(\rho_\theta; \theta))^{-1}$ , with  $n$  the number of independent measurements done on independent replicas of the state.
- **Quantum Fisher information:** The QFI is defined as  $I(\rho_\theta; \theta) = \max_{M_\xi} J(\text{tr}[M_\xi \rho_\theta]; \theta)$ . It characterizes the amount of information about  $\theta$  contained in  $\rho_\theta$ . Like the FI the QFI is a local quantity depending only on the state and its first derivative.
- **Sensitivity and QFI:** The QFI can also be seen as a characterization of the sensitivity of the state with the parameter. It corresponds to the Bures distance between two neighbouring states  $I(\rho_\theta; \theta) = 4 d_B(\rho_\theta, \rho_{\theta+d\theta})^2/d\theta^2$ .
- **Saturation of the bound:** In the one-parameter case the bound can always be reached asymptotically ( $n \rightarrow \infty$ ).
- **Channel estimation and channel QFI:** We want to estimate the parameter  $\theta$  encoded in a quantum channel  $\mathcal{E}_\theta$ . Channel QFI corresponds to the maximal QFI, with maximization over input states  $\rho_0$ :  $C(\mathcal{E}_\theta; \theta) := \max_{\rho_0} I(\mathcal{E}_\theta(\rho); \theta)$ .
- **Hamiltonian parameter estimation:** Estimation of a parameter encoded in a Hamiltonian  $H(\theta)$ . The QFI for a pure state is proportional to the variance of the local generator  $\mathcal{H} := i U_H^\dagger \frac{\partial U_H}{\partial \theta}$ .

## Chapter 4

# Quantum metrology and quantum enhanced measurements

We now turn our attention to quantum metrology — the field of measuring physical quantities in quantum systems — and to quantum enhanced measurements — the use of some specific strategies to increase the sensitivity for a given estimation problem. We choose to distinguish quantum metrology and quantum enhanced measurements from quantum parameter estimation theory (q-pet): In the former we investigate some specific protocols usually related to some specific physical quantities while in the latter we derive the analytic tools. Typically we use the figure of merits developed in q-pet (FI, QFI, channel QFI) to analyse the efficiency of some protocols.

Due to the uttermost importance of metrology in physics, quantum metrology became a huge field and it is not in our agenda to make an exhaustive study of it. Rather we want to present the basics of it along with some specific topics chosen either due to their global importance or because they are connected to our work.

### 4.1 Historical approach of the standard quantum limit and the Heisenberg limit

Before formalizing the framework of quantum metrology we will take a glance of the origin of the field. Apart from the pure historical interest this presentation is also useful from a physical point of view. As we will see in the next sections, two concepts playing a fundamental role in quantum enhanced measurements, the standard quantum limit (SQL) and the Heisenberg limit (HL), suffer problems of definition leading to physical and semantic confusions. An historic survey can only prove useful in this context.

#### 4.1.1 Gravitational wave detection

Quantum metrology owes a lot to general relativity. Not because some concepts of general relativity are enlightening to better understand the quantum world, but because the quest for detecting effects predicted by general relativity has pushed scientists to think on how to measure incredibly small quantities. Theoretically introduced in 1916 by Albert Einstein, gravitational waves (GWs) have been the object of long debates about their physical reality. When it became clear by the end

of the fifties that GWs lead to some measurable effects (see Box 9), their detection became one of the big stakes of modern physics. First because it would constitute a new brilliant demonstration of the validity of general relativity, but also for the advances that this would provide in astronomy and cosmology. Indeed conventional astronomy is based on the observation of electromagnetic radiations, such that if an object does not emit electromagnetic radiation — as it is the case for a black hole — one cannot observe it directly. But if this object emits GWs then we will be able to observe it directly. Moreover GWs propagate almost freely in the matter (a fact that does not help to detect them) which means that observation could help to probe the interior of dense systems.

#### Box 9: The genesis of GW detection

Although clearly appearing in the formalism of general relativity, the reality of GWs has been under debate during the first half of the XXth century. It was under question if these waves were an artefact due to a choice of coordinates or if they had a physical meaning. Einstein himself, who introduced them, developed doubt about their existence. This controversy stayed alive until 1957, the year when most of the community accepted the physical character of GWs [Aguiar, 2011; Saulson, 2011, 2013].

This year Hermann Bondi published a paper considered to be the paper which cleared the controversy [Bondi, 1957]. In this same year, in the Chapel Hill Conference [Bergmann, 1957], Richard Feynman exposed the famous argument of the "sticky bead"<sup>a</sup>. But the story does not finish here. Felix Pirani, a former student of Bondi who was also present at the Chapel Hill Conference, discussed before Feynman how one could use free masses to detect GWs. He also submitted a paper before the conference, [Pirani, 1956, 1957]. As we see it is not completely obvious to whom we should give — if we should — the credit for showing that GWs have a physical reality.

After that, the debate evolved into the question of whether or not binary stars could emit GWs. The detection in 1974 of the first binary pulsar solved definitely all these controversies — and allowed Hulse and Taylor to win the Nobel prize in 1993. The study of the orbit of the system showed that its spatial expansion is decreasing, a phenomena explained by the emission of GWs. Therefore this discovery was the first observation showing the existence of GWs. Nevertheless this was considered (although not by everybody — see footnote 2 in [Kennefick, 2014]) as an indirect observation of GWs, and a direct detection was still sought.

---

<sup>a</sup>This Gedanken experiment is simple: we consider a bead lying on a rod and a GW propagating in the transverse direction. This wave will put the bead in movement, and due to the friction between the bead and the rod the rod will be heated. By measuring the change of temperature one should be able to detect the GW. See [Cooperstock, 2015] for a discussion of the sticky bead argument

The link with quantum mechanics is found by looking at the basic mechanisms that can be used for the detection of GWs. A GW perturbs the metric of space-time. To detect it one should be able to design a device sensitive enough to these perturbations. Two different methods are known and implemented. Historically the first one, resonant-mass detection is based on the measurement of the excitations in a solid caused by a GW: When a GW passes by, the atoms in the solid will try to stick to the new geodesics while the electrostatic forces try to keep the atoms at a fixed position, resulting in the vibration of the solid. The second method is based on free masses: If we consider two free falling masses their distance over time will change due to the passage of the

## 4.1 Historical approach of the standard quantum limit and the Heisenberg limit

gravitational wave.

Due to the distance that lies between us and the potential event that generates the GW their effects on earth are very weak, producing then very small signals, that in order to be detected should be distinguishable from the noise. This challenge motivated researchers to better understand the limits in the detection of a weak classical force. Braginsky [1967] showed that by isolating properly the free masses from the heat bath it is possible to bring the masses to the domain of quantum behaviour, even when working at high temperatures (meaning from few Kelvins to room temperature). Obviously, such an analysis assumes that all the other noises have been discarded. For a particle which is almost free (only subject to the force to be measured and to frictional forces with a relaxation time  $\tau^*$ ), the temperature at which the system starts to behave quantum mechanically is given by  $T \leq \frac{\hbar\tau^*}{2k_B\tau^2}$  where  $\tau$  is the measurement time (equal to room temperature for  $\tau = 10^{-3}$  and  $\tau^* = 4 \times 10^7$ ). In the case of a harmonic oscillator of frequency  $\omega$  the temperature should be  $T \leq \frac{\hbar\omega\tau^*}{2k_B\tau}$  (a constraint satisfied at liquid helium temperature already at the end of the eighties — see [Braginsky et al., 1985]) [Braginsky et al., 1992].

### 4.1.2 Birth of the SQL

The fact that GW detectors are likely to work in the quantum regime implies that *in fine* the ultimate limitation in their sensitivity is set by quantum mechanics. An analysis of the process of detection of the GW reveals the limits that back-action sets on the sensitivity of the detector.

#### Monitoring the position of a particle

The detection of GW using free masses is done by monitoring the position of a free mass over time. Fundamentally this means that we are interested in the *successive* measurements of the position of a free mass. As quantum mechanics teaches us that in general we cannot measure a system without perturbing it, it is clear that this measurement cannot be infinitely accurate. A standard argument to find the limit is based on the Heisenberg uncertainty relation applied to the position and momentum operator, respectively denoted  $\hat{X}(t)$  and  $\hat{P}(t)$ . The original argument is the following: A measurement of the position is performed at a time  $t = 0$  with an uncertainty  $\Delta\hat{X}(0) := \sqrt{\text{Var}[\hat{X}(0)]}$ . Then the position of the mass at a time  $t$  will be  $\hat{X}(t) = \hat{X}(0) + \hat{P}(0)t/m$  and leads to an uncertainty in the repeated measurement of the position at time  $t$  equal to

$$(\Delta\hat{X}(t))^2 = (\Delta\hat{X}(0))^2 + (\Delta\hat{P}(0))^2 t^2 / m^2, \quad (4.1)$$

with  $m$  the mass. Using the Heisenberg uncertainty principle which states that  $\Delta\hat{X}(0)\Delta\hat{P}(0) \geq \hbar/2$ , we obtain

$$(\Delta\hat{X}(t))^2 \geq (\Delta\hat{X}(0))^2 + \left( \frac{\hbar t}{2\Delta\hat{X}(0)m} \right)^2 \geq 2\Delta\hat{X}(0) \frac{\hbar t}{2\Delta\hat{X}(0)m} = \frac{\hbar t}{m}. \quad (4.2)$$

This limit got known at the beginning of the eighties as the SQL for the measurement of a free mass.

#### Quantum harmonic oscillators

When looking at the detection of GWs using resonant-mass detectors the fundamental system to be investigated is the quantum harmonic oscillator [Weber, 1960; Thorne, 1980; Smarr, 1979].

## Box 10: Controversy on the SQL

In [Yuen, 1983a,b], Yuen pointed out a flaw in the original argument. Indeed the proper formula for the uncertainty of  $\hat{X}(t)$  should take in consideration the cross terms between  $\hat{X}(t)$  and  $\hat{P}(t)$ :

$$(\Delta\hat{X}(t))^2 = (\Delta\hat{X}(0))^2 + (\Delta\hat{P}(0))^2 t^2 / m^2 + \frac{\langle\hat{X}(0)\hat{P}(0) + \hat{P}(0)\hat{X}(0)\rangle - 2\langle\hat{X}(0)\rangle\langle\hat{P}(0)\rangle}{m}.$$

In the standard derivation of the SQL there is the implicit assumption that the cross term is at best zero and at least positive. But Yuen showed that there exist states, called *contractive* states, for which the cross term is negative. This lead him to claim that the SQL is not valid.

The next step in the story was the answer of Caves to Yuen's paper [Caves, 1985]. While acknowledging the existence of the flaw in the derivation of the SQL, Caves defended the validity of the SQL: "the flaw lies in the standard argument, not in the SQL". His argument — a heuristic one — was based on the fact that one should not only focus on the intrinsic uncertainty of the state, but also include the finite resolution of the measuring device. Taking this in consideration he was able to derive the SQL. But here again he should make an assumption on the link existing between the precision of the measurement device and the uncertainty of the state.

The debate then evolved in the discussion of the different meaning of the uncertainty in the free-mass monitoring [Ozawa, 1988] and eventually in the discussion of the exact meaning and formulation of the Heisenberg uncertainty relation ([Busch et al., 2013; Ozawa, 2004]). As a result the controversy on the SQL for a free mass was still lively in the XXIth century (see for example [Kosugi, 2010] where the author showed a refined version of the SQL).

Consider an oscillator of mass  $m$  and frequency  $\omega$ . We also introduce the position and momentum operator  $\hat{X}$  and  $\hat{P}$  as well as the complex amplitude  $\hat{X}_1$  and  $\hat{X}_2$ , with the relation  $\hat{X} + i\frac{\hat{P}}{m\omega} = (\hat{X}_1 + i\hat{X}_2)e^{-i\omega t}$ . Like  $\hat{X}$  and  $\hat{P}$ ,  $\hat{X}_1$  and  $\hat{X}_2$  are conjugate observables and therefore obey the Heisenberg uncertainty relation  $\Delta\hat{X}_1\Delta\hat{X}_2 \geq \hbar/(2m\omega)$ .

The way to derive the SQL for the harmonic oscillator [Braginsky, 1970; Giffard, 1976] is slightly less direct than for the free mass, and uses a heuristic argument. The point is that one should carefully think on what is to be measured for detecting a weak force acting on the oscillator. A standard measurement consists in measuring both the amplitude and the phase of the oscillator. This corresponds to a continuous monitoring of the position  $\hat{X}(t) = \hat{X}_1 \cos(\omega t) + \hat{X}_2 \sin(\omega t)$  of the oscillator over time. The uncertainty is thus equal to  $(\Delta\hat{X}(t))^2 = (\Delta\hat{X}_1)^2 \cos^2(\omega t) + (\Delta\hat{X}_2)^2 \sin^2(\omega t)$ . By measuring simultaneously  $\hat{X}_1$  and  $\hat{X}_2$  with the same uncertainty we can monitor the position with a constant precision:  $\Delta\hat{X}_1 = \Delta\hat{X}_2 = \Delta\hat{X}(t)$ . Coupling this equation with the Heisenberg uncertainty relation we obtain the SQL for a harmonic oscillator:

$$\Delta\hat{X}(t) \geq \Delta X_{\text{SQL}} := \sqrt{\frac{\hbar}{2m\omega}}. \quad (4.3)$$

Historically it seems that it is this limit that was called "standard quantum limit" for the first time in the early review of Caves et al. [1980], "standard" referring to the fact that the measurement of

#### 4.1 Historical approach of the standard quantum limit and the Heisenberg limit

both the amplitude and the phase of the oscillator "is the type of measurement made by standard electronic devices."

Finally let us mention that in this review the SQL is also given for the number of quanta  $\hat{N}$  and the phase of the oscillator. Although it is well known that there is no such object as a phase operator in quantum mechanics, the authors circumvent the problem by considering an oscillator with a large number of quanta  $\hat{N} \gg 1$ , with  $N := \langle \hat{N} \rangle$ , and defining the phase as  $\phi := \tan^{-1}(\hat{X}_2/\hat{X}_1)$ . Then the SQL reads

$$\Delta\hat{N} \geq \sqrt{N + 1/4} \quad , \quad \Delta\phi \geq \frac{1}{2\sqrt{N}} . \quad (4.4)$$

##### Box 11: GW detection with resonant-mass antennas [Aguir, 2011].

Soon after the physicists agreed on the reality of GWs, experiments for detecting them were launched. Most of the credit for starting experimental work on this topic is given to John Weber. He was among the first to believe that the technological advances would allow the detection of GWs. At the beginning of the sixties Weber used large aluminium bars at room temperature with piezzo-electric transducers to detect the vibration at the resonant frequency<sup>a</sup>, what is known nowadays as *Weber's bars*. At the end of the sixties Weber announced the first detection of GWs. In the following years many groups tried to reproduce his results but failed to do so, even by using more sensitive bars that were cooled, proving that the events detected by Weber were not GWs.

However, with this first series of experiments Weber opened the path for actual GW detection<sup>b</sup>. The next generations of Weber's bar were cooled at 4 K (liquid Helium temperature), were equipped with better suspension systems for vibration isolation and benefited from better detection equipment (transducer and amplifiers). In the nineties two groups used bars cooled at 0.1 K and at the turn of the century five groups were managing highly sensitive bars, allowing four months of quadruple operations, unfortunately without success.

Eventually, the last generation of resonant-mass detectors uses a spherical geometry. This geometry offers the advantage against bars that its sensitivity does not depend on the direction of the source. It is planned to cool them down to a few millikelvins, and there is hope for reaching — and maybe even beating — the SQL.

<sup>a</sup>Here lies the weak point of Weber bars: they are designed to detect signals at only one frequency, their resonant frequency

<sup>b</sup>For the anecdote, Robert Forward, a student of John Weber, played a crucial role in the development of the first interferometric GW detectors [Moss et al., 1971].

#### Beating the SQL: Quantum non-demolition measurements

Ironically, before Yuen initiated the controversy on the validity of the SQL (see box 10), the discussion was not so much about the SQL itself, whose "existence is firm", but rather on the way to surpass it. In what is considered to be the first paper showing explicitly the SQL, Braginsky and Vorontsov [1974] proposed to monitor the number of phonons in a harmonic oscillator to surpass the SQL. This was the first proposal for what would be known some years later as *quantum non-*

*demolition* (QND) measurements<sup>1</sup>. But the example was flawed [Thorne et al., 1978; Unruh, 1978] as it appeared that such monitoring will not be possible in practice. The first real QND measurements were proposed a few years later [Braginsky et al., 1977; Thorne et al., 1978]. The field received a lot of interest and in 1980 already two reviews were published [Braginsky et al., 1980; Caves et al., 1980]. More mature reviews are [Braginsky et al., 1992; Braginsky and Khalili, 1996; Bocko and Onofrio, 1996].

The way that a QND measurement works is best understood starting from the SQL. In the case of the free-mass measurement we see that what causes problems is the fact that we monitor an observable ( $\hat{X}(t)$ ) whose conjugate observable ( $\hat{P}(t)$ ) affects the dynamics of the system (the Hamiltonian being proportional to  $\hat{P}(t)$ ), causing a disturbance in the repeated measurement. The solution to avoid this limitation is to design a QND observable, *i.e.* an observable which is conserved during the free evolution of the object. A trivial QND observable for the free mass is its momentum. When measuring the momentum one disturbs the position, but this does not prevent to measure again and again the same momentum. When looking at the SQL for a harmonic oscillator we get a different angle to tackle the problem. There the SQL is enforced by the fact that we want to measure at the same time two non-commuting observables. To surpass it one should then find a way to measure the quantity of interest (for example the force acting on the oscillator) without monitoring two conjugate observables.

From a historical point of view it is interesting to note that Landau and Peierls [1931] already discussed successive measurements of a free mass. In this paper one can also find a footnote making reference to what is now called QND measurement. Elsasser also, in 1937 discussed measurements that "do not alter the measured system" [Elsasser, 1937]. We should also cite Bohm who pointed out in 1951 a condition — who proved to be too strong — for a measurement to be non-demolition. These few examples show that the question of perturbation and of QND measurements were already present at the beginning of quantum mechanics, which is not very surprising as the Heisenberg uncertainty relation and their consequences are a striking point in quantum mechanics. But at that time, the possibility of measuring single quantum systems looked so far that all those discussions could be seen as purely formal and academic. It is the combination of the technological advances — and among them the invention of the laser — and the challenge offered by GW detection that pushed physicists to discuss more thoroughly these questions.

### 4.1.3 SQL and HL in interferometry

#### Michelson and Mach-Zender interferometers

Interferometers played a great role in modern physics, starting from the famous double slit experiment of Young at the beginning of the XIXth century. This experiment was at the basis of the recognition of the wave-like character of light. Some eighty years later, it was again an interferometric experiment that allowed Michelson and Morley to discard (up to a certain confidence level) the existence of ether, taking a significant step in the direction of special relativity. Hundred years after it is again a Michelson interferometer that was used for the detection of GWs.

In a basic Michelson interferometer (depicted in Fig. 4.1) the light enters in one port, and is divided in two by a beam splitter (we assume a 50/50 beam splitter). Each beam follows a different

---

<sup>1</sup>Here again the vocabulary took time to get stabilized: The first to use the term "quantum non-demolition" measurements are Thorne, Drever, and Caves [Thorne et al., 1978]. Before, but also after, other terms have been and will be used: "non-perturbative", "non-invasive", "non-destructive" or "back-action evading". Even nowadays not everybody agrees on the vocabulary and some researcher reject the use of the term "non-demolition", arguing that it is misleading [Monroe, 2011].

## Box 12: GW detection with interferometers: GW150914 and GW151226

The development of the first interferometers for GW detection can be traced back to the late sixties/beginning of the seventies with the construction of small prototypes in laboratories. At the beginning of the eighties, two bigger projects, but still too small and noisy to detect any GWs, were developed and showed the possibility of building an interferometer with 1km long arms which should in principle be able to detect GWs. This was the birth of the LIGO (Laser Interferometer Gravitational-Wave Observatory) project in the USA. It took still many years before LIGO sprang up, but eventually in 1997 the construction was achieved and the measurement started in 2002.

At the same epoch other interferometers were also being built or planned to be build. The most important were GEO 600 [Dooley and Collaboration, 2015] in Germany (in collaboration with England) and VIRGO [Acernese et al., 2015] in Italy (in collaboration with France). Having multiple detectors (and working at the same time!) offers different advantages. Metrological advantage as it is easier to make cross verifications to be sure that the event observed was indeed a GW, but also an advantage for astronomy as having different detectors distributed on the earth allows to identify more precisely the origin of the GW.

Obviously, reaching the sensitivity necessary to detect GWs is a tremendous technical challenge. The current generation of interferometers uses some classical tricks (no QND measurement are used for example) to improve their sensitivity. We can cite for example power recycling that consists in placing a partially reflecting mirror between the laser and the beam splitter: As a large part of the power of the laser is sent back to itself by the beam splitter, this mirror allows to resonantly increase the light power in the interferometer. By doing the same at the output port we can increase the signal, a trick known as signal recycling. By placing also such mirrors in the two arms of the interferometer we obtain two Fabry-Perrot cavities which make the light interfering many time with itself, increasing again the signal. All of this adds to very efficient detectors, impressive suspension systems for the free masses and a strong vacuum in the arms of the interferometer [Schnabel et al., 2010; Abbott et al., 2009].

Thanks to all these efforts, advanced LIGO could reach a strain sensitivity of the order of  $10^{-23}/\sqrt{\text{Hz}}$  [Martynov et al., 2016] and finally, on February 11, 2016 the LIGO collaboration jointly with the VIRGO collaboration announced the first detection of a GW on September 9, 2015, due to the merger of two black holes [Abbott et al., 2016b]. This first detection was followed by a second one on December 26, 2015, also due to the merger of two black holes [Abbott et al., 2016a].

In the future, it seems that increasing LIGO sensitivity would require the use of quantum enhanced measurement and QND [Abbott et al., 2016c; Ma et al., 2017]. Notice that the use of squeezed vacuum states has already been implemented in Geo 600 [LSC, 2011]. Increasing the power of the laser is also planned. Finally let us mention the eLISA project: a space interferometer with "arms" long of 2.5 million of kilometres. A test version, called LISA pathfinder, was launched in 2015 and already demonstrated a higher-than-expected sensitivity.

path, bounces off the mirror, and goes back to the beam splitter, where both beams recombine. A displacement of the mirrors results in the appearance of a phase difference  $\phi$  between the two

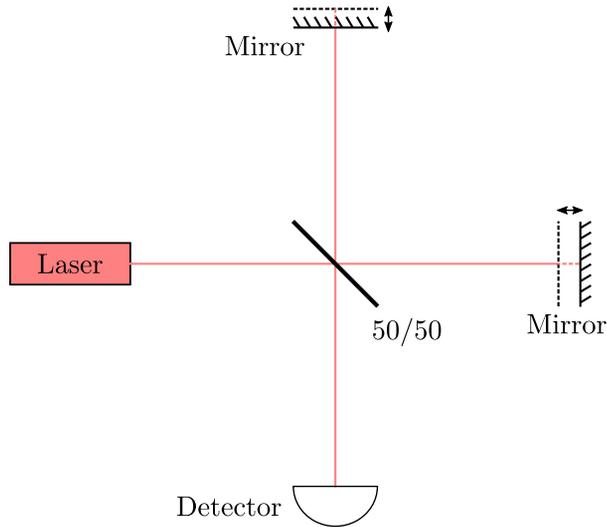


Figure 4.1: Michelson interferometer for detection of GWs. The arrows shows the displacement of the mirrors when a GW with a a direction of propagation perpendicular to the plan of the interferometer passes.

beams. A measurement of the intensity of the light at the detector allows to infer the value of this phase difference. From there comes the idea to use Michelson interferometers to detect gravitational waves, as such a wave will contract one arm of the interferometer and dilate the other. The displacement of the mirror due to the passage of a GW is typically of the order of  $10^{-18}\text{m}$  for an interferometer with 1km long cavities. Their detection implies to track and suppress all of the possible noises.

In Fig. 4.2 we depicted a Mach-Zender interferometer. Mach-Zender interferometers are very useful in this sense that they allow to model most of the different layouts of interferometers. In a Mach-Zender interferometer the optics devices are held fixed and the phase shift is introduced on the path of one of the beams. In this configuration it is also clearer that the vacuum port of the interferometer can play a role. Demkowicz-Dobrzański et al. [2015] wrote a modern review on the topic of quantum limits in interferometers — combining local (based on the FI) and global (Bayesian analysis) approaches. The reader interested by a rather complete analysis of interferometry and quantum optics in general is directed to [Bachor, 2004].

### Two competing noises

As we already mentioned it the use of interferometer for GW detection is based on the free-mass measurement. And we know that QND measurements kept aside, the successive measurement of a free mass is constrained by the SQL. The arguments for deriving the SQL being general they should apply to the metrology with a Michelson interferometer. In such a measurement there exist two fundamental sources of noise in the interferometer: the *shot noise* — also called the photon-counting noise or detection noise — and the *radiation-pressure noise*. The shot noise is due to the particle character of light. The radiation pressure noise is due to the fluctuations in the transfer of momentum from the photons to the mirrors. Obviously the goal is to minimize, if not suppress, all of the noises in the interferometer, and especially these two. But the crucial

#### 4.1 Historical approach of the standard quantum limit and the Heisenberg limit

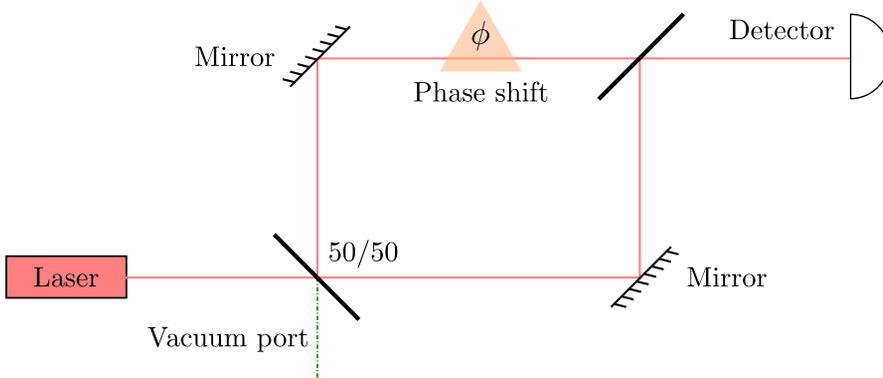


Figure 4.2: Mach-Zehnder interferometer. The Mach-Zehnder interferometer can be used to model most of the different configurations of interferometers. One of the beam suffers a phase shift noted  $\phi$ . Notice also the possibility of using the vacuum port on the interferometer.

point is that the shot noise is inversely proportional to the power of the laser, which means to the average number of photons, while the radiation pressure noise is proportional to the power of the laser. As a result there exists an optimal number of photons in the interferometer as a function of the other parameters. Looking at the minimal phase shift that one can measure when working at the optimal power, and translating this measure of phase shift in a measure of distance, one retrieves the SQL for the free-mass measurement.

This analysis of the SQL in the context of interferometric measurements was carried out in the seventies (see refs. 3,5,8,9 in [Caves, 1980]). While the result looks convincing, the argument has been during many years controversial [Smarr, 1979]. Indeed, all of the noise was believed to have its origin in the laser fluctuations. But if there are fluctuations in the laser it is not clear why these fluctuations should not be distributed evenly between the two arms after the beam splitter. And if they were distributed evenly then the radiation pressure will be indeed fluctuating, but acting on the same way on each mirror causing *in fine* no disturbances in the measurement of the phase shift, and therefore on the mass.

The controversy was solved by Caves [1980] who showed that both noises are intrinsic to the interferometer, and not primarily caused by the laser fluctuations. The proper analysis of the experiment shows that it is the superposition of the fluctuation of the vacuum in the second port and the laser light in the first port which produces such unbalanced fluctuations. Alternatively we can say that it is the scattering due to the beam splitter that provokes the noise. Indeed when  $N$  photons arrive at the beam splitter each of them is scattered randomly, creating a binomial distribution in each arm of the interferometer. It turns out that these two distributions are anti-correlated, which enforces the radiation pressure noise.

#### When the shot noise becomes the SQL ... and when it does not

Caves [1981] also noted that in a typical experiment the optimal laser power is far from being reachable and that as a consequence the radiation-pressure noise would be negligible in comparison with the shot noise. Therefore the sensitivity of the interferometer will not be limited by the SQL but rather by the sole shot noise, which as we have seen scales as  $1/\sqrt{N}$ , where  $N$  is the average number of photons. We want to emphasize that in the context of GW detection with free masses,

this  $1/\sqrt{N}$  scaling does not correspond to a quantum limit as we have been defining them. The quantum limit there is still the SQL which corresponds to the uncertainty obtained with the optimal power of the laser (optimal  $N$ ).

But this is not the full story. Indeed we can also think to use the interferometer not for monitoring free masses but for monitoring a harmonic oscillator, especially the quantum field of light! This brings us outside the field of GW detection, as there the harmonic oscillator should be solid. But the study of SQL for the harmonic oscillator is still general and therefore applies also to the monitoring of a mode of light. An interferometer like the Mach-Zehnder can be used to monitor the phase of the field, and we know that the SQL for the phase of a harmonic oscillator is equal to  $1/(2\sqrt{N})$ . In such interferometer the optics are held fixed and there is no radiation pressure noise. The only fundamental remaining noise is the shot noise, which becomes then a true quantum limit, and, up to numerical prefactors, can be identified with the SQL.

In the literature there is a confusion between shot noise and SQL. Depending on the fields, subfields, authors, years and months the two words have been used indifferently — which as we just saw is not always justified. This is especially true in the field of quantum metrology, where the SQL is sometimes loosely defined as a  $1/\sqrt{N}$  scaling for the sensitivity without much more details. From this point of view we could say that the SQL used in quantum metrology correspond to the SQL for the harmonic oscillator and that the free-mass SQL has been forgotten. In the field of GW detection it seems that people have been more careful and kept the meaning of SQL closer to the original one.

### Beating the SQL: apparition of the Heisenberg limit

In the same paper where he recognized the predominant role of shot noise, Caves [1981] gave also a proposal to overcome the shot noise (we will use the term shot noise here keeping in mind that it is equivalent to the SQL for the harmonic oscillator). In line with his analysis where the noise comes from the vacuum fluctuations in the unused port of the interferometer, he proposed to inject a squeezed vacuum state in the unused port. Despite their name, squeezed vacuum states have in general a finite — potentially large — average number of photons. For a large number of average photons  $N \gg 1$  and optimizing the distribution of the  $N$  photons between the laser light (coherent state) and the squeezed vacuum, one obtains a photon-counting noise scaling as  $1/N^{3/4}$ , which is better than the shot noise scaling  $1/N^{1/2}$ .

The scaling obtained with Caves proposal is still not optimal. Bondurant and Shapiro [1984] investigated the use of squeezed states and considered the use of homodyne detection<sup>2</sup> in interferometers. They used the signal-to-noise ratio (SNR — see Box 8) as a measure of sensitivity. They recover the scaling of Caves ( $1/N^{3/4}$ ) for single mode squeezed states, but most importantly they showed that with homodyne detection the best SNR becomes equal to  $4N(N+1)$ , which for large  $N$  gives a  $1/N$  scaling. This scaling was also retrieved in [Yurke et al., 1986]. It took close to ten years before this limit of  $1/N$  got known as the "Heisenberg limit" [Holland and Burnett, 1993]. The next crucial step was done few years later when Bollinger et al. [1996] showed, using the time-energy uncertainty relation, that the HL is the best limit that can be achieved (see also [Ou, 1997]). They also showed that the HL can be saturated using NOON states, *i.e.*, two-mode states of the form  $(|N, 0\rangle + e^{-i\varphi}|0, N\rangle)/\sqrt{2}$ .

Finally, we should mention that interferometry does not apply only to light. One can also think of atomic interferometers. For example using Ramsey pulses we can design the equivalent

---

<sup>2</sup>In direct detection one measures directly the number of photons while in homodyne detection is done by comparing the output with the the original light entering the interferometer.

## 4.1 Historical approach of the standard quantum limit and the Heisenberg limit

of the Mach-Zehnder interferometer for qubits. It was shown that in this context also the best sensitivity is the HL [Wineland et al., 1992, 1994].

### 4.1.4 Remarks on the shot noise

The discovery of shot noise dates back to the beginning of the XX<sup>th</sup> century. In 1918 Schottky showed theoretically that the noise affecting the electrical current in a vacuum tube has two fundamental origins [Schottky, 1918]. One which is due to the thermal motion of the electrons, and a second one which is directly related to the discrete character of the electric charge, the so-called shot noise. It consists of fluctuations of the intensity of the current which can be modelled by a Poissonian distribution of the probability of transferring an electron from the cathode to the anode (shot noise is also present in p-n junctions, tunnel junctions or Schottky barriers). If we assume that the average number of electrons arriving being detected in a certain interval time is proportional to  $N$ , say  $\alpha N$ , then, the process being Poissonian, the variance is also  $\alpha N$ , and therefore the standard deviation equals  $\sqrt{\alpha N}$ . From this very simple analysis we get an idea of the precision in the estimation of the average intensity through the SNR  $\alpha N/\sqrt{\alpha N} = \sqrt{\alpha N}$ . The SNR increases with the number of electrons, and thus the effect of shot noise will be all the more so than the intensity is low.

This example of a noise raises an important question. What is a noise? The answer depends on what exactly is the signal in which we are interested. If we tackle the problem from a classical point of view, we are interested in measuring the intensity averaged over long times. This can be motivated because we want to estimate physical parameters with a definite relation to the intensity. Then the fluctuations correspond really to a noise — in the sense of a disturbance — which set a bound on the precision with which we can infer any parameter using this signal. From a purely quantum point of view, it is not so clear if these fluctuations should also be considered as a noise. "The noise is the signal" said Landauer. The noise may contain relevant information which we can use to estimate some parameters. In this case the signal could be not the average intensity but the intensity as a function of time. These fluctuations may be useful only if we have a fundamental description of them, meaning that we should have an accurate model of the system, such a way that we can consider them as "the signal". Although the shot noise in electronic systems was introduced long time ago, it took many decades to understand the processes at the origin of this noise, and then giving the possibility to use these fluctuations as a source of information [Beenakker and Schönberger, 2003].

As we have seen shot noise appears also in another branch of quantum physics: quantum optics and photonics. Being related to the particle-like behaviour of the electrons, it is not surprising that shot noise is also present when dealing with photons. Contrarily to the electronic case, the shot-noise limit for photons has been for long understood now, and already in the sixties people noticed that these fluctuations allow one to distinguish between different sources of light.

### 4.1.5 Legacy and limits of metrology for GW detection

This historical review on the two crucial limits encountered in quantum metrology, the SQL and the HL gives raise to important comments. First we should acknowledge the crucial role that the quest of GWs, something that at a first sight is more connected to general relativity or to astronomy and cosmology, played in the modern development of quantum mechanics. From a technical point of view GW detection pushed physicists to build the most sensitive devices ever build on earth. From a theoretical point of view we have seen how some crucial tools and concepts of modern quantum physics can directly be traced back to the effort made for GW detection.

Among them we can cite the fields of QND measurements (*cf.* to Serge Haroche's Nobel prize in 2012), discussions on the exact meaning and formulation of Heisenberg uncertainty relation or continuous measurement.

Secondly we should emphasize that most of the arguments presented in this historical review, whether they are related to the SQL, to QND measurements or to the Heisenberg limit, are kind of weak. Indeed quantum metrology was just in its infancy and retrospectively a lot of concepts appear as dated if not wrong nowadays. For example the SQL is said to be a limit, but at the same time we see that we can break it with QND, showing that it is a limit only because we make strong assumptions (on the kind of measurement, of states, etc) to derive it. We can also point out that all the measurements were described with observables and not POVMs.

Another point lies in the tools used to analyse the sensitivity of the measurement. Mainly the variance of observable was used, and most of the analysis relied on the Heisenberg uncertainty relation, whose exact operational meaning has been debated until recently. It is interesting to see that while the tools of quantum parameter estimation theory already existed at this time (*cf.* [Helstrom, 1976]), it is only in the late nineties that the QFI started to impose itself as the standard measure for determining the sensitivity of a scheme. This offers two fundamental advantages. First it permits to clear the wrong idea that what we can measure should necessary correspond to a quantum observable. This was particularly damageable for the phase as we know that there is no such object as a phase operator. The use of q-pet makes everything clear: all we need is a parameter encoded in a state. The second point is the fundamental character of the QFI. Once that we have the parameter-dependent state we do not need to bother with the measurement nor with the estimator.

## 4.2 Heisenberg limit and standard quantum limit

### 4.2.1 Protocols and resources in quantum metrology

As we just pointed out in the conclusion of the last section, a more robust analysis of quantum metrology should be based on q-pet. To be able to compare different metrological protocols we first have to define them formally. Then, to ensure the comparison between different protocols to be fair, we will have to define a resource count. This is specially important in the quest of defining properly the SQL and the HL, which, as we will see, is still not easy despite of the improved clarity offered by q-pet.

#### Protocols

Quantum parameter estimation theory as presented in the previous chapters provides us with the tools to describe any quantum metrological protocol. In general such protocols can be split into four stages

- 1) preparation of the initial probe  $\rho$ .
- 2) evolution of this probe through a quantum channel  $\mathcal{E}_\theta$  which imprints the parameter  $\theta$  to be estimated into the probe.
- 3) measurement stage done by a POVM  $\{E_\xi\}$ .
- 4) estimation procedure which uses the measurement results.

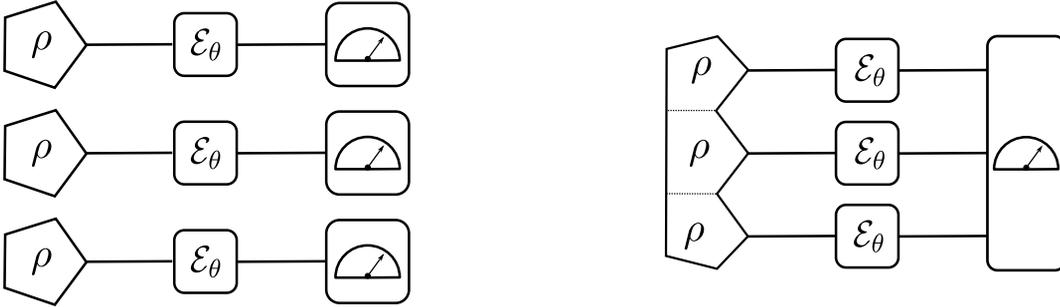


Figure 4.3: Parallel protocol for quantum metrology. In the left figure the probes are in a separable state while in the right figure the probes are in a possibly entangled state.

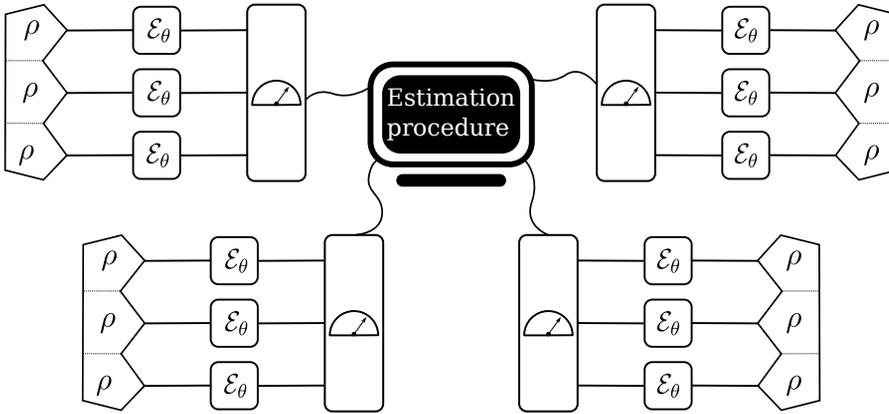


Figure 4.4: Complete quantum metrological protocol with state preparation, evolution, measurement and estimation.

In the previous chapter we said that our operational approach is a two-step approach of quantum metrology: We describe our quantum system and the measurement carried on it to produce a probability distribution and then we use this probability distribution in the framework of parameter estimation (data analysis). In the first chapter we have emphasized that we will work in the framework of the local and frequentist approach, which makes sense in the asymptotic limit of the size of the  $m$ -samples. In our physical estimation theory this means that we have to take the three first steps (1), (2) and (3), and repeat them in an identical and independent fashion  $m$  times. Therefore the situation depicted in Fig. 4.3 is not complete. Instead we should replace it by the situation depicted in Fig. 4.4 and we must add a step in our description of quantum metrological schemes

3') repeat independently  $m$  times the steps 1., 2. and 3.

This extra step insures us that the QFI and the SNR carry an operational meaning. Therefore when we will talk about asymptotic properties, it will be of uttermost importance to remember that this refers to the number of *i.i.d.* repetitions and not to any other typical resources as the number of subsystems or number of sequential repetitions.

It is common in the literature [Giovannetti et al., 2006] to divide the metrological protocols

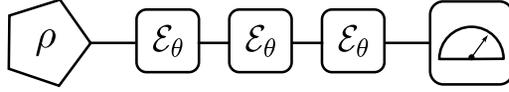


Figure 4.5: Sequential protocol for quantum metrology.

into two branches. On one side we have the parallel protocol, where one applies to a given number of probes the channel in parallel, and on the other side the sequential protocol (see Fig. 4.5) where one considers only one probe and applies a given number of times the channel on it before measuring.

Obviously these two categories are not completely exclusive. We can easily think on a hybrid system where we have more than one probe (parallel aspect) which undergo several time the same channel (sequential aspect). Notice also that in certain cases (unitary evolution for example) the sequential application of channels corresponds to applying the same channel but with a different parameter.

### Defining resources

In order to compare two different schemes we should first agree on the definition of resources, to ensure the comparison to be fair. Due to the huge variety of protocols in quantum metrology it is difficult to make a completely general accounting of resources. We can still try to identify what the common resources taken into account in quantum metrology are, keeping in mind that for specific protocols we will have to identify which one is the most relevant resource.

**Number of probes** The number of probes, usually denoted by  $N$ , is the usual resource considered in quantum metrology. We count only the probes used in one run of the quantum stages, meaning that we do not take into account the independent repetitions. When dealing with systems composed of subsystems the proper way of counting the probe is clear, as for example when considering a collection of qubits. In such case we have a clear tensor product structure of the Hilbert space and the number  $N$  counts the resources. This will be the case when studying the coherent averaging scheme.

The situation gets more complicated when dealing with quantum fields. Suppose that the system considered is an electromagnetic field. It is not clear how to define a probe in such system. We could think on two candidates for this: The number of modes and the number of excitations (photons). We will treat more in details this question when discussing the SQL and the HL in quantum optics.

**Applications of channel** A second common resource considered in quantum metrology is the number of applications of the quantum channel  $\mathcal{E}_\theta$ . In a standard parallel scheme, counting the number of applications of the channel is tantamount to counting probes. The situation gets different when going to a sequential scheme as we apply several times the channel to the same probe.

**Time** The last important resource to consider is the time. As we did for the number of probes we will first consider the running time of one run of the protocol. In most of the situation we do

not consider the time needed to prepare a probe, and only focus on the time of evolution<sup>3</sup>. When dealing with unitary channels the time has a direct meaning since it appears on the evolution operator. Also when working with channels built using master equations the time will usually appear somewhere. When dealing with general channels parametrized in a more mathematical way (meaning that we use a mathematical description of the space of channels rather than referring to a specific physical situation) the time will not appear directly.

**Role of the  $m$  repetitions and the question of scaling** When considering the three resources (probe, channel, time) we focused on the quantum part of the schemes, discarding the  $m$  *i.i.d.* repetitions. Nevertheless a full accounting of the resources should take these repetitions in consideration. It is not difficult to do so: each resource gets multiplied by  $m$ . Although this extra step is trivial in terms of counting resources, it plays a role when thinking about optimality of a protocol. This is closely related to the idea of scaling.

Consider a situation where the QFI is an increasing monotonous function of time. The longer we let the system evolve, the higher is the QFI. The total time is equal to  $T = m \times t$ , where  $t$  is the running time for one evolution. At first sight it may seem equivalent to consider that the resource is  $T$  or  $t$ . But this turns out to be wrong in general. Say that the QFI scales as  $t^\alpha$ . This does not mean that the QFI scales as  $T^\alpha$ . Indeed the QFI will scale as  $m \times t^\alpha$ . From this point of view, if  $\alpha > 1$  the best strategy would be to make only one measurement with running time  $T$  (but then we will not have the  $m$  repetitions that are needed to make the QFI operational). On the other hand if  $\alpha < 1$  then the best strategy will be to make  $t$  as small as possible and to increase  $m$ .

This simple example explains why the scaling is a very important feature in the QFI. In general if we have a scaling which is better than linear, the optimal strategy will be to increase this resource as much as possible in each run. And the limit will be given by technical limitations. Nevertheless we have to keep in mind that the  $m$  repetitions will still be necessary if we want to be sure to saturate the QCRB. If the scaling is lower than linear the best strategy will be to reduce the resources as much as possible in each run, and perform a large number of repetitions.

## 4.2.2 Standard definition of the SQL and the HL

We now agreed on the typical quantum metrological protocols that we will study: A purely quantum part, with the triplet "state+channel+POVM" and its repetition  $m$  times in a *i.i.d.* fashion. This picture is particularly relevant when working with the QFI and the SNR as figures of merit for the metrological efficiency. We we will now try to formalize the concepts of SQL and HL.

In general the SQL and the HL are closely related to the scaling of the sensitivity with a certain resource. The important question is to define which resource we shall use to have a general formalism. A quite reasonable choice will be the number of probes. To avoid ambiguity the number of probes  $N$  will refer to the number of subsystems:  $\mathcal{H} = \bigotimes_{i=1}^N \mathcal{H}_i$ . Then in order to have the freedom to consider an independent or a correlated protocol we must apply the channels independently on each probe. This corresponds to the parallel configuration depicted in Fig. 4.4.

To formalize this we consider a Hamiltonian  $H(\theta)$  acting separately in each subsystems

$$H(\theta) = \sum_{i=1}^N h_i^{(i)}(\theta), \quad (4.5)$$

---

<sup>3</sup>This is not always true. In [Braun et al., 2014b], the authors showed that under certain conditions, photon-subtracted Gaussian states can lead to a divergent QFI although having almost a vanishing average photon number. It is only by taking into account the preparation time that this divergence is cancelled.

where  $h_i^{(i)}(\theta) = \mathcal{I} \otimes \dots \otimes \mathcal{I} \otimes h_i(\theta) \otimes \mathcal{I} \dots \otimes \mathcal{I}$  where  $h_i(\theta)$  is at the  $i$ th position. Due to the fact that the Hamiltonian acts independently on each subsystem, the evolution operator takes the simple form  $U_H \equiv e^{-i t H(\theta)} = \bigotimes_{i=1}^N U_{h,i}$  with  $U_{h,i} := e^{-i t h_i(\theta)}$ . In the same way the local generator  $\mathcal{H}$  (see Sec. 3.6.2) can be written as  $\mathcal{H} = i U_H^\dagger (\partial_\theta U_H) = \sum_{i=1}^N \mathcal{H}_i$ , with  $\mathcal{H}_i := i U_{h,i}^\dagger (\partial_\theta U_{h,i})$  the local generator of the subsystem  $i$ .

### Standard quantum limit

To define the SQL we will investigate the independent scenario. The Hamiltonian has already been chosen to match this criterion. It remains to enforce this condition of independence to the state and to the measurement. Interestingly, according to an argument from Giovannetti et al. [2006], correlations at the measurement stage do not influence the scaling of the result, meaning that joint measurements do not improve the precision. As a result we do not have to restrain the set of possible POVMs and we can thus use the QFI. Considering the pure initial separable state  $|\psi\rangle = \bigotimes_{i=1}^N |\phi\rangle_i$ , the state after evolution is given by  $|\psi(t)\rangle = \bigotimes_{i=1}^N U_{h,i} |\phi\rangle_i$ . Due to the additivity of the QFI we obtain

$$I(|\psi(t)\rangle; \theta) = N t^2 4 \text{Var}[\mathcal{H}_i, |\phi\rangle]. \quad (4.6)$$

The important part is the linear scaling of the QFI with the number of probes:  $I(|\psi(t)\rangle; \theta) \propto N$ . We call this scaling a *SQL scaling*.

We now maximize the QFI over the pure separable states. This amounts to maximizing  $\text{Var}[\mathcal{H}_i, |\phi\rangle]$  over all the possible  $|\phi\rangle$ . To do so we can first notice that the Popoviciu's inequality states that the variance of any random variable  $X$  with upper and lower bound  $b$  and  $a$  respectively is upper bounded by  $(b - a)^2/4$ . Applied to our case we obtain

$$\text{Var}[\mathcal{H}_i, |\phi\rangle] \geq (h_{\max} - h_{\min})^2/4, \quad (4.7)$$

where  $h_{\max}$  and  $h_{\min}$  are the maximal and the minimal eigenvalue of  $\mathcal{H}_i$ . Introducing the superposition of the eigenvectors corresponding to these two eigenvalues  $|\phi(x)\rangle = \sqrt{x} |h_{\max}\rangle + \sqrt{1-x} |h_{\min}\rangle$  we obtain

$$\text{Var}[\mathcal{H}_i, |\phi(x)\rangle] = t^2 x(1-x)(h_{\max} - h_{\min})^2, \quad (4.8)$$

which saturates the bound for  $x = 1/2$ . This shows that the maximal QFI when using separable state is obtained by using the state  $|\phi(1/2)\rangle^{\otimes N}$  and is equal to

$$I(U_H |\phi(1/2)\rangle^{\otimes N}; \theta) = N t^2 (h_{\max} - h_{\min})^2. \quad (4.9)$$

We refer to this QFI as the SQL.

### Heisenberg limit

In contrast to the SQL, where we enforced the independence of the probes, the Heisenberg limit is defined as the QFI obtained when optimizing over the set of pure states, allowing in particular quantum correlations between the subsystems on the initial state. In order to maximize the QFI we use the same method as for the SQL. The maximal and minimal eigenvalues of  $\mathcal{H}$  are  $N h_{\max}$  and  $N h_{\min}$ , and correspond to the eigenvectors  $|h_{\max}\rangle^{\otimes N}$  and  $|h_{\min}\rangle^{\otimes N}$ . Using the entangled state  $|\varphi(x)\rangle = \sqrt{x} |h_{\max}\rangle^{\otimes N} + \sqrt{1-x} |h_{\min}\rangle^{\otimes N}$  as initial state, we obtain for the QFI

$$I(U_H |\varphi(x)\rangle; \theta) = N^2 t^2 x(1-x)(h_{\max} - h_{\min})^2. \quad (4.10)$$

We see that the result is very similar to the one obtain for the SQL, but this time we obtain an  $N^2$  scaling with the resource:  $I(U_H|\varphi(x));\theta) \propto N^2$ . We refer to this scaling as the *HL scaling*. Setting  $x$  to  $1/2$  we obtain the maximal QFI

$$I(U_H|\varphi(1/2));\theta) = N^2 t^2 (h_{\max} - h_{\min})^2. \quad (4.11)$$

We refer to this QFI as the HL.

### Sequential protocol

The choice of the number of subsystems as the resource excludes *de facto* sequential protocols from our analysis of SQL and HL. What would be the interest of discussing the scaling in  $N$  if  $N = 1$ ? In order to circumvent this, we can use as the resource the number of applications  $\tilde{N}$  of the channel  $\mathcal{E}_\theta$ . With this choice we can study in the same way pure parallel schemes (channel applied once and only once to each subsystem), pure sequential schemes but also a mix of them.

To investigate the pure sequential protocol we can use our previous result. We have with only one subsystem,  $N = 1$ , and we apply  $\tilde{N}$  times the channel on this subsystem. The key point is that this is equivalent to applying the channel once but with a duration of  $\tilde{t} = \tilde{N}t$ . Since in the SQL we already have a term  $t^2$  we see that in this scheme the QFI scales as  $\tilde{N}^2$ , corresponding to a HL scaling.

### 4.2.3 SQL and HL in quantum optics

As we have seen the SQL shares a long intricate history with interferometers and quantum optics in general. In the case of the HL the link is even stronger, as the HL was introduced for the first time as the optimal sensitivity offered by an interferometer. Ironically our definition excludes this case due to the difficulty of studying in a unified formalism the metrology for finite and infinitely dimensional systems<sup>4</sup>. The problem is closely related to the definition of resources. As pointed out, for a quantum field we can think of two different fundamental ways of defining the subsystems.

On one hand we are tempted to say that the subsystems are the photons, as particles. The problem with this choice is that our subsystems do not correspond to a tensorial partition of the system when we work with second quantification (Fock space) as it is usually the case. Here a system with  $N$  photons is in general not described by a state of the form  $\bigotimes_{i=1}^N |\psi_i\rangle_i$ . On the other hand we can define the subsystems to be the modes of the light. Then the partition of the system corresponds to the natural tensorial partition of our Hilbert space in second quantification. But then we lose our interpretation of subsystems being the photons.

This problem becomes crucial when discussing the role of entanglement in quantum metrology. A treatment of this question implies to have an extensive look on the way of defining separability for indistinguishable particles, a task whose study would take us too far from the main topic of this thesis (for a review on this topic see [Braun et al., 2017] and [Demkowicz-Dobrzański et al., 2015] for the opposite point of view). As we are not primarily interested in the question of entanglement in quantum metrology the question for us is more a semantic one, which influences the way of defining the SQL and the HL.

Despite the fact that it does not fit in our previously defined formalism, we will consider that the subsystems are the photons. The notions of SQL and HL will not be derived based on the dichotomy separable *versus* entangled but rather classical *versus* non-classical. The notion of

---

<sup>4</sup>For the sake of simplicity we emphasize the case of the harmonic oscillator, and especially electromagnetic fields.

classicality in quantum optics can be defined in different ways. One can use the P-function or in a less formal way rely on the photon distribution. To our concern we have to remember that coherent states are the classical states of light, while any other exotic state like squeezed states or Fock states are truly quantum.

Although we decided to consider photons as our resource, we still have to take in consideration the possibility of having many modes. In general we should consider phase-shift Hamiltonians of the form  $H(\phi) = \phi \sum_i \omega_i \hat{N}_i$  where the sum over  $i$  correspond to the sum over the modes. It turns out that in full generality we can concentrate — at least formally — in one mode and thus to consider the Hamiltonian

$$H(\phi) = \phi \omega \hat{N} . \quad (4.12)$$

### SQL and coherent states

The coherent states of the light correspond to the most classical states encountered in quantum optics. They correspond to the light emitted by a laser way above its threshold. They are usually denoted as  $|\alpha\rangle$  and are parametrized by two real numbers, the real and the imaginary part of  $\alpha$ . In the Fock basis  $\{|0\rangle, |1\rangle, \dots\}$  they are written as

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle = e^{-|\alpha|^2/2} e^{\alpha \hat{a}^\dagger} |0\rangle , \quad (4.13)$$

where  $\hat{a}^\dagger$  and  $\hat{a}$  are the creation and annihilation operators whose action on the Fock states reads  $\hat{a}|n\rangle = \sqrt{n}|n-1\rangle$  and  $\hat{a}^\dagger|n+1\rangle = \sqrt{n+1}|n+1\rangle$ . The coherent states are eigenvectors of the creation operator with eigenvalue  $\alpha$ :  $\hat{a}^\dagger|\alpha\rangle = \alpha|\alpha\rangle$ . The average number of photons in a coherent state is equal to  $N := \langle \alpha | \hat{N} | \alpha \rangle = |\alpha|^2$  with  $\hat{N} := \hat{a}^\dagger \hat{a}$  the photon number operator.

Being the most classical states of the light it is natural to use them to define the SQL. The QFI when starting with a coherent state reads

$$I(e^{-i t H(\phi)} |\alpha\rangle; \phi) = t^2 \omega^2 N . \quad (4.14)$$

We see that using the QFI and the QCRB we retrieve the usual SQL for the phase of a harmonic oscillator in the form  $\text{Var}[\hat{\phi}_{\text{est}}] \geq 1/N$ .

### Trying to define the HL: the problem of the variance

Here again we want to define the HL as the highest QFI that we can reach with our Hamiltonian. And here pops up the second problem that forces us to treat separately the finite dimensional and the infinitely dimensional case: If we try to optimize the QFI for a quantum field we obtain an infinite QFI as the Hamiltonian is not bounded.

A way to circumvent the problem will be to introduce a cut-off. Indeed, we can claim that the energy is necessarily bounded, and introduce a cut-off in the energy. This can be done by either limiting the average energy, or by truncating the basis by discarding states in the basis with an energy above the threshold. We will see that both solutions lead to different types of problems.

When limiting the maximal average energy we can write states as normalized superposition of all Fock states:  $|\psi\rangle = \sum_{n=0}^{\infty} p_n |n\rangle$  with  $\sum_{n=0}^{\infty} |p_n|^2 = 1$ . The limitation is done by imposing the restriction

$$\langle \psi | \hat{N} | \psi \rangle \leq n_{\text{lim}} . \quad (4.15)$$

This means that we consider only the states that have an average energy lower than  $E_{\text{lim}} = \omega n_{\text{lim}}$ . The problem with this cut-off is that we can still construct states with infinite variance. Consider for example the state

$$|\psi_a\rangle = \frac{a|0\rangle + (a^2 + 1)|n_{\text{lim}}\rangle}{\sqrt{a^2 + 1}}. \quad (4.16)$$

The energy of this state is equal to  $E_{\text{lim}}$  but its variance is equal to

$$\text{Var}[\hat{N}, |\psi_a\rangle] = \frac{(n_{\text{lim}}(a^2 + 1))^2}{a^2 + 1} - \left( \frac{n_{\text{lim}}(a^2 + 1)}{a^2 + 1} \right)^2 \quad (4.17)$$

$$= \frac{(n_{\text{lim}}(a^2 + 1))^2(a^2 + 1 - 1)}{(a^2 + 1)^2} = n_{\text{lim}}a^2. \quad (4.18)$$

By making  $a$  arbitrarily large we end up with a diverging variance again, showing that this kind of cut-off is not sufficient to circumvent our problem.

A second possibility — the one that we would adopt here — is to restrain our basis:

$$\{|n\rangle\} \rightarrow \{|n\rangle, n \leq 2n_{\text{c.o.}}\}. \quad (4.19)$$

This is equivalent to bound the Hamiltonian. The effective Hamiltonian is now  $H(\phi)^{\text{c.o.}} = \phi \sum_{n=1}^{n_{\text{c.o.}}} \omega \hat{N}$  and therefore the variance of any state is bounded by  $\phi^2 \omega^2 n_{\text{c.o.}}^2$ . The problem when bounding the operator directly like this is that we exclude some important states. For example we cannot anymore construct coherent states with this basis, which makes the distinction SQL-HL slightly weak from a formal point of view (we use two frameworks that are exclusive). In practice these problems look quite academic as the construction of states with infinite variance and finite energy appears difficult to implement.

Eventually, taking for the sake of notation  $N := n_{\text{c.o.}}$ , we can define the HL through the QFI obtained starting with a ON state [Braun, 2011]

$$I(e^{-i t H(\phi)}; \phi) | \text{ON} \rangle = t^2 \omega^2 N^2, \quad (4.20)$$

where the ON state is defined as  $|\text{ON}\rangle := (|0\rangle + e^{-i\varphi} |N\rangle) / \sqrt{2}$ .

### Multi-mode analysis

We claimed that there was no need in considering many modes for the definition of the SQL and the HL, and that is was enough to focus on only one mode. This can be shown easily by modifying slightly the way we define the cut-off. We start from the Hamiltonian

$$H(\phi) = \phi \sum_i \omega_i \hat{N}_i. \quad (4.21)$$

We can rewrite this Hamiltonian on the form  $H(\phi) = \phi G$  with  $G = \sum_{j=1}^{\infty} E_j |j\rangle \langle j|$  and introduce a cut-off on the basis  $G \rightarrow G^{\text{c.o.}} := \sum_{j=0}^{j_{\text{lim}}} E_j |j\rangle \langle j|$ . Then the HL is defined as the highest QFI achievable which is equal to  $t^2 E_{\text{lim}}^2$  and which is saturated by using the state  $(|0\rangle + e^{-i\varphi} |j_{\text{lim}}\rangle) / \sqrt{2}$ . The key point is to notice that it does not matter how the state  $|j_{\text{lim}}\rangle$  is constructed. It can be by dividing the energy in many modes or by concentrating all the energy in one mode. All what matters is the maximal energy allowed to be use to construct the states.

### 4.3 Quantum enhanced measurements I: Enhancing the sensitivity with complex dynamics

We present in this section some strategies that help to improve the estimation of parameters in quantum systems. We focus on methods that are close to the one studied in Part II and III of the thesis. This is the case with non-linear schemes as much as strategies based on the "power of one qubit" which share some similarities with the coherent averaging scheme presented in Part II.

#### 4.3.1 Non-linear schemes and $k$ -body metrology

In the standard derivation of the HL [Giovannetti et al., 2006] the Hamiltonian is of the form  $\sum_{i=1}^N h^{(i)}(\theta)$ , where  $h^{(i)}(\theta)$  acts only on one subsystem. This means that the Hamiltonian is linear with respect to the number of subsystems. Albeit this has been assumed for the derivation of the SQL and the HL, it is by no means the only kind of Hamiltonian that we can imagine. It is thus interesting to see how the introduction of non-linear Hamiltonians impacts the QFI.

As usual when dealing with quantum metrology it is difficult to write a formalism that encompasses both finite and infinitely dimensional systems. Indeed we still struggle with the fact that for an optical system what is commonly referred to subsystems will be the different modes and not the photons in one mode. Historically non-linearity appeared first for optical systems, and then for solid state systems. Following the chronological order we start by non-linear Hamiltonians in quantum optics.

#### Non-linearity in quantum optics

Optical systems, and among them interferometers, have been the most popular systems in the early study of quantum metrology. In such systems most of the research focuses on the estimation of phase shift. Considering non-linear media leads naturally to non-linear Hamiltonians, with the paramount example of Kerr-type transformations with Hamiltonians proportional to  $(a^\dagger a)^2$ . The formal study of such Hamiltonians was initiated by Luis who found that they lead to a better scaling with the number of photons [Luis, 2004; Beltrán and Luis, 2005].

In these papers the authors did not use the QFI but the SNR with specific observables. As the SNR does not have a fundamental meaning as the QFI, we translate these results into the formalism of QFI. We also introduce the general polynomial Hamiltonian

$$H_k^{\text{ph}}(\theta) = \theta P[a_i \hat{N}^i; k], \quad (4.22)$$

where  $P[a_i \hat{N}^i; k] := \sum_{i=1}^k a_i \hat{N}^i$  is a polynomial of  $\hat{N}$  (photon number operator) of order  $k$ . Notice that for simplifying the discussion, and in line with the original work from Luis [2004] we picked a linear dependence *on the parameter*  $\theta$ , making the Hamiltonian a phase shift Hamiltonian. To study the effect of such Hamiltonians from a metrological point of view we will look at the QFI for  $\theta$  for coherent states and for optimal states.

We start by considering a coherent state  $|\alpha\rangle$ . After evolution we obtain the state  $U_k^{\text{ph}}(\theta)|\alpha\rangle$  with  $U_k^{\text{ph}}(\theta) := e^{-i\theta H_k^{\text{ph}}(\theta)}$ . The QFI is then given by

$$I(U_k^{\text{ph}}(\theta)|\alpha; \theta) = 4t^2 \sum_{i,j=1}^k a_i a_j \left\{ T_{i+j}(|\alpha|^2) - T_i(|\alpha|^2) T_j(|\alpha|^2) \right\}, \quad (4.23)$$

### 4.3 Quantum enhanced measurements I: Enhancing the sensitivity with complex dynamics

where

$$T_i(x) := \sum_{n=0}^{\infty} x^n n^i / n! = \sum_{q=0}^i \left\{ \begin{matrix} i \\ q \end{matrix} \right\} x^q, \quad (4.24)$$

with  $\left\{ \begin{matrix} i \\ q \end{matrix} \right\}$  the *Stirling number* defined as  $\left\{ \begin{matrix} i \\ q \end{matrix} \right\} := \frac{1}{q!} \sum_{p=0}^q (-1)^{q-p} \binom{q}{p} p^i$ . We are mainly interested in the leading order term with respect to  $\alpha$  since we want to check the scaling offered by the usage of this Hamiltonian. The highest order seems to be  $|\alpha|^{4k}$ . But since  $\left\{ \begin{matrix} i \\ i \end{matrix} \right\} = 1$ , this term exactly vanishes. The next leading term is of the order  $4k - 2$  and reads  $4t^2 |\alpha|^{4k-2} a_k^2 \left( \binom{2k}{2} - 2 \binom{k}{2} \right)$ . Using  $|\alpha|^2 = n_{\text{ph}}$  and the fact that  $\binom{2k}{2} - 2 \binom{k}{2} = k^2$  we find the behaviour of the QFI for a large number of photons

$$I(U_k^{\text{ph}}(\theta) | \alpha; \theta) \underset{|\alpha| \gg 1}{\simeq} 4(tk a_k)^2 n_{\text{ph}}^{2k-1}. \quad (4.25)$$

If the SQL is defined as the best sensitivity using classical states of light we are thus left with the result that using non-linear Hamiltonians of order  $k$  the SQL corresponds to the QFI being proportional to  $n_{\text{ph}}^{2k-1}$ .

We now check what is the highest QFI that we can achieve with this Hamiltonian. This amounts in calculating the channel QFI of a phase-shift Hamiltonian, an easy task for pure states (see Sec. 3.6). Given a maximal number  $2n_M$  of photon in the state, the maximal QFI is equal to

$$C(H_k^{\text{ph}}(\theta); \theta) = 4t^2 \sum_{i,j=1}^k a_i a_j n_M^{i+j}, \quad (4.26)$$

whose leading term is  $4t^2 a_k^2 n_M^{2k}$ . If we define the HL as the best sensitivity allowed (remembering that we have to introduce a cut-off in the Hamiltonian) we see that the scaling is equal to  $n_M^{2k}$  to be compared to the  $n_{\text{ph}}^{2k-1}$  scaling offered by coherent states.

#### **$k$ -body Hamiltonians for metrology**

While in quantum optics non-linear Hamiltonians can appear as slightly exotic, in solid state physics, due to the interaction between the particles, non-linear Hamiltonians are the norm. It is thus natural to consider the estimation of the parameter attached to the interaction terms. Boixo et al. [2007] investigated to which extend metrology with  $k$ -body Hamiltonians can lead to an increased sensitivity. Following their study we consider the non-linear Hamiltonian

$$H_k^{\text{nl}}(\theta) = \theta \sum_{i_1=1}^N \cdots \sum_{i_k=1}^{i_{k-1}-1} H_{i_1}^{(i_1)} \otimes \cdots \otimes H_{i_k}^{(i_k)}, \quad (4.27)$$

where  $A^{(i)}$  is an operator acting on the  $i$ th probe and where we did not write the remaining identity operators for the sake of concision. Notice that in comparison with the non-linear Hamiltonian (4.22) that we used for optical systems, here only the term of maximal order is kept. In the original paper the authors take into consideration also other terms in the Hamiltonian (they use a Hamiltonian of the form  $H_k^{\text{nl}}(\theta) + \tilde{H}$ , where  $\tilde{H}$  includes the free Hamiltonian of the particles as well as possible interaction of a lowest order), and then give only an upper bound to the QFI. We will come back to this topic later on as this is at the base of our proposal for Hamiltonian extensions<sup>5</sup> presented in Chapter 9. When all the systems and their Hamiltonians are identical,

<sup>5</sup>Notice also that even if the authors went further from the standard linear Hamiltonian, they still considered a phase-shift Hamiltonian. We will see how getting rid of this assumption allows us to consider useful Hamiltonian extensions.

we have  $H_{i_r} = H$ . Then the optimal QFI is given by [Boixo et al., 2007]

$$I(e^{-i t H_k^{n_1}(\theta)} |\psi_{\text{opt}}\rangle; \theta) = t^2 \binom{N}{k}^2 \frac{(h_{\max} - h_{\min})^2}{4}, \quad (4.28)$$

where  $h_{\max}$  and  $h_{\min}$  are the maximal and minimal eigenvalue of  $H$ , with corresponding eigenvectors  $|h_{\max}\rangle$  and  $|h_{\min}\rangle$ . For  $N \gg k$  we can approximate the binomial coefficient as  $\binom{N}{k} \simeq n^k/k!$ . Doing so the leading term in  $N$  is then  $\frac{t^2}{(k!)^2} \frac{(h_{\max} - h_{\min})^2}{4} N^{2k}$  which scales as  $N^{2k}$ . In line with our analysis of Hamiltonian parameter estimation for phase shift (see Sec. 3.6) we know that the optimal state is the entangled state  $|\psi_{\text{opt}}\rangle = (|h_{\max}\rangle^{\otimes N} + e^{-i\varphi} |h_{\min}\rangle^{\otimes N})/\sqrt{2}$ .

We now turn to the study of separable states. Starting with the separable pure state  $|\psi_N\rangle = |\psi\rangle^{\otimes N}$  the QFI becomes

$$I(e^{-i t H_k(\theta)} |\psi_N\rangle; \theta) = t^2 \left\{ \langle H \rangle^{2k} \binom{N}{k} \left( \binom{N-k}{k} - \binom{N}{k} \right) + \sum_{i=1}^k \langle H \rangle^{2(k-i)} \langle H^2 \rangle^i \binom{N}{i} \binom{N-i}{k-i} \binom{N-k}{k-i} \right\}, \quad (4.29)$$

with  $\langle H \rangle = \langle \psi | H | \psi \rangle$ . We can check that the case  $k = 1$  gives back the usual SQL for Hamiltonians without interaction, *i.e.*,  $N \text{Var}[H, |\psi\rangle]$ . Using again the approximation of the binomial coefficient for large  $N$  we find for the leading order of the QFI with separable states

$$I(e^{-i t H_k(\theta)} |\psi_N\rangle; \theta) \underset{|\alpha| \gg 1}{\propto} N^{2k-1}. \quad (4.30)$$

When  $k = N$  the QFI takes the simple form  $4(\langle \psi | H^2 | \psi \rangle^N - \langle \psi | H | \psi \rangle^{2N})$ . If we use the state  $|\psi\rangle = (|h_{\max}\rangle - |h_{\min}\rangle)/\sqrt{2}$  then the QFI becomes  $(\frac{h_{\max}^2 + h_{\min}^2}{2})^N - (\frac{h_{\max} + h_{\min}}{2})^{2N}$ . The origin of the energy having no physical meaning, we can always shift the energy such that  $h_{\max} = -h_{\min} = \tilde{h}$ . Then the QFI is equal to  $\tilde{h}^{2N} = e^{N \ln(\tilde{h}^2)}$  showing that the QFI scales exponentially with the number of subsystems! This result is in line with the results of Roy and Braunstein [2008] (which actually inspired the work of Boixo et al. [2007] — the apparent violation of causality in the publication dates being due to the publication of preprints). There the authors studied the Hamiltonians  $\theta K$  and  $\theta G$  defined implicitly through  $\bigotimes_{i=1}^N (\sigma_x - i \sigma_y) = K + i G$ . They showed that the separable states  $|0 \cdots 0\rangle$  or  $|1 \cdots 1\rangle$  maximize the variance of  $K$  and  $G$  and proved that this maximal variance is equal to  $2^{N-1}$ .

### 4.3.2 More on the SQL and the HL

#### Problems with the definition of the HL and the SQL

We have seen in the historical review how the term HL was used for first time in the mid nineties. Holland and Burnett [1993] used this term in the specific framework of interferometry: "*A possible mechanism for improving the sensitivity is to drive the interferometer with nonclassical states of light as the  $1/\sqrt{N}$  level of relative phase fluctuations for a coherent source is well above the Heisenberg limit of  $1/N$  rad*". Then this term got popularized, and people started to use it whenever a term  $N^2$  appeared for the calculation of a sensitivity (measured by the QFI or the SNR), without taking so much care of what  $N$  exactly was. In the same way the SQL was loosely define as a QFI proportional to  $N$ . In the Sec. 4.2.2 we formalized these two limits in a

### 4.3 Quantum enhanced measurements I: Enhancing the sensitivity with complex dynamics

way that could appear general enough (although we had to separate finite-dimensional systems from infinite-dimensional ones). The SQL was defined as the maximal QFI when starting with separable states (or coherent states for a quantum field). At the same time the HL was presented as the optimal sensitivity that one should target.

With non linear systems we can achieve way better scaling than the  $N^2$  scaling, namely scaling  $N^{2k}$  with  $k$ -body metrology. With the loose definition of the HL plus the idea that it is the best scaling, these sub-HL scaling appeared as an incredible enhancement. But it is not clear why we should compare the sensitivity obtained with a non-linear Hamiltonian to the sensitivity obtained using a linear one. *A priori* the two parameters that will be estimated carry a very different physical meaning. It is true that in some cases the same parameter may appear in two different terms of the Hamiltonian (say the free Hamiltonian and the two-body interaction). In this case the comparison will be valid, but then one should not rely only on scaling but on the precise details of the model before claiming any advantage.

In our opinion what should be compared is rather the sensitivity obtained using different states with the same Hamiltonians. As we have seen, from this point of view non-linear Hamiltonians lead to the same conclusion as linear ones: There exists an advantage by a factor  $N$  when using entangled states instead of separable ones.

#### Query complexity and average energy

To avoid these problems one should define a universal resource count, whose definition is independent of the details of the metrological scheme studied. In [Giovannetti et al., 2006] (and more formally introduce in [Zwierz et al., 2010, 2012]) was introduced the use of query complexity as a resource count. Query complexity, denoted  $Q$ , originates in quantum computing, and is easily visualized using the quantum circuit representation of quantum computing. Any metrological scheme can actually be represented as a quantum circuit. In this context, an important class of gates are the black boxes (also called oracles). A black box is a quantum gate whose internal behaviour is unknown but which acts always in the same way on the states. This definition matches our requirement of metrological situation in the frequentist point of view, where the black box represents the unitary channel that imprints the parameter. The query complexity is defined as the number of uses of the black box in the circuit.

We see that this choice of resource count is close to the "number of applications of the channel". The difference is that with query complexity one goes back to the fundamental interactions that imprint the parameter, the cost being that this approach works only with unitary gates, namely only for Hamiltonian parameters. In [Zwierz et al., 2012] it was shown that for Hamiltonians of the form  $H_k(\theta) = \theta \sum_{i_1=1}^N \cdots \sum_{i_k=1}^{i_{k-1}} H_{i_1} \otimes \cdots \otimes H_{i_k}$  the QFI scales as  $I(H_k(\theta); \theta) \propto Q^2$ . We see that with this resource count we can define in a quite general fashion the HL scaling as a  $Q^2$  scaling in the QFI. The problem is that in some situation the query complexity is ill defined. This is for example the case when dealing with quantum fields.

#### 4.3.3 Power of one qubit

We now turn our attention to metrological protocols inspired by the field of quantum computation. The most renowned architecture for quantum computers is based on the use of  $q$  qubits initialized in a pure state. Using a finite set of 1-qubit and 2-qubits gates, standard quantum computers can lead to an exponential speed-up in comparison to a classical computer [Nielsen and Chuang, 2011]. In general it is not possible to keep the qubits pure during the computation, errors will affect the qubits. To fight those errors, strategies based on the use of extra qubits and known as quantum

error correction have been proposed (such methods find application in quantum metrology too, see Sec. 4.6.2). Knill and Laflamme [1998] proposed an architecture for quantum computers using only one qubit in a pure state and all the others in a completely mixed state, known as the DQC1 (Deterministic Quantum Computation with one qubit) protocol. Although less powerful than standard quantum computers, DQC1 protocols seem<sup>6</sup> to allow an exponential speed-up for some tasks. The circuit b) in Fig. 4.6 shows for example a DQC1 protocol used for calculating the renormalized trace ( $\text{tr}[U] / 2^n$ ) of an operator  $U$  [Datta et al., 2005]. To obtain the speed up with this algorithm it is actually enough to have a control qubit with a non-vanishing purity [Datta et al., 2005]. DQC1 protocols are especially relevant for nuclear magnetic resonance in liquid systems, since there one uses thermal qubits which are then highly mixed [Jones, 2001, 2011].

The idea of using an assembly of mixed qubits along with one pure control qubit (or at least with non-vanishing purity) can also be applied to quantum metrology. Taking the circuit b) in Fig. 4.6 and using a gate  $U = U_\theta = e^{-i t \theta H}$  (black box), one obtains by measuring only the control qubit a variance for the estimator of  $\theta$  scaling as  $\text{Var}[\hat{\theta}_{\text{est}}] \propto 1/(nt^2)$ , which is the SQL. Boixo and Somma [2008] proposed a protocol to achieve this scaling using Bayesian rules and adaptive measurements.

Cable et al. [2016] designed a more general protocol (see circuit a) in Fig. 4.6), also based on DQC1, to estimate a phase-shift parameter. Apart from the usual qubit control and the set of  $n$  completely mixed qubits, there are in addition  $m$  partially mixed qubits with purity  $\varepsilon$  and  $l$  other pure qubits. Crucially it is still only on the control qubit that measurements are performed. The QFI for this scheme is equal to

$$I_{\text{DQC1}} = n + m(1 - \varepsilon^2) + (1 + l + \varepsilon m)^2. \quad (4.31)$$

This QFI is to be compared to the one obtained if we applied the gates  $U_\theta$  independently to each qubit, which reads  $1 + n + l + m$ . Importantly the authors showed that one can reach the QFI by performing adaptive measurements only on the control qubits (*c.f.* the  $V_r$  gates in the circuit). The analysis of  $I_{\text{DQC1}}$  shows the role of every kind of qubits used in the protocol:

- the  $n$  completely mixed qubits produce an SQL term.
- the  $m$  partially mixed qubits contribute with a HL scaling proportional to their purity and a SQL scaling which decreases with increasing purity.
- the  $l$  pure qubits contribute with a HL scaling.

## 4.4 Quantum enhanced measurements II: Ancilla assisted metrology

Up to here we have seen two quantum enhanced measurement schemes. First, through the very definition of the SQL and the HL we saw how the use of entanglement between probes for bounded Hamiltonians typically increases the QFI in comparison to an *i.i.d* scheme. Then we also saw how we can use more complex dynamics to increase the sensitivity. Although our formal description of a quantum metrological protocol was based on general quantum channels  $\mathcal{E}_\theta$ , the quantum enhanced

---

<sup>6</sup>"seems", because for the tasks that are solved in polynomial time with DQC1 there is no classical algorithm *known* that scales also in polynomial time. To claim an actual exponential speed-up one should show that there exists no classical algorithm solving the task in polynomial time.

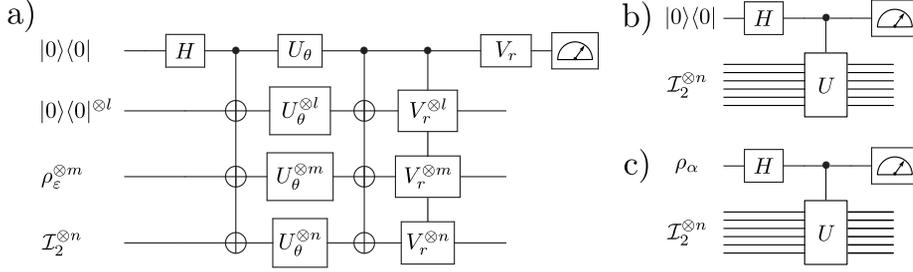


Figure 4.6: a) Metrological protocol based on the "power of one qubit" [Cable et al., 2016]. The first qubit is the measured qubit. The protocol includes also  $n$  other pure qubits,  $m$  partially mixed qubits (purity equal to  $\varepsilon$ ) and  $l$  totally mixed qubits. The preparation stage is composed of Hadamard gates and of C-not gates. Then comes the evolution stage with the  $1 + n + m + l$  black boxes  $U_\theta$  and finally the measurement procedure is done through an adaptive method ( $V_r = e^{-i\theta_r \sigma_z}$  where  $\theta_r$  is the estimate after  $r$  runs) allowing ones to reach the QFI by measuring  $\sigma_x$  on the control qubit. b) DQC1 protocol for computing the trace of the operator  $U$ . c) Generalized DQC1 protocol: the control qubit has only purity  $\alpha$ :  $\rho_\alpha = (\mathcal{I}_2 + \alpha Z)/2$ .

measurement that we saw only used unitary channels. This can be justified (and attacked) by the fact that unitary evolution corresponds to a non-noisy evolution, and it is always important to understand properly the ideal case. So what happens if we want to also consider noisy evolutions?

There are two answers to this question, the duality being in a certain fashion related to the discussion of noise raised in the context of interferometry. First we can consider that the noise is the signal. Consequently the metrological task becomes the estimation of the parameter  $\theta$  in  $\mathcal{E}_\theta$ , taken as non-unitary. This is directly in line with our formal description of a quantum metrological protocol. This approach is quite natural in quantum information. Indeed systems are always affected by noises, and if we cannot suppress them we should at least be able to identify them properly.

A second paradigm would be to use quantum channels to introduce a disturbance in the system. We start from a standard situation where one is interested in the estimation problem of a channel — unitary or not —  $\mathcal{E}_\theta$  and we then include a second channel  $\Gamma$  that plays the role of the noise, acting before or after the original channel. This still fits our formal definition of quantum metrological protocol with the new channel  $\Gamma \circ \mathcal{E}_\theta$  or  $\mathcal{E}_\theta \circ \Gamma$ . In this section we focus on the first case, where noise is the signal, postponing the study of how noise affects estimation to the end of the chapter.

#### 4.4.1 Channel extension for quantum metrology

To extend quantum metrology to general channels we can learn from the theory of channel estimation developed in Sec. 3.5. Although we developed the framework of channel extension in a formal context, it naturally carries a simple operational meaning, as the extension corresponds to the use of an ancilla on which we act with the identity, from there the term of ancilla assisted metrology (Fig. 4.7).

There is in particular an equation that shows how the channel extension corresponds to a quantum enhanced strategy:

$$C(\mathcal{E}_\theta(\rho); \theta) \leq C(\mathcal{E}_\theta \otimes \text{Id}; \theta). \quad (4.32)$$

This means that, *in terms of channel QFI*, extended channels cannot be worse than original

channels. Then, when considering the optimal case, we can trivially say that allowing the use of an ancilla cannot decrease the sensitivity. This is quite clear as we can always start with a separable state, and thus both schemes are equivalent. But way more interesting is the fact that we already said that usually this inequality is not tight. Meaning that the extended channel QFI is in general higher than the original channel QFI, leading in the context of metrology to a genuine improvement in the sensitivity.

In terms of enhancement we should actually go one step back, from channel QFI to QFI. In the context of parameter estimation we have been pushing in the direction of optimization, over POVMs first and then over input states. This was done in order to obtain figure of merits that are uniquely given by a physical object (state or channel respectively). But in terms of metrology this is not always desirable as it may be extremely hard to implement. Sometimes it is desirable to not look at a too optimised figure of merit. Ironically in metrology, if one wants to propose an experimentally achievable protocol, one needs to travel the way back: consider a specific input state (from channel QFI to QFI), a specific POVM (from QFI to FI) and even a specific estimator (from FI to the variance of this specific estimator). Going back to the ancilla assisted schemes this means that even if in some case the inequality (4.32) is tight, it does not mean that ancillas are useless. They can still lead in some non-optimal but realistic case to an improvement.

#### 4.4.2 Enhancing the sensitivity with ancillas

We just saw how what we learn in channel estimation theory is directly transposable to design quantum enhanced protocols. Historically, the flow was the other way around: People started to propose enhanced schemes based on the use of ancillas which created the need of developing a proper framework. The idea of using ancillas entangled with the probe had already a long history in quantum information before being applied to quantum metrology. An important source of inspiration at this time was the EPR paradox. Bell inequalities and Aspect experiments showed the true non-local character of correlation in quantum theory. As it was clear that such correlations could not be used for establishing supraluminal communication physicists have been looking to which task correlations could benefit.

An important discovery in this direction was done by Bennett and Wiesner [1992] with the so-called superdense coding protocol. Superdense coding allows one to share two bits of information by sending (but not *using*) only one qubit. The idea is as follows: An EPR pair (Bell state) is created and shared among two users for some reason called Alice and Bob. If Alice wants to send two bits of information to Bob she can apply a transformation to her part of the state and send it back to Bob. By an appropriate reading procedure Bob is able to retrieve the message (either 00, 01, 10 or 11).

Another application of entanglement in information theory, concomitant of ancilla assisted

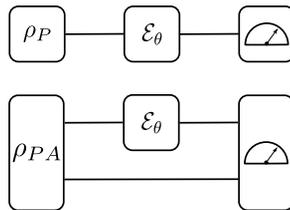


Figure 4.7: Ancilla assisted metrology.

metrology, and closer to it than superdense coding, was developed at the beginning of the century. Acín [2001] studied how well we can distinguish between two unitary operations  $U_1$  and  $U_2$ . We know that distinguishing between two states that are non-orthogonal requires an infinite number of copies. The authors showed that in order to distinguish between two unitary operations there exists a finite number of copies for which we can perfectly distinguish them. Crucially, achieving this result requires the use of the extended channels  $U_i \otimes \mathcal{I}$ . This result was also found in [D'Ariano et al., 2001], where the authors were mainly interested in the estimation of a unitary evolution using covariant measurements. We see that much of the essence of metrology is present here, with the main difference that we try to distinguish objects separated by a finite distance while in metrology the objects are parametrized and infinitesimally close. In a similar way, it was shown at this time that tomography of quantum unitary channels can be optimized by using ancillas [Acín et al., 2001].

Eventually, quantum enhanced protocols based on entanglement with ancillas studied in the modern framework of quantum parameter estimation appeared at the same epoch. The fundamental paper [Fujiwara, 2001b] was probably the first one to tackle the problem of estimation of a non-unitary channel with an ancilla assisted scheme. In this "Quantum channel identification problem" Akio Fujiwara studied the estimation of a depolarizing channel<sup>7</sup>  $\mathcal{E}_{\text{dep}}(p)$ , acting on a qubit (see Sec. 3.5). He was able to identify different regimes of optimality depending on the value of  $p$ . This was because he did not consider the number of invocations of the channel as the primary resource for comparison but rather the number of qubits used. Then for different values of  $p$  it turned out that it is better to use either (i)  $\mathcal{E}_{\text{dep}}(p) \otimes \text{Id}$  on a maximally entangled state or (ii)  $\mathcal{E}_{\text{dep}}(p) \otimes \mathcal{E}_{\text{dep}}(p)$  on a maximally entangled state, or (iii)  $\mathcal{E}_{\text{dep}}(p) \otimes \mathcal{E}_{\text{dep}}(p)$  on a separable state. In most of the subsequent works on channel estimation, the primary resource considered became the number of invocations of the channel. Indeed being the unknown object, it is most of the time the channel and not the state that is supposed to be the costly and limiting resource. By taking this approach it turns out that it is only the strategy (i) that is optimal for all values of  $p$ . In [Fujiwara, 2001a] the same author looks at a different channel identification problem. Inspired by superdense coding, he investigated the estimation of an  $\text{SU}(2)$  operation acting on a qubit in a multi-parameter setting. Also here it turns out that the best strategy is to use a maximally entangled state. But the same result does not have necessarily the same cause: Here it is necessary to use an ancilla due to the number of parameters to estimate. Full characterization of the dynamics requires three parameters and only two parameters can be imprinted in a single pure qubit. The vector character of the estimation problem is crucial here to explain the improvement achieved by using an ancilla.

Finally let us make a remark on the number of ancillas one should use. From the point of view of the QFI it turns out that using only one ancilla — with a Hilbert space of the same dimension as the Hilbert space of the probe, say  $d_1$  — is always enough. Imagine we want to use many ancillas, or equivalently one ancilla with a Hilbert space with dimension larger than  $d_1$ . Thanks to the theorem 2.3 we know that the state of the total system can always be written as  $|\psi\rangle = \sum_{i=1}^{d_1} p_i |\alpha_i\rangle \otimes |\beta_i\rangle$ . This means that we can use an effective ancilla with dimension equal to  $d_1$  and that increasing the dimension cannot help to increase furthermore the QFI.

### 4.4.3 Ancilla assisted with more than one probe

In the line of the founder paper on ancilla assisted metrology a lot of work has focused on the effect of ancillas on the estimation of quantum channels. Being the simplest model in quantum

---

<sup>7</sup>Notice that Fujiwara defines the depolarizing channel for  $p \in [-1/3, 1]$  while we define it only for  $p \in [0, 1]$ .

information theory, qubits have been a favored model for investigations [Fujiwara, 2004; Fujiwara and Imai, 2003]. Early results have been focusing on physically important channels, like the generalized amplitude damping or Pauli channels. Pauli channels play a very important role in information theory since the three fundamental errors affecting qubits in computation are specific cases of Pauli channels (see Sec. 2.6.3). Since two of them will be the scope of Chapter 8 we postpone their study to it.

Up to now the examples of ancilla assisted schemes were based on protocols involving one probe. But as we know in quantum metrology there can be enhancement by using several probes in parallel and entangling them. We can extend this idea to ancilla assisted schemes to arrive at a very general protocol: We start with a joint state  $\rho$  of  $N$  probes and  $N$  ancillas ( $\rho \in \mathcal{B}(\mathcal{H}^{\otimes 2N})$ ). We then apply the channel independently on each probe and act with the identity on the ancillas ( $(\mathcal{E}_\theta \otimes \text{Id})^{\otimes N}$ ). The resulting state is then treated as usual (measurement and estimator).

With the possibility of using multiple probes in parallel comes again the discussion of the scaling with the resources. When using the ancilla assisted protocol only with independent probe-ancilla subsystems —*i.e.*, for the case of a pure state, starting with  $|\psi\rangle = \bigotimes_{i=1}^N |\psi_{i,i+N}\rangle_{i,i+N}$  where  $|\psi_{i,i+N}\rangle_{i,i+N}$  is a pure bipartite state of the  $i$ th probe and the  $i$ th ancilla (which is the  $i + N$ th subsystem) — the channel QFI is always proportional to  $N$ . A very important result from Fujiwara and Imai [2008] gives a condition under which a channel allows an estimation with an  $N^2$  scaling:

**Theorem 4.1 — SCALING FOR CHANNEL ESTIMATION.**

Consider a quantum channel  $\mathcal{E}_\theta$ . Then the channel QFI for the extended channel used in a parallel setting with  $N$  probes scales at most as  $N^2$ :

$$C((\mathcal{E}_\theta \otimes \text{Id})^{\otimes N}; \theta) \leq \mathcal{O}(N^2). \quad (4.33)$$

Moreover if there exists a  $q$ -Kraus decomposition  $\{A(\theta)\}$  of  $\mathcal{E}_\theta$  that fulfils

$$\sum_{k=1}^q \dot{A}_k(\theta)^\dagger A_k(\theta) = 0, \quad (4.34)$$

then the channel QFI scales linearly with the number of probes:  $C((\mathcal{E}_\theta \otimes \text{Id})^{\otimes N}; \theta) = \mathcal{O}(N)$ . If in complement of (4.34) the channel QFI obeys

$$C(\mathcal{E}_\theta \otimes \text{Id}; \theta) = 4 \left\| \sum_{k=1}^q \dot{A}_k(\theta)^\dagger \dot{A}_k(\theta) \right\|_\infty, \quad (4.35)$$

then we have  $C((\mathcal{E}_\theta \otimes \text{Id})^{\otimes N}; \theta) = N C(\mathcal{E}_\theta \otimes \text{Id}; \theta)$ .

In the same paper, Fujiwara and Imai [2008] showed an important results for full-rank channels<sup>8</sup>:

**Theorem 4.2 — SCALING OF FULL RANK CHANNELS.**

Any full-rank channel  $\mathcal{E}_\theta^{\text{full}}$  leads to a channel QFI scaling linearly with the number of probes  $C((\mathcal{E}_\theta^{\text{full}} \otimes \text{Id})^{\otimes N}; \theta) = \mathcal{O}(N)$ .

This theorem has direct application when studying the channel estimation for qubits. The geometry of these channel has been extensively studied and nice geometrical representation of the

---

<sup>8</sup>Consider a channel acting on the bounded operator  $\mathcal{B}(\mathcal{H})$  where the dimension of  $\mathcal{H}$  is equal to  $d$ . Then a full-rank channel is a channel that cannot be written with less than  $d^2$  Kraus operators.

set of channel are known. When going to unital channels this theorem tells us for example that depolarizing channel has a linear scaling with the number of probes (in general all of the channels lying inside the tetrahedron of the Fig 2.2 will have a linear scaling).

Theorem 4.1 is extremely insightful, as it shows that the maximum scaling with the number of probes, when applying only once the channel to each probe, is equal to  $N^2$ . This is completely general and one may ask why we do not define the SQL and the HL using this formalism? Indeed the formalism is sound, and the theorem shows that there cannot be a better scaling than  $N^2$ . The problem is that this formalism does not allow one to define in a convenient way the subsystems. Its basic ingredient is the channel  $\mathcal{E}_\theta$  which acts on  $\mathcal{S}(\mathcal{H})$ . The division in subsystems is then enforced by the choice of channel. Typically, if we want to analyse a  $k$ -body Hamiltonian within this framework we cannot use as the resource the number of particles, rather we have to say that the total system corresponds to one resource.

## 4.5 Correlations in quantum metrology

In our discussions on quantum limits and on enhancement there is one concept that came back again and again: independence. It is for example based on the concept of independence that we constructed the opposition SQL *versus* HL. This shows the importance of the role of correlations in quantum metrology (actually in all the fields of quantum information). As correlation, and especially entanglement, is a resource difficult to generate, understanding if entanglement is present and necessary or not in a given quantum enhanced scheme is an important question. We will thus study to which extent correlations are needed in quantum metrology. In line with our study of quantum enhanced measurements we will first look at the correlation between the probes and then look at the correlation between probes and ancillas.

### 4.5.1 Entanglement between probes

The question of the use of entanglement among probes is at a first sight the simplest to deal with. By the very definition of the SQL and the HL we see the role played by entanglement in Hamiltonian parameter estimation. Indeed, the SQL is defined as the highest QFI obtained using separable states while the HL is defined as the highest QFI using arbitrary states, and especially entangled ones. And we have seen that for both linear and non-linear Hamiltonians the use of entangled states provides an advantage by a factor  $N$  in the QFI. Notice that the example of non-linear metrology is also used to claim that we can achieve a quantum enhancement without entanglement [Datta and Shaji, 2012]. Indeed if one compares the HL for a linear Hamiltonian (so using an entangled state) to the SQL for a non-linear Hamiltonian (so using a separable state) we obtain an increase of the QFI by a factor  $N^{2k-1}$ . As we already emphasized, this comparison is at least questionable. For a complete discussion on the use of entanglement in quantum metrology see [Braun et al., 2017].

When going further than Hamiltonian parameter estimation the study of the use of entanglement between the probes becomes more difficult. Formally we can use the theorems 4.1 and 4.2 to get some information about the scaling. If this scaling is linear we know that entanglement between the subsystems can only improve the QFI by a constant factor. Notice that this method suffers from the drawback of these theorems: In this formalism the subsystems are defined as the system to which we apply the channel  $\mathcal{E}_\theta$ . If these systems are divided in subsystems also, then these theorems do not help to discuss the entanglement between these subsystems.

We should also say a word about the question of entanglement in sequential protocols. Since we have only one probe in a pure sequential scheme it is hard to discuss entanglement. Still we have seen that in a sequential protocol for Hamiltonian parameter estimation we also obtain a scaling  $N^2$  with the number of application of the unitary channel. In this context Maccone [2013] showed how we can use entanglement to transform a parallel strategy to a sequential one.

### Box 13: Detecting entanglement with the QFI

Entanglement can play a key role in quantum metrology for enhancing the precision. It is nevertheless not the only way to build a bridge between entanglement and metrology. Indeed, if in some specific cases enhancing the precision *requires* entanglement, then this enhanced precision *witnesses* entanglement.

Pezzé and Smerzi [2009] showed that in order to detect entanglement in an assembly of  $N$  qubits we can use the QFI. We define a fictitious angular momentum  $J_{\mathbf{n}} = \sum_{i=1}^N \sigma^{(i)}$ , where  $\sigma$  is a Pauli matrix pointing in the  $\mathbf{n}$  direction. Defining the quantity  $\chi^2 := N/I(e^{-i\theta J_{\mathbf{n}}} \rho_{\text{in}} e^{i\theta J_{\mathbf{n}}}; \theta)$ , sufficient condition for entanglement is  $\chi^2 < 1$ . This condition is equivalent to say  $I(e^{-i\theta J_{\mathbf{n}}} \rho_{\text{in}} e^{i\theta J_{\mathbf{n}}}; \theta) > N$ , *i.e.* when the QFI is larger than the SQL. This witness can be related to witnesses using spin squeezing [Tóth, 2012]. We see that in general if one defines the SQL as the best QFI obtainable with separable states, the condition of having the QFI larger than the SQL is a sufficient condition for entanglement. One can also go beyond this simple analysis. Hyllus et al. [2012] found a sufficient condition involving the QFI to detect  $(q+1)$ -particle entanglement<sup>a</sup> in the context of two-modes interferometer. One can also use a set of local variables  $\{A^{(i)}\}$  to witness entanglement [Gessner et al., 2016]: a state  $\rho$  is entangled if  $I(\rho; \sum_{i=1}^N A^{(i)}) \leq 4 \sum_{i=1}^N \text{Var}[A^{(i)}, \rho]$ .

In [Strobel et al., 2014], a way of calculating the FI without tomographic reconstruction was investigated and demonstrated experimentally. The method uses the fact that the FI corresponds to a metric in the parameter space. By performing a measurement and then varying the state one can calculate the Heilinger distance between the probability distributions. By taking the curvature of the square of the distance we recover the FI. The experimental setup is based on an array of binary Bose Einstein condensates of  $^{87}\text{Rb}$  and the authors were able to detect non-Gaussian entangled states.

<sup>a</sup> A  $(q+1)$ -particle entangled state is a state that is  $q+1$  producible but not  $q$  producible. A pure state is  $q$  producible when it can be written as a tensor product of states that are all state with less than  $q$  particles, meaning that  $|\psi\rangle = \bigotimes_{i=1}^m |\psi_i\rangle$  where all the  $|\psi_i\rangle$  are state of at most  $q$  particles.

## 4.5.2 Discord for metrology

### Discord

Through the **XX**th century it was thought that non-classical correlations were synonym of entanglement. At the beginning of the new century the situation changed: The advances in the study of correlations in quantum mechanics lead to the discovery of a new kind of quantum correlations, called discord. Henderson and Vedral [2001] characterized the classical, the quantum, and the total correlations in a bipartite system. They defined total correlation by generalizing to the quantum world the classical mutual information. Entanglement was defined as the entropy between the

state under consideration and the nearest separable state. The real advance lies in their definition of classical correlations. By giving a set of four axioms they defined a measure of classical correlation, and interestingly they observed that for some states, "the whole is greater than the sum of the parts": The addition of classical correlation and entanglement does not reproduce the total correlation. Ollivier and Zurek [2001] made a similar observation and introduced the term of discord. They took two different quantum generalizations of classical mutual information, one being the obvious generalization and the other corresponding to the classical information as defined by Henderson and Vedral [2001]. The discord is then defined as the difference between the total correlation and the classical correlation. For pure states discord is equivalent to entanglement, but for mixed states, separable states — by definition non entangled — can have a non-vanishing discord, showing that there exist genuine quantum correlations in mixed separable states (a recent review on quantum discord can be found in [Modi et al., 2012]).

### Correlation and Power of one qubit

When noticing that quantum computers were able to beat classical computers, researchers naturally tried to understand what the physical mechanism behind this speed-up was. It was shown that entanglement is necessary, qualitatively and quantitatively. Qualitatively in the sense of the presence of entanglement over an adequate large number of qubits, and quantitatively in the sense of a certain amount of entanglement given a partition of the system. This was shown for quantum computing with an architecture based on pure states. The discovery of the DQC1 protocol, based on the use of completely mixed states, changed the story radically.

In the DQC1 protocol the initial state is separable and then one should look for entanglement in the final state. For the bipartition control qubit/mixed qubits there is no entanglement, but for the other bipartitions (*i.e.* when the control qubit is not alone) the entanglement is equal to a non-vanishing constant independent of the number of qubits used [Datta et al., 2005]. As the number of qubits becomes large, the amount of entanglement present in the system becomes proportionally smaller. The authors also studied a generalized DQC1 protocol where the control qubit is no longer pure, but rather a mixed qubit with purity  $\alpha$ :  $\rho_\alpha = (|0\rangle\langle 0| + \alpha Z)/2$  (see circuit c) in Fig. 4.6). For any non-vanishing purity ( $\alpha \neq 0$ ) this generalized DQC1 still provides a speed-up for the calculation of the trace of  $U$ , at the cost of an increasing necessary number of runs. Without being able to show it, their results suggest that separable states do not allow for a speed-up and that entanglement can be present even with a tiny fraction of purity.

The lack of entanglement between the control qubits and the set of mixed qubits was somehow surprising as this is the most natural partition for this system. But as we saw, lack of entanglement does not imply a lack of quantum correlations as soon as the states involved are mixed. Datta et al. [2008] showed that there is indeed a non-vanishing discord for any values of  $\alpha$  present in the final state, and that for a large number of qubits this amount of discord is independent of the number of qubits. This result demonstrated that for certain instances quantum correlations different from entanglement are the genuine quantum resource.

As DQC1-based protocols can also be used in the context of metrology (see Sec. 4.3.3), it is natural to study also in this framework the exact role of correlation in the enhancement. As we have seen, Cable et al. [2016] proposed a generalized DQC1 protocol for metrology, which allows one to surpass the SQL by measuring only one qubit. They characterized the enhancement offered by such protocols and tracked the presence of entanglement and discord. When  $l = n = 0$ , meaning that we just have the control qubit and a set of  $m$   $\varepsilon$ -mixed qubits, the output state is entangled (this can be seen from the fact that the QFI is higher than the SQL — see box 13. It was also found there that for a certain regime the state is separable but has a finite discord, giving

again hints about the use of quantum correlations different than entanglement for metrology.

Modi et al. [2011] studied several schemes based on the power of one qubit. They compared an *i.i.d.* situation (which gives the usual SQL) to three other protocols, two quantum and one classical. In all of them all the qubits have the same finite purity, and only the first one is measured (DQC1-like protocol). The study of the enhancement obtained over the *i.i.d.* strategy coupled to the analysis of the correlations present in the states demonstrates that while classical correlations do not help to increase the sensitivity, quantum correlations do, and this even when purity is so low that entanglement vanishes, showing again the important role of quantum correlations beyond entanglement in quantum information processes.

### 4.5.3 Role of correlations in ancilla assisted schemes

There are two fundamentally different ways to use correlations in metrology: Correlations between probes and correlations between ancillas and probes. Up to now we focused on correlations between probes and in Hamiltonian parameter estimation. We will now review the role played by entanglement in ancilla assisted schemes and see how discord can be used when we miss some knowledge on an interferometric setup.

#### Obvious role of entanglement in ancilla assisted metrology

In the framework of channel estimation we saw how extended channels can lead to a better sensitivity. This improvement is directly related to correlations between the ancillas and the probes, as starting with a separable state is equivalent to having no ancillas. It does not mean neither that entanglement implies necessarily enhancement. For certain channels entanglement with the probe does not bring any advantage and in certain cases one even shows that any amount of entanglement degrades the QFI [Fujiwara, 2004].

An intermediary result in [Fujiwara and Imai, 2008] can be used to characterize when entanglement is useful in an ancilla-assisted protocol. Consider the channel  $\mathcal{E}_\theta$  acting on  $\mathcal{S}(\mathcal{H})$ . We define the function

$$f(\sigma, \{A_j(\theta)\}) = \text{tr} \left[ \sigma \sum_{k=1}^q \dot{A}_k^\dagger(\theta) \dot{A}_k(\theta) \right], \quad (4.36)$$

with  $\{A_j(\theta)\}$  a  $q$ -Kraus decomposition of the channel  $\mathcal{E}_\theta$  and  $\sigma$  an element of  $\mathcal{S}(\mathcal{H})$ . The following theorem [Fujiwara and Imai, 2008] tells us when the channel can benefit from probe-ancilla entanglement:

**Theorem 4.3 — UTILITY OF CHANNEL EXTENSIONS .**

*Consider the channel  $\mathcal{E}_\theta$  acting on  $\mathcal{S}(\mathcal{H})$ . If the function  $\min_{\{A_j(\theta)\}} f(\sigma, \{A_j(\theta)\})$  takes its maximum on the boundary of  $\mathcal{S}(\mathcal{H})$ , meaning if the function is maximized by a pure state, then the channel does not benefit in terms of channel QFI from entanglement between probe and ancilla:*

$$C(\mathcal{E}_\theta \otimes \text{Id}; \theta) = C(\mathcal{E}_\theta; \theta). \quad (4.37)$$

We can actually use this result to show that unitary channels do not profit from ancilla-probe entanglement. For a unitary channel with local generator  $\mathcal{H} = iU^\dagger \dot{U}$  the function  $f$  is equal to  $\text{tr}[\sigma \mathcal{H}^2]$ . This is a convex function with respect to the state and thus takes its maximum on the boundaries, showing that we cannot increase the channel QFI of a unitary with channel extension. We will see in Chapter 9 how we can go over this result using Hamiltonian extensions.

### Black box metrology and discord

In the standard framework of quantum metrology we assume that the only unknown part is the parameter to be estimated. We can go further this idealized point of view and consider the situation where we lack also information about the exact form of the dynamics. Such a situation was studied in [Girolami et al., 2014; Girolami, 2015]. There the authors studied a phase-shift estimation problem with a unitary evolution  $U = \exp(-i\varphi H_\alpha)$  where  $\varphi$  is the parameter to be estimated. They considered the situation where initially only the spectra of the generator  $H_\alpha$  is known, but not its eigenbasis. When the form of the Hamiltonian is exactly known there is no need for ancillas to reach the best sensitivity as we saw it in the last section.

Crucially this changes when we lack knowledge about the Hamiltonian. We thus introduce the possibility of using an ancilla. The situation is thus the following: Two parties (Bob and Alice, Bob having the probe and Alice the ancilla) prepare a bipartite state. Then Bob sends his state through the black-box. When this is done, a third person communicates the full form of the Hamiltonian to Bob and Alice. Eventually they can decide which measurement to perform to estimate  $\varphi$ . It was thus shown that the QFI corresponding to the worst case scenario is a genuine measure of discord. This means that in this case the discord of the state measures the guaranteed sensitivity obtained.

## 4.6 Noise in quantum metrology

We have seen how we can exploit genuine quantum effects to improve the sensitivity in measuring physical quantities. While we have been analysing both ideal and non-ideal evolutions to imprint the parameter, meaning unitary evolution and evolution through non unitary quantum channels, we did not consider yet actual noise in the system. However this question of noise is of crucial interest in practical metrology as it is impossible to avoid completely noise in a laboratory. It is extremely important to characterize the noise, its effect on the applicability of metrological protocols, and, if appropriate, how to deal with it.

### 4.6.1 Effect of the noise

The first study of the impact of noise on a metrological protocol was carried out by Huelga et al. [1997]. There the authors studied the impact of the noise in the context of Ramsey interferometry [Wineland et al., 1992]. Especially they considered GHZ states with  $N$  particles which allow to reach the HL. The noise was modelled by a Markovian master equation (no memory effect in the environment) including a dephasing term. They showed that the presence of such noise, no matter how small, completely destroys the HL scaling and brings back the sensitivity at an SQL level. By numerical calculations they were able to identify some partially entangled states that perform better than the SQL but only improve it by a constant factor, keeping the SQL scaling.

This result clearly demonstrated the importance of noise for any practical proposal of a quantum enhanced measurement. Still, the study of noise for quantum metrology became really popular only ten years later. It is in general difficult to make general statements about the effect of noise. Much progress has been done in this field by using tools not primarily developed for it. For example one can use the upper bound for the QFI and channel QFI derived in Sec. 3.5 to study the effect of the noise. Indeed the upper bound found by Fujiwara and Imai [2008] (or equally the one from Escher et al. [2011]) can be used to determine if a quadratic scaling with the number of invocations of the channel is present. For example if the total channel (say  $\Gamma \circ \mathcal{E}_\theta$ ) is full rank we

know from Theorem 4.2 that the scaling is linear, and any possible original improvement is lost.

Such approach was taken in [Demkowicz-Dobrzański et al., 2012] (see also [Kolodyński, 2014]). There the authors considered four channels: depolarizing, dephasing and spontaneous emission of a qubit and loss in an interferometer. Using the channel QFI of the extended channel (see Theorem 3.7) and a semi-definite optimization, they showed that keeping the HL scaling for large  $N$  requires a level of decoherence scaling as  $1/N$ . Knysh et al. [2011] considered the loss in an interferometer and also found that for a large number of photons we only obtain a constant factor enhancement.

While these results suggest that it is the norm to lose the scaling advantage when noise is present, there are some exceptions. Ji et al. [2008] defined a criterion, called "programmability", and showed that programmable channels lead to an SQL scaling for the QFI. In particular this implies directly that the smallest amount of depolarizing noise destroys any advantage given by entangling probes. But they were also able to find some non-unitary channels that allow to reach the HL. It was also shown that for the phase estimation with qubits it is possible to reach a QFI scaling with  $N^{5/3}$  despite the fact that the qubits are affected by a local, *i.e.* acting-independently on each probe, transverse noise (like the bit-flip and bit-phase flip channel) [Chaves et al., 2013; Brask et al., 2015].

Another example where advantage can be maintained is given by considering non-Markovian noises. It was shown in [Chin et al., 2012; Matsuzaki et al., 2011] that for a local non-Markovian noise it is possible to reach a scaling  $N^{3/2}$ , which outperforms the SQL. Achieving this scaling — known as the Zeno scaling — requires to interrogate the system at a time faster than the typical frequency of the dynamics of the bath. It was then shown that this enhancement holds for a large class of noises, and that it is truly the short time dynamics of the system, the Zeno dynamics, that allows the enhancement [Smirne et al., 2016; Macieszczak, 2015].

The example of the Zeno scaling shows that in the presence of noise the evolution time (also called interrogation time) plays often a crucial role. Kiilerich and Mølmer [2015], using a model of a driven qubit, showed that for noise free estimation, the best strategy is to measure the system as rarely as possible, while in the presence of noise (taken as dephasing) it is better to measure at precisely determined time intervals.

A striking example of the importance of properly taking into account the noise when discussing quantum enhanced measurement protocols can be found in the context of GW detection. We already mentioned that the strategy to reduce the shot noise proposed by Caves [1981], namely injecting a squeezed vacuum in the unused port of the interferometer, has been implemented in GEO600 [LSC, 2011]. Demkowicz-Dobrzański et al. [2013] showed that the enhancement of sensitivity was close to the fundamental bound calculated theoretically, when taking into consideration the losses of the interferometer. They also showed that for high laser powers (comparatively to the losses) Caves's strategy is close to be optimal (meaning that introducing more fancy states would be of almost no help).

## 4.6.2 Fighting the noise

We have seen how noise affects the sensitivity of the quantum enhanced metrological protocols. We now turn our attention to some specific methods to fight this deleterious influence of noise.

### Optimal state preparation

The first way for fighting the effect of noise is to reconsider the initial probes. For example we have seen how, when dealing with Ramsey interferometry, the optimal state in the noiseless case

turns out to be very inefficient in presence of noise. Therefore, if there is no proof that the optimal case (cf. channel QFI) leads to a scaling not better than the one obtained with independent probe, it is important to check if some states that are not optimal in the noiseless case do not bring an advantage in the noisy case. This is the approach we took for the estimation of some specific channel under the presence of loss of particles (see Chapter 8).

This optimization can be done case by case for specific evolutions — for example by studying the different noises that can affect a qubit used for a phase estimation [Chapeau-Blondeau, 2015]. This strategy was also adopted in [Fröwis et al., 2014; Dorner et al., 2009] to the case of interferometry with photons, and the optimal state was optimized taking in consideration the non-used port, what bring us to the discussion of ancillas to fight the noise.

### Using ancillas

Ancillas have proven to be useful in many cases in ideal quantum metrology. Not surprisingly ancillas can also help when noise is present. Even if their introduction does not change the scaling, they can contribute to a non-negligible increase by a constant factor of the QFI. Recently, Huang et al. [2016] showed how introducing ancillas in the framework of phase estimation using qubits affected by Pauli channels leads to a better sensitivity.

In [Demkowicz-Dobrzański and Maccone, 2014] a thorough analysis of the role of entanglement between probes but also with ancillas in noisy metrology was carried out. While for the noiseless case we have the equivalence of parallel and sequential strategy in terms of sensitivity, it turns out that the presence of noise favors the parallel strategy. By analysing some specific kinds of noises, the authors are in position to conjecture that the best scheme is the parallel strategy with ancillas.

### Quantum error correction and dynamical decoupling

Initially developed for correcting the mistakes that unavoidably affect the qubit in a quantum computer, error correction codes can also be applied to quantum metrology [Dür et al., 2014; Kessler et al., 2014; Arrad et al., 2014]. Although we separate this method from the use of ancillas, it is a specific case of the latter. The basic idea of quantum error correction in quantum computing is to use more qubits to code some redundant information, in such a way that if some qubits are affected by errors then we can afterwards correct the error.

Quantum error correction is specially relevant to exploit the time as a resource for enhanced measurement. Indeed, such strategies allows one to extend the sensing time. This has been experimentally realised in [Unden et al., 2016] with spin in NV-diamonds center, a promising magnetometer (see chapter 9).

## Summary Chapter 4

- **Quantum metrological protocol:** *(i)* Preparation of the probes, *(ii)* evolution through the channel, *(iii)* measurement and estimation. Then the steps *(i)*, *(ii)*, *(iii)* are repeated  $m$  times.
- **Standard quantum limit (SQL):** Defined either as a QFI scaling with  $N$  (scaling as  $1/\sqrt{N}$  for the standard deviation) with the number of probes or applications of the channel, or defined as the best sensitivity obtained using separable states.
- **Heisenberg limit (HL):** Defined either as a QFI scaling with  $N N^2$  (scaling as  $1/N$  for the standard deviation) with the number of probes or applications of the channel, or defined as the best sensitivity obtained using entangled states.
- **Query complexity:** Equal to the number of black boxes used in the quantum circuit representation of the metrological protocol. With the query complexity  $Q$  we can define the HL as a scaling  $Q^2$  for the QFI.
- **Entanglement:** For Hamiltonian parameter estimation (with bounded Hamiltonians) entanglement between probes helps to increase the QFI by a factor  $N$ . For non-unitary channels, entanglement between the probe and an ancilla can lead to a larger QFI.
- **Noisy metrology:** In general, even a small amount of noise can result in the loss of the scaling advantage provided by entanglement between probes. This has still to be studied case-by-case, as some protocols are resistant to a certain type of noise.

## Part II

# Coherent averaging



# Chapter 5

## Coherent averaging protocol and metrology<sup>1</sup>

### 5.1 Coherent averaging protocol

The coherent averaging protocol was first introduced by Braun and Martin [2011]. It participates to the effort of searching quantum enhanced measurement protocols that can handle the presence of noise. Up to a certain extent this is equivalent to looking in protocols that do not require to use entangled initial states, as entangled states are very sensitive to decoherence. Interestingly, the coherent averaging protocol is not only resistant to decoherence, but can even use decoherence to reach a Heisenberg Limit (HL) scaling.

In this chapter, after presenting in detail the protocol and its Hamiltonian, we will apply the perturbative estimation theory developed in Sec. 3.7 to assess the performance of coherent averaging for the estimation of several parameters. We will try to keep the results as general as possible in terms of the Hamiltonian.

#### 5.1.1 Protocol

The coherent averaging protocol is based on a star topology (see Fig. 5.1): A central system, the so-called *quantum bus*, is connected *via* pairwise interactions to  $N$  probes. The probes are independent and do not interact with each other.

#### Hamiltonian

We consider the Hilbert space  $\mathcal{H}$ , and its tensorial decomposition  $\mathcal{H} = \mathcal{H}_{p,1} \otimes \cdots \otimes \mathcal{H}_{p,N} \otimes \mathcal{H}_b$ , where  $\mathcal{H}_{p,i}$  is the Hilbert space of the  $i$ th probe while  $\mathcal{H}_b$  stands for the Hilbert space of the quantum bus. The Hamiltonian of the system  $H(x, \omega_p, \omega_b)$  is the sum of two Hamiltonians, the

---

<sup>1</sup>This chapter is based on: "Coherent averaging", Fraïsse, J. M. E. and Braun, D. (2015), *Annalen der Physik*, 527(9-10):701–712. All the figures and parts of the discussion are reproduced from there. ©2015 Annalen der Physik

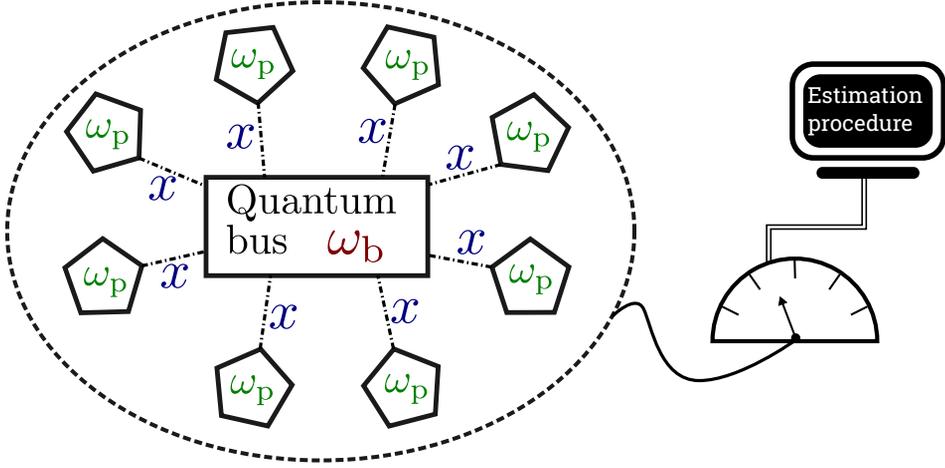


Figure 5.1: Coherent averaging protocol for full-system estimation. The quantum bus interacts pairwise with all the probes. A measurement is performed on the total system to estimate one of the three parameter  $x$ ,  $\omega_p$ ,  $\omega_b$ .

free evolution Hamiltonian  $\delta H_0(\omega_p, \omega_b)$  and the interaction Hamiltonian  $\varepsilon H_{\text{int}}(x)$ . It reads

$$\begin{aligned} H(x, \omega_p, \omega_b) &= \delta H_0(\omega_p, \omega_b) + \varepsilon H_{\text{int}}(x) \\ &= \delta \left( \sum_{i=1}^N F_i^{(i)}(\omega_p) + F_R^{(0)}(\omega_b) \right) + \varepsilon \left( \sum_{i=1}^N \sum_{\nu=1}^k S_{i,\nu}^{(i)}(x) R_\nu^{(0)} \right), \end{aligned} \quad (5.1)$$

where  $A^{(i)} = \mathcal{I}_1 \otimes \cdots \otimes \mathcal{I}_{i-1} \otimes A \otimes \mathcal{I}_{i+1} \otimes \cdots \otimes \mathcal{I}_N \otimes \mathcal{I}_0$  and  $A^{(0)} = \mathcal{I}_1 \otimes \cdots \otimes \mathcal{I}_N \otimes A$ . The Hamiltonian  $F_i(\omega_p)$  governs the free evolution of the probe  $i$ . The free evolution of the quantum bus is given by  $F_R(\omega_b)$ . The interaction part consists of a sum of pairwise interactions between the quantum bus and a probe  $i$ . The most general pairwise interaction is given by a sum of tensor products between operators acting on a probe — the  $S_{i,\nu}(x)$  — and operators acting on the bus —  $R_\nu$ . We introduced the two dimensionless *scale parameters*,  $\varepsilon$  and  $\delta$ , for the purpose of perturbation theory. They allow us to easily identify the different regimes (weak, medium, and strong interaction) and will be specially useful for the ZZXX model.

This Hamiltonian can serve to describe very different models. In particular we did not specify the dimension of the Hilbert space of the probes or of the bus. If this later is taken to be a reservoir to which we have only a partially access, coherent averaging turns out to be a decoherence model. In the opposite direction we can take for the bus a very small system as it is the case in Chapter 6 and 7.

### Initial state

We focus on pure initial states  $|\psi\rangle \in \mathcal{H}$ . Let  $\{|\psi_{j_i}\rangle_i\}$  be a basis of the Hilbert space  $\mathcal{H}_{p,i}$  and  $\{|\xi_j\rangle_0\}$  a basis of the Hilbert space  $\mathcal{H}_b$ . Then any vector of  $\mathcal{H}$  can be written as

$$|\psi_0\rangle = \sum_{\{j_i\}, \beta} c_{\{j_i\}, \beta} \bigotimes_{i=1}^N |\psi_{j_i}\rangle_i \otimes |\xi_\beta\rangle_0, \quad (5.2)$$

where  $\sum_{\{j_i\}} \equiv \sum_{j_1} \sum_{j_2} \cdots \sum_{j_N}$  and  $c_{\{j_i\},\beta} = c_{j_1, \dots, j_N, \beta}$ . The state after evolution is given by

$$|\psi(t, x, \omega_p, \omega_b)\rangle = e^{-itH(x, \omega_p, \omega_b)} |\psi_0\rangle. \quad (5.3)$$

### 5.1.2 Metrology for coherent averaging

#### Parameter to be estimated

In the Hamiltonian (5.1) we are interested in three different parameters:

1. The parameter encoded in the interaction Hamiltonian, noted  $x$ . We call it the *interaction parameter*. It is this parameter that was considered in [Braun and Martin, 2011].
2. The parameter encoded in the free evolution Hamiltonian of the probes, noted  $\omega_p$ . We call it the *free evolution parameter of the probes*.
3. The parameter encoded in the free evolution Hamiltonian of the bus, noted  $\omega_b$ . We call it the *free evolution parameter of the bus*.

We depicted the parameter dependence of the coherent averaging protocol in the Fig. 5.1. Notice that although there are three parameters, we stay on the framework of scalar parameter estimation. This means that when we estimate one parameter the two others are supposed to be perfectly known (at least if we want the QFI to have an operational meaning and not only to serve as an upper bound).

#### Performance and resource count

As we have seen in our presentation of quantum metrology we have to decide what the most relevant resource count is. In this case the number of probes  $N$  looks like a reasonable choice. It turns out that the number of probes is also equal to the query complexity of the system for the estimation of  $x$  or  $\omega_p$ : there are  $N$  invocations of the fundamental free evolution Hamiltonian of a probe and also  $N$  invocations of the fundamental pairwise interaction. For the study of  $\omega_b$  the query complexity is equal to one.

**HL and SQL for  $x$**  In Braun and Martin [2011] only the estimation of  $x$  was considered. When the coherent averaging mimics a decoherence model, the estimation of  $x$  corresponds to the estimation of a parameter characteristic of the decoherence. From there stems the title of the original publication "Heisenberg-limited sensitivity with decoherence-enhanced measurements".

From a metrological point of view the estimation of  $x$  raises some difficulties. The problem is similar to the one raised by  $k$ -body metrology (see Sec. 4.3.1): Is it fair to compare a non-linear protocol to a linear one? Indeed we have seen that in general a parameter characteristic of an interaction has a different nature than a parameter characteristic of a free evolution. This led us to temperate the idea that non-linear metrology outperforms metrology with linear Hamiltonians. We also saw how the controversy can be solved by defining the query complexity as the resource count.

Here the query complexity  $Q$  for  $x$  is equal to the number of probes,  $Q = N$ . From this point of view we define the HL scaling as a QFI proportional to  $N^2$  and the SQL scaling as a QFI proportional to  $N$ . Since the parameter  $x$  characterises a two-body interaction, one could argue that the best protocol to estimate a two-body interaction is the one where all the particles interact with each other, which eventually gives a HL scaling in  $N^4$ . But the comparison is not totally

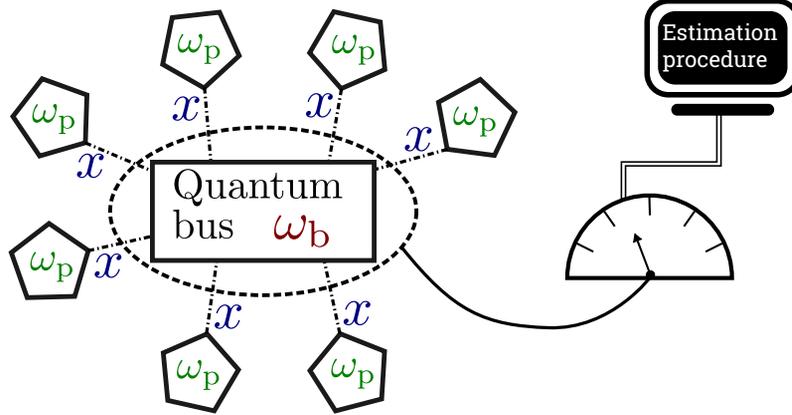


Figure 5.2: Coherent averaging protocol for local estimation. Instead of measuring the all system only the quantum bus is measured.

fair as in the coherent averaging protocol we have in general an interaction between two systems of different nature.

**HL and SQL for  $\omega_p$**  When estimating the free parameter of the probes,  $\omega_p$ , the situation is drastically different. We have  $N$  probes that undergo the unitary evolution in an independent way. The comparison with the linear protocol in quantum metrology is direct, and the number of probes is the relevant resource, again equal to the query complexity. Reaching the HL in the linear case requires to use entangled states. It would then be interesting to study if we can achieve an HL scaling for  $\omega_p$  starting with a separable state.

The comparison is also direct for the estimation of  $\omega_b$ . In the absence of interaction the situation is trivial, we have a single system that undergoes its free evolution.

### Full-system estimation and local estimation

One of the interesting features of coherent averaging is that it allows to reach an HL scaling by performing a measurement on the sole bus. We refer to this situation as *local estimation*<sup>2</sup>. When there can be some ambiguity, we will refer to the standard estimation — where we measure the total system — as *full-system estimation*. Local estimation may play an important role for some models where we have good access to the bus but not to the probes. To study this situation Braun and Martin [2011] used the signal-to-noise ratio (SNR) as a figure of merit. Considering an observable  $A^{(0)}$  that acts non-trivially only on the bus, the SNR  $S(A^{(0)}, |\psi_\theta\rangle; \theta)$  with  $\theta \in \{x, \omega_p, \omega_b\}$  is a good measure of the sensitivity that can be obtained. The problem with this figure of merit is that it strongly depends on the choice of the observable.

We propose a more fundamental figure of merit that we call *local QFI*. It corresponds to the QFI obtained by using only the reduced state of the quantum bus and is denoted as  $I^{(0)}(\rho_\theta; \theta)$  or  $I_\theta^{(0)}$  for short. It is possible to show that the local QFI corresponds to the maximal FI reachable when using the state of the total system but measuring only the bus (see Box 14).

<sup>2</sup>The reader should not confuse this with the use of the word "local" in parameter estimation theory.

## 5.2 Perturbative parameter estimation for coherent averaging

To show how these different figures of merit are connected we consider a bipartite system  $\mathcal{H}_1 \otimes \mathcal{H}_0$  and a state  $\rho_\theta \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_0)$ . The QFI for this state is  $I(\rho_\theta; \theta)$ . For the local estimation with measurement on the second subsystem the local QFI is defined as

$$I^{(0)}(\rho_\theta; \theta) := I(\rho_{\theta,0}; \theta), \quad (5.4)$$

where  $\rho_{\theta,0} = \text{tr}_1[\rho_\theta]$  is the reduced state of the second subsystem. Finally considering an observable  $A^{(0)}$  the SNR is  $S(A^{(0)}, \rho_\theta; \theta)$ .

Due to the monotonicity property of the QFI (see Theorem 3.4 in Sec. 3.3.3) we have

$$I(\rho_\theta; \theta) \geq I^{(0)}(\rho_\theta; \theta),$$

since the partial trace corresponds to a quantum channel. Using the chain of equality  $\text{tr}[\mathcal{I}_1 \otimes A^{(0)} \rho_\theta] = \text{tr}_0[\text{tr}_1[\mathcal{I}_1 \otimes A^{(0)} \rho_\theta]] = \text{tr}_0[A^{(0)} \text{tr}_1[\rho_\theta]] = \text{tr}_0[A^{(0)} \rho_{\theta,0}]$  with  $\rho_{\theta,0} = \text{tr}_1[\rho_\theta]$  we have  $S(A^{(0)}, \rho_\theta; \theta) = S(A, \rho_{\theta,0}; \theta)$ . Since the SNR corresponds to the inverse of the variance of a specific estimator it is bounded by the QFI and we have

$$I^{(0)}(\rho_\theta; \theta) = I(\rho_{\theta,0}; \theta) \geq S(A, \rho_{\theta,0}; \theta). \quad (5.5)$$

Combining all these results we obtain

$$I(\rho_\theta; \theta) \geq I^{(0)}(\rho_\theta; \theta) \geq S(A^{(0)}, \rho_\theta; \theta). \quad (5.6)$$

which with the short notation reads  $I_\theta \geq I_\theta^{(0)} \geq S_\theta^{A^{(0)}}$ .

## 5.2 Perturbative parameter estimation for coherent averaging

### 5.2.1 Interaction pictures, PT I and PT II

Since it is very difficult to find exact expressions for the QFI using the general coherent averaging Hamiltonian, we use perturbation theory and make several assumptions.

One of them is to start with a pure separable state with all the probes in the same state:

$$|\psi_0\rangle = \bigotimes_i^N |\varphi\rangle_i \otimes |\xi\rangle_0. \quad (5.10)$$

Additionally we consider that all the probes evolve under the same Hamiltonian:  $F_i(\omega_p) = F_j(\omega_p)$ ,  $\forall i, j$  and  $S_{i,\nu}(\omega_p) = S_{j,\nu}(\omega_p)$ ,  $\forall i, j, \nu$ . Since all the probes are in the same state and undergo the same evolution, it will prove useful in terms of notation to introduce a generic Hilbert space  $\mathcal{P}$  representing a generic probe. The state in this space is denoted as  $|\varphi\rangle_P$  and the free and interaction Hamiltonians as  $F(\omega_p)$  and  $\{S_\nu(x)\}_\nu$  respectively.

For the sake of concision we write the partial derivative with respect to an arbitrary parameter  $y$  as  $\partial_y$ . We will in general not write all the variables in the Hamiltonians in the interaction picture, but only emphasize the one that we are interested in at the moment. For example when estimating  $x$  we denote the two Hamiltonians in the interaction pictures as  $\varepsilon H_{\text{int},1}(t, x)$  and  $\delta H_{0,1}(t, x)$ . The state  $|\psi(t, x, \omega_p, \omega_b)\rangle$  is denoted simply as  $|\psi(t)\rangle$ .

In Chapter 3 we saw how we can use perturbation theory in Hamiltonian parameter estimation. Especially we showed that for a perturbed Hamiltonian the estimation of a parameter encoded

## Box 14: Local QFI

We introduced the local QFI as the QFI for a reduced state and we claimed that this local QFI is a good figure of merit for the sensitivity when measuring only one subsystem. We can actually show that this figure of merits corresponds to the maximal FI that can be achieved by measuring only the subsystem. To do so we go back to the proof — the physical one — of the QFI. Consider a bipartite system  $\mathcal{H}_1 \otimes \mathcal{H}_0$  and a state  $\rho_\theta$  element of  $\mathcal{H}_1 \otimes \mathcal{H}_0$ . We assume that only the second system can be measured, meaning that all the POVMs are of the form  $\{\mathcal{I}_1 \otimes M_\xi^{(0)}\}$ . The FI is given by

$$J(\text{tr}[\rho_\theta (\mathcal{I}_1 \otimes M_\xi^{(0)})] ; \theta) = \int d\xi \frac{1}{\text{tr}[\rho_\theta (\mathcal{I}_1 \otimes M_\xi^{(0)})]} \text{tr} \left[ \frac{\partial \rho_\theta}{\partial \theta} (\mathcal{I}_1 \otimes M_\xi^{(0)}) \right]^2. \quad (5.7)$$

The important point is to notice that the last term in the integral can be written  $\text{tr}_0 \left[ \text{tr}_1 \left[ \frac{\partial \rho_\theta}{\partial \theta} \right] M_\xi^{(0)} \right]$ . By interchanging the order of the derivative and of the partial trace with respect to the first system we obtain  $\text{tr}_0 \left[ \frac{\partial \text{tr}_1[\rho_\theta]}{\partial \theta} M_\xi^{(0)} \right]$ . Introducing the reduced state  $\rho_{\theta,0} = \text{tr}_1[\rho_\theta]$  of the second system and applying the same procedure to the denominator in the integral we obtain

$$J(\text{tr}[\rho_\theta (\mathcal{I}_1 \otimes M_\xi^{(0)})] ; \theta) = \int d\xi \frac{1}{\text{tr}_0[\rho_{\theta,0} M_\xi^{(0)}]} \text{tr}_0 \left[ \frac{\partial \rho_{\theta,0}}{\partial \theta} M_\xi^{(0)} \right]^2. \quad (5.8)$$

which is nothing else than the FI for the reduced state of the second system  $J(\text{tr}_0[\rho_{\theta,0} M_\xi^{(0)}] ; \theta)$ . Using the quantum Cramér-Rao theorem we obtain

$$J(\text{tr}[\rho_\theta (\mathcal{I}_1 \otimes M_\xi^{(0)})] ; \theta) = J(\text{tr}_0[\rho_{\theta,0} M_\xi^{(0)}] ; \theta) \leq I(\rho_{\theta,0}; \theta) = I^{(0)}(\rho_\theta; \theta), \quad (5.9)$$

which shows that the local QFI is the maximal FI reachable when measuring only the second subsystem.

in the perturbation gives completely different results than the estimation of a parameter encoded in the unperturbed Hamiltonian. For a Hamiltonian  $H(\lambda, \theta) = H_0(\lambda) + V(\theta)$  with  $V$  small, the estimation of  $\theta$  corresponds to PT I and the estimation of  $\lambda$  corresponds to PT II. In PT I there is no zeroth order term in the QFI as in the absence of any perturbation there is nothing to estimate. In PT II the zeroth order term in the QFI is equal to the the QFI in the absence of perturbation. Especially within the range of validity of PT II the scaling is dominated by the zeroth order term.

Going back to coherent averaging, mixing PT I and PT II with the choice of the parameter we identify four formally different cases. First we can choose to take the interaction Hamiltonian or the free evolution Hamiltonian as the perturbation. In the former case we are in the regime of weak interaction and in the latter case in the strong interaction regime. Second, we can choose the parameters to be estimated, either  $x$  or  $\omega_p$  and  $\omega_b$ . Depending on the choice of the perturbation and of the parameter we will have to use PT I or PT II. The four cases are summarized in Table

## 5.2 Perturbative parameter estimation for coherent averaging

| Regime              | Interaction Hamiltonian  | Parameter                | Perturbation                                       |
|---------------------|--|--------------------------|--|
| Weak interactions   | $H_{\text{int,I}} = \varepsilon e^{i t \delta H_0} H_{\text{int}} e^{-i t \delta H_0}$               | $x$                      | PT I : " $\varepsilon \ H_{\text{int}}\  \ll 1$ "  |
|                     |  | $\omega_p$ or $\omega_b$ | PT II: " $\delta \ H_0\  \ll 1$ "                  |
| Strong interactions | $H_{0,\text{I}} = \delta e^{i t \varepsilon H_{\text{int}}} H_0 e^{-i t \varepsilon H_{\text{int}}}$ | $x$                      | PT II : " $\varepsilon \ H_{\text{int}}\  \ll 1$ " |
|                     |  | $\omega_p$ or $\omega_b$ | PT I: " $\delta \ H_0\  \ll 1$ "                   |

Table 5.1: The four cases for perturbation theory with the coherent averaging protocol

5.1.

In the case of PT II we give an expression for the zeroth order term, which is the unperturbed QFI. The higher order terms (see Appendix C.2) only bring some small corrections that do not modify the essence of the result. This questions the utility of PT II for metrology, as we are mainly concerned about the global behaviour of the QFI and not by small deviations. Still perturbation theory can be useful for trying to predict at which point perturbation theory breaks and the results are not valid anymore. Indeed outside the regime of validity of PT II the unperturbed QFI may not reproduce faithfully the behaviour of the QFI.

### 5.2.2 Perturbative parameter estimation for the interaction parameter

$x$

#### Weak interactions — PT I

We seek to estimate the parameter  $x$  encoded in the interaction Hamiltonian, while treating this Hamiltonian as a perturbation. This situation, which corresponds to PT I, is the one that was originally studied by Braun and Martin [2011].

Using Eq. (3.147) we obtain for the QFI up to second order in perturbation theory

$$I_x^{H_{\text{int,I}}} = 4 \int_0^t dt_1 \int_0^{t_1} dt_2 K_{|\psi_0\rangle}(\partial_x H_{\text{int,I}}(x, t_1), \partial_x H_{\text{int,I}}(x, t_2)) + \mathcal{O}(\varepsilon^3). \quad (5.11)$$

The correlation function can be calculated and reads [Braun and Martin, 2011]:

$$\begin{aligned} K_{|\psi_0\rangle}(\partial_x H_{\text{int,I}}(x, t_1), \partial_x H_{\text{int,I}}(x, t_2)) = \\ \varepsilon^2 \sum_{\mu, \nu}^k \left\{ NK_{|\varphi\rangle_P}(\partial_x S_\nu(x, t_1) \partial_x S_\mu(x, t_2))_0 \langle \xi | R_\nu(t_1) R_\mu(t_2) | \xi \rangle_0 \right. \\ \left. + N^2 K_{|\xi\rangle_0}(R_\nu(t_1) R_\mu(t_2))_P \langle \varphi | \partial_x S_\nu(x, t_1) | \varphi \rangle_P \langle \varphi | \partial_x S_\mu(x, t_2) | \varphi \rangle_P \right\} \quad (5.12) \end{aligned}$$

where the Hamiltonians in the interaction picture are defined as:

- $S_\nu(x, t) = e^{i t F(\omega_p)} S_\nu(x) e^{-i t F(\omega_p)}$  (probe part of the interaction — we used the generic probe space)
- $R_\nu(t) = e^{i t F_R(\omega_b)} R_\nu e^{-i t F_R(\omega_b)}$  (bus part of the interaction)

Importantly we see that there is a term of order  $N^2$  in the correlation function. Provided that its prefactor does not vanishes it shows that we can reach an HL scaling when measuring  $x$  in the regime of weak interactions. Remarkably it is the case although we started from a separable state.

### Strong interactions — PT II

We now look at the use of PT II with the estimation of  $x$ , meaning that it is the free Hamiltonian that plays the role of the perturbation and that the interaction dominates the dynamics.

The leading term in the QFI is equal to

$$I_x^{H_{0,I}} = t^2 (\langle \psi_0 | (\partial_x H_{\text{int}}(x))^2 | \psi_0 \rangle - \langle \psi_0 | \partial_x H_{\text{int}}(x) | \psi_0 \rangle^2) + \mathcal{O}(\delta). \quad (5.13)$$

By developing this result we obtain a similar form as we obtained for the weak interaction, but without the integrals over time:

$$I_x^{H_{0,I}} = \varepsilon^2 t^2 \sum_{\mu, \nu}^k \left\{ N K_{|\varphi\rangle_P} (\partial_x S_\nu(x) \partial_x S_\mu(x))_0 \langle \xi | R_\nu R_\mu | \xi \rangle_0 \right. \\ \left. + N^2 K_{|\xi\rangle_B} (R_\nu R_\mu)_P \langle \varphi | \partial_x S_\nu(x) | \varphi \rangle_P \langle \varphi | \partial_x S_\mu(x) | \varphi \rangle_P \right\} + \mathcal{O}(\delta). \quad (5.14)$$

Again we have an HL scaling term in the QFI.

### 5.2.3 Perturbative parameter estimation for free evolution parameters

$\omega_P$

#### Weak interactions — PT II

In the case of weak interaction the estimation of  $\omega_p$  is treated using PT II. When we look at the dominant term, which corresponds to the case where we completely neglect the interaction, we obtain:

$$I_{\omega_p}^{H_{\text{int},I}} = 4Nt^2 ({}_P \langle \varphi | (\partial_{\omega_p} F(\omega_p))^2 | \varphi \rangle_P - ({}_P \langle \varphi | \partial_{\omega_p} F(\omega_p) | \varphi \rangle_P)^2) + \mathcal{O}(\varepsilon). \quad (5.15)$$

This result was expected since in this case we are estimating  $N$  independent probes: Due to the additivity of the QFI we get the usual SQL prefactor  $N$  in front of the variance of the derivative of the state.

#### Strong interactions — PT I

PT I for the estimation of  $\omega_p$  correspond to the regime of strong interactions. To calculate the QFI we need  $H_{0,I}$ , the free Hamiltonian in the interaction picture:

$$H_{0,I} = e^{i t H_{\text{int}}} H_0 e^{-i t H_{\text{int}}} \quad (5.16)$$

$$= \delta \left( \sum_{i=1}^N e^{i t H_{\text{int}}} F_i^{(i)}(\omega_p) e^{-i t H_{\text{int}}} + e^{i t H_{\text{int}}} F_R^{(0)}(\omega_b) e^{-i t H_{\text{int}}} \right). \quad (5.17)$$

Using again Eq. (3.147) the QFI is

$$I_{\omega_p}^{H_{0,I}} = 4 \int_0^t dt_1 \int_0^t dt_2 K_{|\psi_0\rangle} \left( \partial_{\omega_p} H_{0,I}(\omega_p, t_1), \partial_{\omega_p} H_{0,I}(\omega_p, t_2) \right) + \mathcal{O}(\delta^3). \quad (5.18)$$

The derivative with respect to  $\omega_p$  of the free Hamiltonian in the interaction picture reads

$$\partial_{\omega_p} H_{0,I} = \delta \sum_{i=1}^N e^{i t H_{\text{int}}} \partial_{\omega_p} F_i^{(0)}(\omega_p) e^{-i t H_{\text{int}}}. \quad (5.19)$$

## 5.2 Perturbative parameter estimation for coherent averaging

To compute the correlation function we should assume that  $[R_\nu, R_\mu] = 0, \forall \nu, \mu$ . Then we obtain

$$K_{|\psi_0\rangle} \left( \partial_{\omega_p} H_{0,I}(\omega_p, t_1), \partial_{\omega_p} H_{0,I}(\omega_p, t_2) \right) = \delta^2 \left\{ N_0 \langle \xi | K_{|\varphi\rangle_P} \left( \partial_{\omega_p} F_I(\omega_p, t_1), \partial_{\omega_p} F_I(\omega_p, t_2) \right) | \xi \rangle_0 + N^2 K_{|\xi\rangle_0} \left( {}_P \langle \varphi | \partial_{\omega_p} F_I(\omega_p, t_1) | \varphi \rangle_P, {}_P \langle \varphi | \partial_{\omega_p} F_I(\omega_p, t_2) | \varphi \rangle_P \right) \right\}, \quad (5.20)$$

with

$$\partial_{\omega_p} F_I(\omega_p, t) = e^{i\epsilon t \sum_\nu S_\nu \otimes R_\nu} \partial_{\omega_p} F(\omega_p) e^{-i\epsilon t \sum_\nu S_\nu \otimes R_\nu}. \quad (5.21)$$

Notice that  $\partial_{\omega_p} F_I(\omega_p, t)$  is an operator acting on the space of the probe *and* of the quantum bus. Therefore, elements like  ${}_P \langle \varphi | F_I(\omega_p, t) | \varphi \rangle_P$  are still operators acting on the quantum bus, and in a similar way,  ${}_0 \langle \xi | K_{|\varphi\rangle_P} (F_I(\omega_p, t_1), F_I(\omega_p, t_2)) | \xi \rangle_0$  is still an operator on the probe space.

This is an important result as it shows that we can obtain an HL scaling for the parameter of the free evolution of the probe while starting with a separable state.

### 5.2.4 Perturbative parameter estimation for free evolution parameters

$\omega_B$

#### Weak interaction — PT II

If we completely neglect the interaction Hamiltonian we get for the QFI

$$I_{\omega_b}^{H_{\text{int},I}} = 4t^2 ({}_0 \langle \xi | (\partial_{\omega_b} F_R(\omega_b))^2 | \xi \rangle_0 - {}_0 \langle \xi | \partial_{\omega_b} F_R(\omega_b) | \xi \rangle_0^2) + \mathcal{O}(\epsilon). \quad (5.22)$$

#### Strong interaction — PT I

We have to consider the free Hamiltonian in the interaction picture  $H_{0,I}$ . However this time the derivative will be with respect to  $\omega_b$ , and thus the useful Hamiltonian is

$$\partial_{\omega_b} F_{R,I}^{(0)}(\omega_b, t) = e^{i\epsilon t \sum_{i,\nu} S_{i,\nu}^{(i)} R_\nu^{(0)}} \partial_{\omega_b} F_R^{(0)}(\omega_b) e^{-i\epsilon t \sum_{i,\nu} S_{i,\nu}^{(i)} R_\nu^{(0)}} + \mathcal{O}(\delta). \quad (5.23)$$

It is difficult to go further in the simplification since the exponentials make the total quantity an operator acting non-trivially in the full Hilbert space.

### 5.2.5 Perturbative parameter local estimation in PT I

Up to here we focused only on the full-system estimation. It is also possible to obtain perturbative results for the SNR. In [Braun and Martin, 2011] a perturbative formula for the SNR in PT I was derived using two assumptions: First that the initial state is an eigenstate of the observable:  $A|\xi\rangle = a_\xi|\xi\rangle$ , and second that the observable commutes with the free Hamiltonian of the bus:  $[A, F_R] = 0$ . Under these two conditions it was shown that the SNR for the estimation of  $x$  can be written as

$$S_{x, H_{\text{int},I}}^{A(0)} = \frac{\left| \int_0^t dt_1 \int_0^t dt_2 \partial_x \sum_{\nu,\mu} \chi_{\nu\mu}(N, x, t_1, t_2) {}_0 \langle \xi | R_\nu(t_1) [A, R_\mu(t_2)] | \xi \rangle_0 \right|^2}{\int_0^t dt_1 \int_0^t dt_2 \sum_{\nu,\mu} \chi_{\nu\mu}(N, x, t_1, t_2) {}_0 \langle \xi | [R_\nu(t_1), A] [A, R_\mu(t_2)] | \xi \rangle_0} + \mathcal{O}(\epsilon), \quad (5.24)$$

## Coherent averaging scheme and metrology

where

$$\chi_{\nu\mu}(N, x, t_1, t_2) = NK_{|\varphi\rangle_P}(S_\nu(x, t_1), S_\mu(x, t_2)) + N^2_P \langle \varphi | S_\nu(x, t_1) | \varphi \rangle_P \langle \varphi | S_\mu(x, t_2) | \varphi \rangle_P .$$

We provide in the appendix (see Sec. C.3) the SNR for  $x$  when the two assumptions are relaxed.

## Summary Chapter 5

- **Coherent averaging protocol:** A central system, the quantum bus, is connected to  $N$  probes with pairwise interactions. The probes do not interact with each other. We consider only separable initial states.
- **Parameters:** Three parameters: One encoded in the pairwise interactions,  $x$ . One encoded in the free evolution of the probes,  $\omega_p$ . One encoded in the free evolution of the quantum bus,  $\omega_b$ .
- **Local estimation:** Corresponds to the situation where we only have access to the quantum bus for the measurement. Define the local QFI  $I_\theta^{(0)}$  as the QFI for the reduced state of the bus. We also consider the SNR for a local (acting only on the quantum bus) observable  $A^{(0)}$ . It holds that  $I_\theta \geq I_\theta^{(0)} \geq S_\theta^{A^{(0)}}$ .
- **Perturbation theory I:** Estimate the parameter encoded in the perturbation. For weak interactions we estimate  $x$  and for strong interaction we estimate  $\omega_p$ .
- **Perturbation theory II:** Estimate the parameter encoded in the unperturbed Hamiltonian. For weak interactions we estimate  $\omega_p$  and for strong interactions we estimate  $x$ . The zeroth order term dominates and the perturbation cannot change the global behaviour of the QFI.
- **Estimation of  $x$ :** Perturbation theory predicts an HL scaling for weak and strong interactions.
- **Estimation of  $\omega_p$ :** Perturbation theory predicts an HL scaling for strong interactions.



# Chapter 6

## Qubit model: ZZZZ Hamiltonian<sup>1</sup>

Up to now, our analysis of the coherent averaging protocol was fully based on perturbation theory. In this section we introduce a specific model of coherent averaging that can be solved exactly, in which both the probes and the quantum bus are qubits. It can be considered as a toy model which helps to test the validity of perturbation theory and to explore cases that are not accessible via perturbation theory.

### 6.1 Hamiltonian and state of the system

We consider the following Hamiltonian:

$$H(x, \omega_p, \omega_b) = \delta \left( \frac{\omega_p}{2} \sum_{i=1}^N Z^{(i)} + \frac{\omega_b}{2} Z^{(0)} \right) + \varepsilon \left( \frac{x}{2} \sum_{i=1}^N Z^{(i)} Z^{(0)} \right), \quad (6.1)$$

where  $Z$  is the Pauli matrix defined in Eq. (2.44). The interaction Hamiltonian commutes with the free Hamiltonian and we can calculate the state after the evolution as well as the QFI exactly.

As we did in the previous chapter we consider a pure state, symmetric under the exchange of probes:

$$|\psi_0\rangle = \bigotimes_{i=1}^N (\cos(\alpha)|0\rangle_i + \sin(\alpha) e^{i\phi} |1\rangle_i) \otimes (\cos(\beta)|0\rangle_0 + \sin(\beta) e^{i\gamma} |1\rangle_0), \quad (6.2)$$

where  $|0\rangle, |1\rangle$  are “computational basis states”, *i.e.*  $Z|0\rangle = |0\rangle$  and  $Z|1\rangle = -|1\rangle$ . The symmetric state of  $N$  qubits can be represented as a spin  $N/2$  (see Sec. 2.6.2 and especially Eq. (2.51)). Using the spin notation the Hamiltonian is written as

$$H(x, \omega_p, \omega_b) = \delta \left( \frac{\omega_p}{2} J_z \otimes \mathcal{I} + \frac{\omega_b}{2} Z^{(0)} \right) + \varepsilon \frac{x}{2} (J_z \otimes \mathcal{I}) Z^{(0)}, \quad (6.3)$$

where  $J_z$  is the spin operator in the  $z$ -direction whose effect on a basis state is given by  $J_z|j, m\rangle = m|j, m\rangle$ .

---

<sup>1</sup>This chapter is based on: "Coherent averaging", Fraïsse, J. M. E. and Braun, D. (2015), *Annalen der Physik*, 527(9-10):701–712. All the figures and parts of the discussion are reproduced from there. ©2015 Annalen der Physik

## 6.2 Exact solution for full-system estimation

We first look at the the full-system estimation. After the evolution the state of the system is given by

$$\begin{aligned}
 |\psi(t)\rangle = & \bigotimes_{i=1}^N \left( \cos(\alpha) e^{-i\delta\omega_p t/2} e^{-i\varepsilon x t/2} |0\rangle_i + \sin(\alpha) e^{i\phi} e^{i\delta\omega_p t/2} e^{i\varepsilon x t/2} |1\rangle_i \right) \\
 & \otimes \cos(\beta) e^{-i\delta\omega_b t/2} |0\rangle_0 \\
 & + \bigotimes_{i=1}^N \left( \cos(\alpha) e^{-i\delta\omega_p t/2} e^{i\varepsilon x t/2} |0\rangle_i + \sin(\alpha) e^{i\phi} e^{i\delta\omega_p t/2} e^{-i\varepsilon x t/2} |1\rangle_i \right) \\
 & \otimes \sin(\beta) e^{i\gamma} e^{i\delta\omega_b t/2} |1\rangle_0 . \quad (6.4)
 \end{aligned}$$

Using this state we can calculate the QFI for any parameter  $\theta \in \{x, \omega_p, \omega_b\}$ ,

$$I_\theta \equiv I(|\psi(t)\rangle; \theta) = 4(\langle \partial_\theta \psi(t) | \partial_\theta \psi(t) \rangle - |\langle \psi(t) | \partial_\theta \psi(t) \rangle|^2) . \quad (6.5)$$

### 6.2.1 Full-system estimation of $x$

The QFI for  $x$  is equal to

$$I_x = N^2 t^2 \varepsilon^2 \cos^2(2\alpha) \sin^2(2\beta) + N t^2 \varepsilon^2 \sin^2(2\alpha) . \quad (6.6)$$

This QFI is a sum of an SQL-like term and an HL-like term, whose respective weights are determined by the polar angle of the probes,  $2\alpha$ , and the polar angle of the bus,  $2\beta$ . The fact that only the polar angle plays a role is explained by the form of the Hamiltonian: it is just composed by  $Z$  Pauli matrices. All that matters is the  $z$ -component of the Bloch vector, which for a pure state depends only on the polar angle.

In terms of sensitivity  $\alpha$  and  $\beta$  have an opposite role. The highest QFI is reached for  $\beta = \pi/4$  and  $\alpha = 0$  (notice that this is true modulo  $\pi$ ). In this case the QFI reaches the HL ( $\varepsilon^2 N^2 t^2$ ) and the state after evolution is equal to

$$|\psi(t)\rangle = \bigotimes_{i=1}^N |0\rangle_i \otimes \left( e^{-iNt\varepsilon x/2} e^{-i\delta\omega_b t/2} |0\rangle_0 + e^{iNt\varepsilon x/2} e^{i\delta\omega_b t/2} |1\rangle_0 \right) / \sqrt{2} . \quad (6.7)$$

This expression shows that it is the phase accumulation on the quantum bus that allows to reach the HL while starting with a separable state. We understand why  $\beta = \pi/4$  and  $\alpha = 0$  corresponds to the best state: It makes the quantum bus in a balanced superposition of the eigenstates of  $Z$ , which maximizes the contrast needed to read the phase, and it makes the probes in an eigenstate of  $Z$  which allows to completely transfer the phase to the quantum bus. The state (6.7) is a separable state. Having no entanglement in the initial state nor in the state after the evolution, we are sure that the enhancement is not due to entanglement.

If  $\alpha = \pi/4$  no phase is transferred to the quantum bus and we only obtain the SQL. In a similar way, if  $\beta = 0$  there is no "contrast"; all the phase transferred is a global irrelevant phase and we only get an SQL term. In the worst case we have  $\alpha = \pi/4$  and  $\beta = 0$ : The entire phase is transferred but it cannot be read out and the QFI is equal to zero.

We depicted in Fig. 6.1 the QFI for three different states: One that reaches the HL, one that reaches the SQL, and one leading to a QFI with bot an SQL term and an HL term.

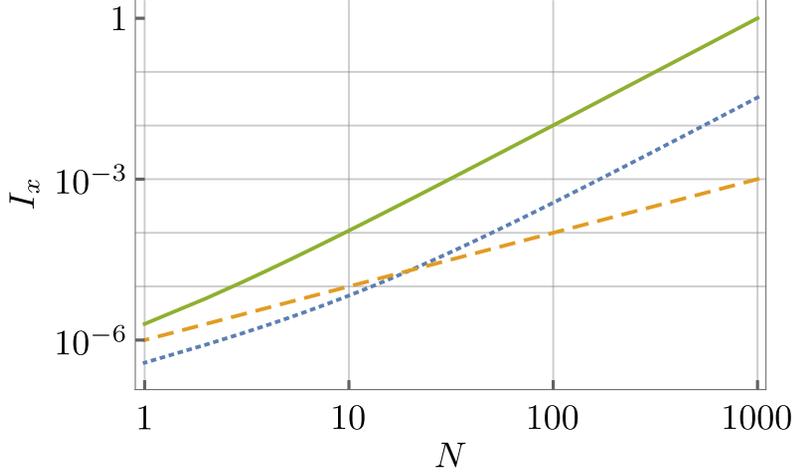


Figure 6.1: Full-system QFI for  $x$  with the ZZZZ model (log-log scale). The parameters are  $t = \omega_p = \omega_b = \delta = 1$ ,  $\gamma = 0$ , and  $\varepsilon^3$ . The green plain line corresponds to the optimal state ( $\beta = \pi/4$ ,  $\alpha = \pi/4$ ), the orange dashed line to  $\alpha = \pi/4$ , and the blue dotted line to  $\beta = 2\pi/5$ ,  $\alpha = \pi/5$ .

### 6.2.2 Full-system estimation of $\omega_p$ or $\omega_b$

The QFI for the parameter of the free evolution of the probes is equal to

$$I_{\omega_p} = N\delta^2 t^2 \sin^2(2\alpha). \quad (6.8)$$

In contrast to the estimation of  $x$  there is only an SQL term. We cannot reach the HL for  $\omega_p$  with this simple model. The QFI is the same as the one that we would obtain by using  $N$  independent qubits. It is only a function of  $\alpha$ : The closer  $\alpha$  is from  $\pi/4$  the higher the QFI and the closer it is from 0 the lower the QFI.

The situation is similar for the estimation of the parameter of the free evolution of the bus. The QFI is equal to the one we would obtain without the probes:

$$I_{\omega_b} = \delta^2 t^2 \sin^2(2\beta). \quad (6.9)$$

## 6.3 Exact solution for local estimation

In this section we look at the local estimation. Since we can reach the HL for the estimation of  $x$  we want to know if we can achieve this by measuring only the bus or if a measurement of the probes is required.

### 6.3.1 Local QFI

In order to calculate the local QFI we should first calculate the reduced density matrix  $\rho_{\text{bus}}(t) = \text{tr}_{\text{probes}}[|\psi(t)\rangle\langle\psi(t)|]$  by tracing out all the probes :

$$\rho_{\text{bus}}(t) = \begin{pmatrix} \cos^2(\beta) & \sin(2\beta) e^{-i(\gamma+\delta\omega_b t)}/2 \\ \sin(2\beta) e^{i(\gamma+\delta\omega_b t)}/2 & (\cos^2(\alpha) e^{-i\varepsilon x t} + \sin^2(\alpha) e^{i\varepsilon x t})^N \\ (\cos^2(\alpha) e^{i\varepsilon x t} + \sin^2(\alpha) e^{-i\varepsilon x t})^N & \sin^2(\beta) \end{pmatrix}. \quad (6.10)$$

The eigenvalues are given by:

$$\lambda_{\pm} = \frac{1}{2} \left( 1 \pm \sqrt{\cos^2(2\beta) + \sin^2(2\beta) |\cos^2(\alpha) e^{i\varepsilon x t} + \sin^2(\alpha) e^{-i\varepsilon x t}|^{2N}} \right), \quad (6.11)$$

and the corresponding eigenvectors by:

$$|v_{\pm}\rangle = \frac{w}{\mathcal{N}_{\pm}} |0\rangle + \frac{y_{\pm}}{\mathcal{N}_{\pm}} |1\rangle, \quad (6.12)$$

with

$$w = \sin(2\beta) e^{-i(\gamma+\delta\omega_b t)} (\cos^2(\alpha) e^{-i\varepsilon x t} + \sin^2(\alpha) e^{i\varepsilon x t})^N, \quad (6.13)$$

$$y_{\pm} = -\cos(2\beta) \pm \sqrt{\cos^2(2\beta) + \sin^2(2\beta) |\cos^2(\alpha) e^{i\varepsilon x t} + \sin^2(\alpha) e^{-i\varepsilon x t}|^{2N}}, \quad (6.14)$$

$$\mathcal{N}_{\pm} = \sqrt{2} \left\{ (\cos^2(2\beta) + \sin^2(2\beta) |\cos^2(\alpha) e^{i\varepsilon x t} + \sin^2(\alpha) e^{-i\varepsilon x t}|^{2N}) \mp \cos(2\beta) \sqrt{\cos^2(2\beta) + \sin^2(2\beta) |\cos^2(\alpha) e^{i\varepsilon x t} + \sin^2(\alpha) e^{-i\varepsilon x t}|^{2N}} \right\}^{1/2}. \quad (6.15)$$

**Estimation of  $x$**  From the Eq. (6.11-6.15) we see that the local QFI for  $x$  has *a priori* both a quantum and a classical contribution (respectively coming from the eigenvectors and from the eigenvalues). Due to the length of the equations it is difficult to write the QFI in a concise way. We thus opt for making a study for different values of the initial state.

- $\beta = 0$ : This corresponds to the case where the bus starts in the state  $|0\rangle$ . After the evolution, the reduced density matrix is equal to  $\rho_{\text{bus}} = |0\rangle\langle 0|$ , so nothing has changed and there is nothing to estimate, the QFI is null.

$$I_x^{\text{bus}}(\beta = 0) = 0. \quad (6.16)$$

- $\beta = \pi/4$ : This corresponds to the case where the bus starts in the state  $(|0\rangle + |1\rangle)/\sqrt{2}$ . Consider the two cases:

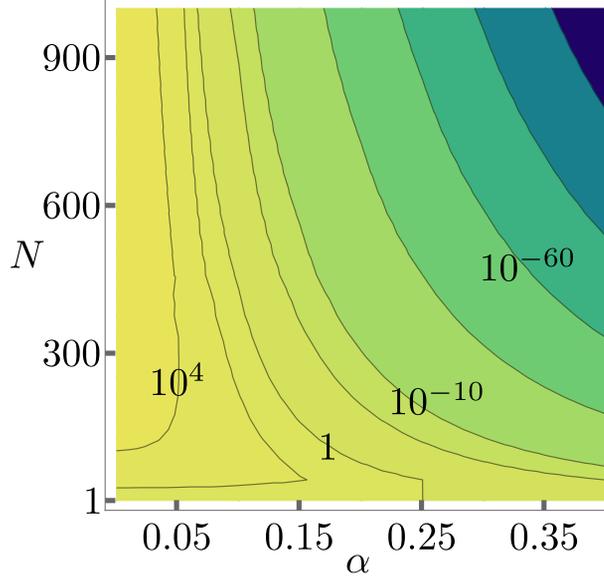


Figure 6.2: Local QFI for the ZZZZ model for  $x$  as a function of  $\alpha$  and  $N$ , with  $\varepsilon = \delta = 1$ ,  $t = 1$ ,  $x = 1$ ,  $\omega_0 = 1$ ,  $\omega_1 = 1$ ,  $\beta = \pi/4$ ,  $\phi = 0$ ,  $\varphi = 0$ . The contours are at  $I_x^{(0)} = \{10^4, 10^2, 1, 10^{-5}, 10^{-10}, 10^{-30}, 10^{-60}, 10^{-100}, 10^{-150}\}$ .

- $\alpha = 0$ , the probes start in the state  $|0\rangle$ . In this case the QFI is purely quantum and it is equal to

$$I_x^{\text{bus}}(\beta = \pi/4, \alpha = 0) = N^2 \varepsilon^2 t^2, \quad (6.17)$$

which is the HL. The local QFI is equal to the full-system QFI: Indeed we have seen that for  $\beta = \pi/4$  and  $\alpha = 0$  the state after evolution is separable and all the dependence on  $x$  is on the bus. Tracing out the probes has no influence on the estimation.

- $\alpha = \pi/4$ , the probes start in the balanced superposition  $(|0\rangle + |1\rangle)/\sqrt{2}$ . In this case the QFI is purely classical and is equal to

$$I_x^{\text{bus}}(\beta = \pi/4, \alpha = \pi/4) = \frac{\varepsilon^2 t^2 N^2 \tan^2(\varepsilon x t) \cos^{2N}(\varepsilon x t)}{1 - \cos^{2N}(\varepsilon x t)}. \quad (6.18)$$

The QFI has a non-polynomial dependency on the number of probes  $N$ . We see that we have the characteristic  $N^2$  factor of the HL scaling, but there is also a power to the  $N$  on the cosine. In order to get the scaling we can make a development at large  $N$ . For this we rewrite the QFI as

$$I_x^{\text{bus}}(\beta = \pi/4, \alpha = \pi/4) = \frac{\varepsilon^2 t^2 N^2 \tan^2(\varepsilon x t)}{e^{N|\log(\cos^2(\varepsilon x t))|} - 1}. \quad (6.19)$$

When  $N$  becomes large, the exponential in the denominator will take over the  $N^2$  power in the numerator and the QFI goes to zero.

In the Fig. 6.2 we show the local QFI when starting with  $\beta = \pi/4$  — *a priori* the most favourable state of the bus — as a function of  $\alpha$  and  $N$ . It appears that apart for the case where

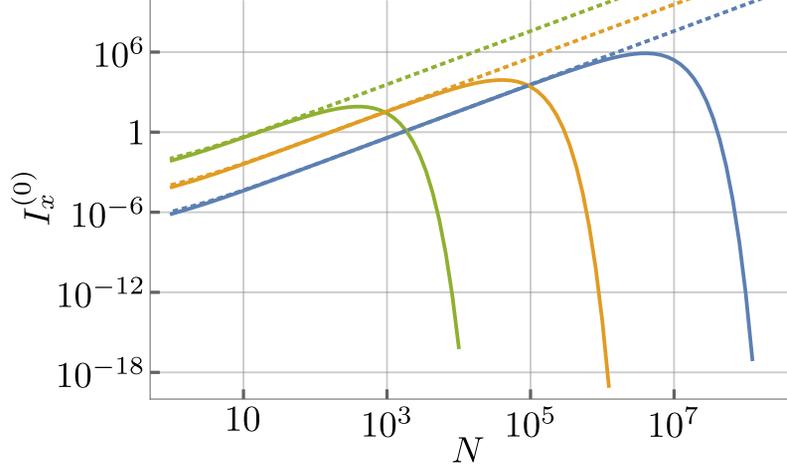


Figure 6.3: Local QFI for the ZZZZ model for  $x$  as a function of  $N$ . The parameters are  $t = \omega_p = \omega_b = x = \delta = 1$ ,  $\gamma = 0$ ,  $\beta = \pi/6$ , and  $\alpha = \pi/8$ . The plain lines represent the local QFI: The green line corresponds to  $\varepsilon = 10^{-1}$ , the orange line to  $\varepsilon = 10^{-2}$ , and the blue line to  $\varepsilon = 10^{-3}$ . The dotted lines correspond to the respective full-system QFI.

$\alpha$  is very close to zero, the local QFI goes to zero for large  $N$ . The closer  $\alpha$  is from  $\pi/4$ , the faster it converges to zero is. Summarized, the plot indicates that it is only for a very specific state that we can reach the HL by measuring only the bus, and that any small changes in the initial states destroy the advantage for large  $N$ .

In the Fig. 6.3 we plot both the local and the full-system QFI as a function of  $N$  for a typical state (avoiding the special cases like  $\{\alpha, \beta\} \in \{\pi/4, 0\}$ ) and for different values of the scale parameter  $\varepsilon$ . The full-system QFI corresponds to the dotted lines. We see that the higher  $\varepsilon$  is, the higher is the QFI. For the local estimation the situation is similar only for low values of  $N$ . Indeed, after a threshold whose value depends on  $\varepsilon$ , the local QFI goes to zero. The threshold corresponds roughly to  $N = 1/\varepsilon^2$ .

**Estimation of  $\omega_p$  and  $\omega_b$**  Although we can also reach the SQL for  $\omega_p$  it is still interesting to check if we can achieve it by measuring only the bus. Unfortunately, a glance to the reduced state of the bus (6.10) is enough to assess that it is impossible to estimate the value of  $\omega_p$  by measuring only the quantum bus, since the parameter  $\omega_p$  does not appear in it:

$$I_{\omega_p}^{\text{bus}}(\beta, \alpha) = 0. \quad (6.20)$$

For the estimation of  $\omega_b$  it is hard to find any advantage offered by the coherent averaging protocol: the QFI is not enhanced by the presence of the probes and it is not easier to measure  $N$  qubits than only one. For the sake of completeness we still provide the local QFI for  $\omega_b$  which reads

$$I_{\omega_p}^{\text{bus}}(\beta, \alpha) = \delta^2 t^2 \sin^2(2\beta) |\cos^2(\alpha) e^{i\varepsilon x t} + \sin^2(\alpha) e^{-i\varepsilon x t}|^{2N}. \quad (6.21)$$

### 6.3.2 Local observable estimation

We now investigate the efficiency of the scheme when one uses a specific observable to estimate the parameter. More precisely we will look at the SNR when measuring  $X^{(0)}$ . Since the dynamics of

the system consists of a rotation of the qubits around the  $z$ -axis, a measurement on the  $xy$ -plane should be efficient to capture the phase information.

With the help of Eq. (6.4) we can compute the expectation value of  $X^{(0)}$

$$\langle \psi(t) | X^{(0)} | \psi(t) \rangle = \sin(2\beta) \left\{ e^{i(\gamma + \delta\omega_b t)} (\cos^2(\alpha) e^{-i\varepsilon x t} + \sin^2(\alpha) e^{i\varepsilon x t})^N + e^{-i(\gamma + \delta\omega_b t)} (\cos^2(\alpha) e^{i\varepsilon x t} + \sin^2(\alpha) e^{-i\varepsilon x t})^N \right\}. \quad (6.22)$$

Using that  $X^{(0)2} = \mathcal{I}$  we can write the SNR as

$$S_\theta^{X^{(0)}} = \frac{|\partial_\theta \langle \psi(t) | X^{(0)} | \psi(t) \rangle|^2}{1 - \langle \psi(t) | X^{(0)} | \psi(t) \rangle^2}, \quad (6.23)$$

where  $\theta \in \{x, \omega_p, \omega_b\}$ .

**Estimation of  $x$**  Again, the expression for a general state is cumbersome and we just look at some specific initial states:

- For  $\beta = 0$  we know from Eq. (6.16) that we cannot estimate  $x$ , which eventually shows that the SNR also vanishes,

$$S_x^{X^{(0)}}(\beta = 0) = 0. \quad (6.24)$$

- For  $\beta = \pi/4$  we will again look at the two extreme cases for  $\alpha$ :

- When  $\alpha = 0$  we are in the optimal situation in terms of the QFI and the local QFI. The SNR reads

$$S_x^{X^{(0)}}(\beta = \pi/4, \alpha = 0) = N^2 t^2 \varepsilon^2, \quad (6.25)$$

which is also the HL. This shows that we can reach the HL by simply measuring  $X$  on the bus.

- When  $\alpha = \pi/4$  the SNR is equal to

$$S_x^{X^{(0)}}(\beta = \pi/4, \alpha = \pi/4) = \frac{N^2 \varepsilon^2 t^2 \tan^2(\varepsilon x t)}{\cos^{-2}(\gamma + \delta\omega_b t) \cos^{-2N}(\varepsilon x t) - 1}, \quad (6.26)$$

which is essentially equal to the local QFI with an extra term which is a function of  $\gamma$  and  $\omega_b$ . Especially the behaviour with  $N$  is similar: for large  $N$  the SNR goes to zero.

## 6.4 Perturbation theory as an example

Although we were able to solve exactly the problem with this very simple Hamiltonian, it is instructive to use the perturbation theory to see how it works, and when it breaks. As it is the parameter for which the ZZZZ model provides the biggest enhancement, we focus on  $x$ .

### 6.4.1 Full-system estimation

#### Perturbation theory I

Since the interaction Hamiltonian commutes with the free Hamiltonian, the operators are the same in both the interaction and the Schrödinger picture. Especially we have  $H_{\text{int,I}} = \varepsilon H_{\text{int}}$  and the QFI in PT I for  $x$  is equal to

$$\begin{aligned} I_x^{H_{\text{int,I}}} &= 4t^2 \varepsilon^2 K|_{\psi_0} (\partial_x H_{\text{int,I}}(x), \partial_x H_{\text{int}}(x)) + \mathcal{O}(\varepsilon^3) \\ &= 4\varepsilon^2 t^2 \text{Var}[H_{\text{int,I}}(x)]|_{\psi_0} + \mathcal{O}(\varepsilon^3). \end{aligned}$$

This turns out to be equal to the exact solution. To show it we can calculate the local generator for this Hamiltonian  $\mathcal{H} = i e^{i t H(x, \omega_p, \omega_b)} \partial_x e^{-i t H(x, \omega_p, \omega_b)}$ . The crucial point here is that the interaction Hamiltonian commutes with the free Hamiltonian and that each term in the Hamiltonian is of a phase-shift kind. This makes it easy to compute the derivative of the evolution operator:  $\partial_x e^{-i t H(x, \omega_p, \omega_b)} = -i t \varepsilon H_{\text{int}} e^{-i t H(x, \omega_p, \omega_b)}$ . It follows that the local generator is proportional to the interaction Hamiltonian,  $\mathcal{H} = t \varepsilon H_{\text{int}}$  and that the perturbative solution is equal to the exact one:

$$I_x^{H_{\text{int,I}}} = I_x.$$

#### Perturbation theory II

In PT II for the estimation of  $x$  we consider that it is the free Hamiltonian that is the perturbation. Using the Eq. (3.150-3.152) we see that the terms of order one and two in the QFI vanish as  $H_{0,I} = H_0$  which is independent of  $x$ . Only the zeroth order term survives and it is equal to the exact QFI:

$$I_x^{H_{0,I}} = I_x.$$

### 6.4.2 Local estimation

For the local estimation we have to look at the SNR for  $X^{(0)}$ . Notice that we cannot use the perturbative solution derived in [Braun and Martin, 2011] (Eq. (5.24)), as the assumptions are not fulfilled ( $X$  does not commute with the free evolution Hamiltonian of the bus, which is proportional to  $Z$ ). We thus have to use the formula given in Appendix C.3.

We first look at the term of order  $\varepsilon^2$  which reads

$$S_{x, H_{\text{int,I}}}^{X^{(0)}} = \frac{\varepsilon^2 N^2 \cos^2(2\alpha) \sin^2(2\beta) (\sin(\gamma) \cos(\delta\omega_b t) - \cos(\gamma) \sin(\delta\omega_b t))^2}{1 - \sin^2(2\beta) (\sin(\gamma) \sin(\delta\omega_b t) + \cos(\gamma) \cos(\delta\omega_b t))^2} + \mathcal{O}(\varepsilon^3). \quad (6.27)$$

For the optimal case,  $\beta = \pi/4$  and  $\alpha = 0$ , the perturbative SNR reads

$$S_{x, H_{\text{int,I}}}^{X^{(0)}}(\alpha = 0, \beta = \pi/4) = N^2 t^2 \varepsilon^2 + \mathcal{O}(\varepsilon^3), \quad (6.28)$$

which is equal to the exact solution (6.25).

When  $\alpha = \pi/4$  the perturbative solution is equal to zero up to order two:

$$S_{x, H_{\text{int,I}}}^{X^{(0)}}(\alpha = \pi/4) = 0 + \mathcal{O}(\varepsilon^3). \quad (6.29)$$

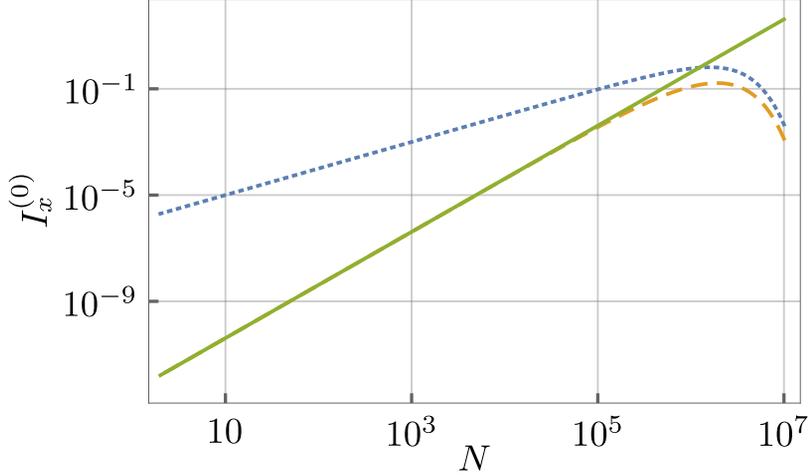


Figure 6.4: Local estimation of  $x$  in the ZZZZ model:  $\beta = \pi/4$ ,  $\alpha = \pi/4$ ,  $t = \omega_p = \omega_b = \delta = x = 1$ ,  $\gamma = 0$  (log-log scale). The scale parameter for the interaction is equal to  $\varepsilon = 10^{-3}$ . The dotted line corresponds to the exact local QFI, the dashed line to the exact SNR  $S_x^{X^{(0)}}$  and the plain line to the perturbative solution up to order four for the SNR (Eq. (6.30)).

This is especially the case for  $\beta = \pi/4$  and  $\alpha = \pi/4$ , whereas the exact solution does not give a vanishing SNR. If we expand the right hand side of (6.26) up to order four in  $\varepsilon$  we obtain

$$S_{x, H_{\text{int}, I}}^{X^{(0)}}(\alpha = \pi/4, \beta = \pi/4) = \frac{N^2 t^4 x^2 \varepsilon^2}{\tan^2(\gamma + \delta \omega_b)} + \mathcal{O}(\varepsilon^5). \quad (6.30)$$

The crucial point is that this term is a pure HL scaling term, while we know that in the exact case we have also an exponential decrease of the SNR with the number of probes. The situation is depicted in Fig. 6.4. Especially we see that for low values of  $N$  the perturbative result for the SNR is in excellent agreement with the exact result. This is expected as we took  $\varepsilon = 10^{-3}$ . But we see that for large  $N$  the perturbative solution completely fails to reproduce the real behaviour of the SNR.

This is extremely important as it shows that the condition for the regime of validity of the perturbation theory cannot be reduced to a condition on the sole scale parameters  $\delta$  and  $\varepsilon$ , but should also include the other parameters and especially the number of probes  $N$ . To make this point clearer we plot in Fig. 6.5 the difference between the exact and the perturbative solution for the SNR. As the perturbative solution is given up to order four, we expect the difference to scale as  $\varepsilon^5$  or more. For  $N = 10$  (bottom lines) we see that the difference scales as  $\varepsilon^6$  for low values of  $\varepsilon$ , before scaling as  $\varepsilon^4$  after  $\varepsilon \simeq 0.2$ . This fits our naive description of the regime of validity based on the sole use of the scale parameters. But when we look at the difference for  $N = 10^7$  (top lines) the situation is drastically different: The scaling in  $\varepsilon^6$  already breaks down at  $\varepsilon = 10^{-3}$ . This clearly shows that  $N$  should be included in the regime of perturbation.

Heuristically a good condition for the validity of perturbation seems to be  $\varepsilon^2 N \ll 1$  for PT I for  $x$ , as the problems in Fig. 6.5 and 6.4 appear roughly at  $\varepsilon^2 N \simeq 1$ . Notice that this limit also corresponds to the failure of local estimation against full estimation in Fig. 6.3 (where both are calculated exactly). There is also a simple analytical argument speaking to introduce  $N$  in the

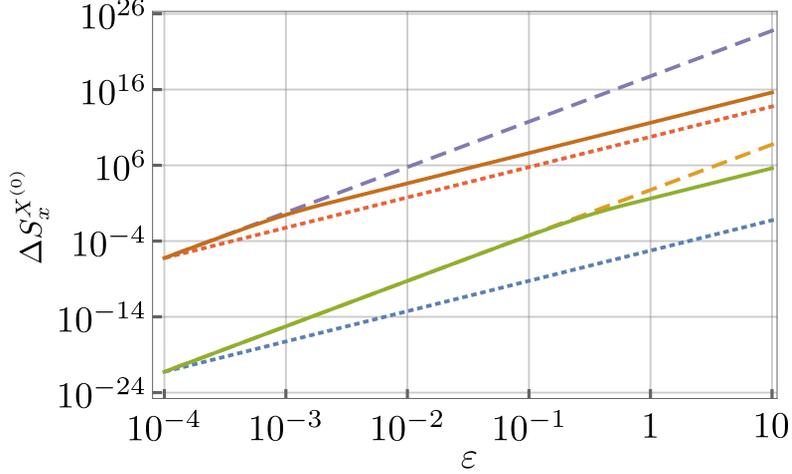


Figure 6.5: Effect of the number of probes  $N$  on the validity of the perturbation theory. The vertical axis corresponds to the difference between the exact result for the SNR and the perturbative result at the order four. The horizontal scale correspond to the scale parameter  $\varepsilon$ . The plot is in a log-log scale. The plain lines correspond to the difference for  $N = 10$  (bottom) and  $N = 10^7$  (top). The dashed lines correspond to a scaling in  $\varepsilon^6$  and the dotted line to a scaling in  $\varepsilon^4$ . The parameters are  $\beta = \pi/4$ ,  $\alpha = \pi/4$ ,  $t = \omega_p = \omega_b = \delta = 1$ ,  $\gamma = 0$ .

condition of validity of perturbation theory. If we look at the term of order three in perturbation theory we see that it is dominated by a factor  $\varepsilon^3 N^3$ . Since we want this term to be small in comparison to the leading order term which scales as  $\varepsilon^2 N^2$  we need that  $\varepsilon N \ll 1$ . Since we consider a large number of probes, this condition is stricter in terms of  $\varepsilon$  than the heuristic one derived from the observations.

## 6.5 Probes in a thermal state

In the previous section we consider pure initial states, where all the probes are in the same state. In order to answer the question how a lack of purity of the initial state affects our results, we now take the  $N$  probes to be in a thermal state. This corresponds to the case where the probes have interacted with a reservoir at a certain temperature  $T$  for long enough. Notice however that we suppose that the interaction with the reservoir was prior to the coherent averaging scheme itself. We do not study a decoherence effect affecting the system during the interaction, rather a thermalisation that happened at the preparation stage.

This resembles the DQC1 protocol where all qubits but one are in a fully mixed state. We have seen in Sec. 4.3.3 that DQC1 based metrological protocols allow to reach the SQL, or to surpass it, by measuring only the control qubit.

### 6.5.1 State and evolution

We still consider symmetric states for the probes. Each probe starts in the state  $\rho_{\text{th}}$  defined as

$$\rho_{\text{th}} = \frac{1}{Z} \begin{pmatrix} e^{-\beta_{\text{th}}\omega_p/2} & 0 \\ 0 & e^{\beta_{\text{th}}\omega_p/2} \end{pmatrix}, \quad (6.31)$$

where  $Z = e^{-\beta_{\text{th}}\omega_p/2} + e^{+\beta_{\text{th}}\omega_p/2}$ ,  $\beta_{\text{th}} = 1/(k_B T)$ ,  $T$  is the temperature, and  $k_B$  is the Boltzmann constant. We consider that  $\delta = 1$  during the thermalisation stage. The quantum bus is still in the pure state  $|\xi\rangle = \cos(\beta)|0\rangle + \sin(\beta)e^{i\gamma}|1\rangle$ . The initial state of the full system is thus

$$\rho_{0,\text{th}} = \rho_{\text{th}}^{\otimes N} \otimes |\xi\rangle_{00}\langle\xi|. \quad (6.32)$$

It is easier to work in the effective spin  $N/2$  basis  $\{|j, m\rangle \equiv |m\rangle\}$ . The state can be written as a direct sum:

$$\rho_{0,\text{th}} = \frac{1}{Z^N} \bigoplus_{m=0}^N \sqrt{\binom{N}{m}} e^{-(N-2m)\beta_{\text{th}}\omega_p/2} | \frac{N}{2} - m \rangle \langle \frac{N}{2} - m | \otimes |\xi\rangle_{00}\langle\xi|. \quad (6.33)$$

The state after evolution during a time  $t$  is given by

$$\begin{aligned} \rho_{\text{th}}(t) = & \bigoplus_{m=0}^N \sqrt{\binom{N}{m}} \frac{e^{-(N-2m)\beta_{\text{th}}\omega_p/2}}{Z^N} | \frac{N}{2} - m \rangle \langle \frac{N}{2} - m | \otimes \left\{ \cos^2(\beta)|0\rangle_{00}\langle 0| + \sin^2(\beta)|1\rangle_{00}\langle 1| \right. \\ & \left. + \cos(\beta)\sin(\beta)(e^{-i(\gamma+\delta t\omega_b+(N-2m)\varepsilon tx)}|0\rangle_{00}\langle 1| + e^{i(\gamma+\delta t\omega_b+(N-2m)\varepsilon tx)}|1\rangle_{00}\langle 0|) \right\}. \end{aligned} \quad (6.34)$$

Each block spanned by  $\{| \frac{N}{2} - m \rangle \otimes |0\rangle_0, | \frac{N}{2} - m \rangle \otimes |1\rangle_0$  has only one non-zero eigenvalue

$$\lambda_{1,m}^{\text{th}} = \frac{e^{-(N-2m)\beta_{\text{th}}\omega_p/2}}{Z^N}, \quad \lambda_{0,m}^{\text{th}} = 0, \quad (6.35)$$

the corresponding eigenvectors being

$$|v_{1,m}^{\text{th}}\rangle = | \frac{N}{2} - m \rangle \otimes (\cos(\beta)e^{-i(\gamma+t\delta\omega_b+(N-2m)\varepsilon tx)}|0\rangle_0 + \sin(\beta)|1\rangle_0), \quad (6.36)$$

$$|v_{0,m}^{\text{th}}\rangle = | \frac{N}{2} - m \rangle \otimes (\sin(\beta)e^{-i(\gamma+t\delta\omega_b+(N-2m)\varepsilon tx)}|0\rangle_0 - \cos(\beta)|1\rangle_0). \quad (6.37)$$

### 6.5.2 Full-system estimation for thermal states

We start by analysing the efficiency of the protocol when one has access to the full system. Since the density matrix  $\rho_{\text{th}}(t)$  is a direct sum of  $2 \times 2$  matrices we can calculate directly the QFI by diagonalizing each block and compute the corresponding QFI.

**Full-system estimation of  $x$**  Looking at Eq. (6.35-6.37) we see that the QFI for  $x$  has only its quantum contribution since the parameter  $x$  does not appear in the eigenvalues. The calculation of the QFI for  $x$  reads

$$I_x^{\text{th}} = \sin^2(2\beta)\varepsilon^2 t^2 \left( N^2 \tanh^2(\beta_{\text{th}}\omega_p/2) + N(1 - \tanh^2(\beta_{\text{th}}\omega_p/2)) \right). \quad (6.38)$$

Remarkably we can still achieve the HL scaling for the estimation of  $x$  when all our probes are in a thermal state. The dependence of HL terms on the temperature is given by the hyperbolic function  $\tanh^2(\beta_{\text{th}}\omega_p/2) = \tanh^2(\frac{\omega_p}{2k_{\text{B}}T})$ , while the SQL term depends on the temperature as  $1 - \tanh^2(\beta_{\text{th}}\omega_p)$ . The function  $\tanh^2(\frac{\omega_p}{2k_{\text{B}}T})$  goes from zero for infinite temperatures to one for vanishing temperature, leading to a QFI dominated by the SQL term for high temperatures, and by the HL term for low temperatures.

**Full-system estimation of  $\omega_p$**  As we can see in the Eq. (6.35) there is a dependence on  $\omega_p$  in the eigenvalues of the density matrix but there is no dependence on  $\omega_p$  in the eigenvectors: The QFI  $I_{\omega_p}^{\text{th}}$  is purely classical in this case. Its calculation reads

$$I_{\omega_p}^{\text{th}} = N\beta_{\text{th}}^2(1 - \tanh^2(\beta_{\text{th}}\omega_p/2))/4. \quad (6.39)$$

There is only an SQL scaling term and no time dependence. We should not mistake the origin of this SQL term. It is not due to the free evolution of the probes, but only to the initial  $\omega_p$ -dependence of the thermal state. Since the initial state of the probe has no coherence, in the eigenbasis of  $Z$  the free evolution does not imprint any phase on the probes. We could completely skip the coherent averaging protocol and get the same result.

**Full-system estimation of  $\omega_b$**  The QFI for  $\omega_b$  when starting with a thermal initial state of the probes is equal to

$$I_{\omega_b}^{\text{th}} = \delta^2 t^2 \sin^2(2\beta). \quad (6.40)$$

As we see this is equal to the QFI starting with pure states. From this point of view the initial state of the probes do not play any role in the full-system estimation of  $\omega_b$ .

### 6.5.3 Local estimation for thermal states

As we did in the case of pure initial states, we study the local estimation when one has just a restricted access to the system and can measure only the quantum bus. The calculation of the reduced density matrix of the bus reads:

$$\rho_{\text{th,bus}}(t) = \begin{pmatrix} \cos^2(\beta) & \left( e^{-i\epsilon t x - \beta_{\text{th}}\omega_p/2} + e^{i\epsilon t x + \beta_{\text{th}}\omega_p/2} \right)^N \\ \left( e^{i\epsilon t x - \beta_{\text{th}}\omega_p/2} + e^{-i\epsilon t x + \beta_{\text{th}}\omega_p/2} \right)^N & \times e^{-i(\gamma + \delta\omega_b t)} \sin(2\beta)/(2Z^N) \\ \times e^{i(\gamma + \delta\omega_b t)} \sin(2\beta)/(2Z^N) & \sin^2(\beta) \end{pmatrix}. \quad (6.41)$$

Remarkably, the reduced density matrix has the same form as the one for the pure state (6.10) when setting

$$\cos^2(\alpha_{\text{th}}) = e^{-\beta_{\text{th}}\omega_p/2} / Z \quad \text{and} \quad \sin^2(\alpha_{\text{th}})^2 = e^{\beta_{\text{th}}\omega_p/2} / Z, \quad (6.42)$$

and identifying the *mixing angle*  $\alpha_{\text{th}}$  with the original angle  $\alpha$  of the qubit. This implies that for any initial pure product state (6.2) there exists an initial thermal state that leads to the same local QFI for  $x$  and for  $\omega_b$ . Especially we reach the HL for a zero temperature and the SQL for an infinite temperature. For a finite temperature the local QFI will have an HL-like term and an SQL-like term. Still we should moderate our enthusiasm as we saw that it seems that apart when

starting exactly with  $\beta = \pi/4$  and  $\alpha = 0$  the local QFI goes in general to zero for large  $N$ , and  $\alpha_{\text{th}}$  is equal to zero only at zero temperature, which would mean that at finite temperature the HL scaling breaks down.

**Estimation of  $\omega_p$**  For the local estimation of  $\omega_p$  the identification with the state (6.10) does not hold anymore since  $\alpha_{\text{th}}$  depends on  $\omega_p$ . While for the pure state it is impossible to estimate  $\omega_p$  by measuring only the bus, for thermal states it becomes possible. This is because the initial thermal state depends already on  $\omega_p$ .

## Summary Chapter 6

- **ZZZZ model:** Probes and quantum bus are qubits. Hamiltonian is  $H = \delta \left( \frac{\omega_p}{2} \sum_{i=1}^N Z^{(i)} + \frac{\omega_b}{2} Z^{(0)} \right) + \varepsilon \left( \frac{x}{2} \sum_{i=1}^N Z^{(i)} Z^{(0)} \right)$ . Since the free interaction and the perturbation commute, and since all the terms are of phase-shift kind, we can solve the dynamics exactly. We start with a separable initial state.
- **Full-system estimation:** We can reach the HL for  $x$ , and in general obtain an HL scaling for a large class of initial states. We obtain only an SQL scaling for  $\omega_p$ .
- **Local estimation:** We can reach the HL for  $x$  by measuring only the bus. The problem is that apparently for almost any other state but the optimal one the local QFI goes to zero for large  $N$ .
- **Limits of the perturbation theory:** Comparison with perturbation theory helps to show that the regime of validity depends on the number of probes. Raises the question of the validity of perturbation theory for a large number of probes.
- **Thermal states:** When starting with the probes in an initial thermal state we can still reach the HL in full-system estimation. For local estimation the situation is formally equivalent to the pure state case, with the effective polar angle of the probes being a function of the temperature.

# Chapter 7

## Qubit model: ZZXX Hamiltonian<sup>1</sup>

In the previous section we considered a model of coherent averaging where the free Hamiltonian commutes with the interaction Hamiltonian. The dynamics of the system could be solved exactly allowing us to calculate the QFI for each parameter in full-system estimation and in local estimation. We could also calculate perturbative solutions to have a better idea of the limits of the perturbation theory. In this simple setup, the coherent averaging scheme improved the scaling for the estimation of  $x$ , but not for the estimation of  $\omega_p$  and  $\omega_b$ .

In this chapter we introduce a model of coherent averaging where the free Hamiltonian does not commute with the interaction Hamiltonian.

### 7.1 General qubit model of coherent averaging

Before going to the ZZXX Hamiltonian we look at a general model for coherent averaging based on qubits. We want to identify the conditions — using perturbation theory — under which we can expect an HL scaling for the estimation of  $\omega_p$ .

#### 7.1.1 General qubit model

We introduce the general Hamiltonian for a qubit-based coherent averaging protocol:

$$H(x, \omega_p, \omega_b) = \delta H_0 + \varepsilon H_{\text{int}} \quad (7.1)$$

$$= \delta \left( \frac{\omega_p}{2} \sum_{i=1}^N F_p^{(i)} + \frac{\omega_b}{2} F_b^{(0)} \right) + \varepsilon \frac{x}{2} \sum_{i=1}^N S_p^{(i)} \otimes R_b^{(0)}, \quad (7.2)$$

with  $F_p = \mathbf{f}_p \cdot \boldsymbol{\sigma}$ ,  $F_b = \mathbf{f}_b \cdot \boldsymbol{\sigma}$ ,  $S_p = \mathbf{s}_p \cdot \boldsymbol{\sigma}$ , and  $R_b = \mathbf{r}_b \cdot \boldsymbol{\sigma}$ .

We study this Hamiltonian in the framework of PT I. We look at the estimation of  $\omega_p$  for strong interactions. We do not look at PT II as there we already know that the leading order term corresponds to the unperturbed QFI, prohibiting an HL term for  $\omega_p$ .

---

<sup>1</sup>This chapter is based on: "Coherent averaging", Fraïsse, J. M. E. and Braun, D. (2015), *Annalen der Physik*, 527(9-10):701–712. All the figures and parts of the discussion are reproduced from there. ©2015 Annalen der Physik

### 7.1.2 QFI for $\omega_p$ in PT I

We seek to calculate the QFI for  $\omega_p$  in formalism I of perturbation theory. Using the formula (3.147) we obtain

$$\begin{aligned}
 I_{\omega_p}^{H_0, I} = & \frac{N^2}{x^2} \langle \Upsilon_1 \rangle^2 (1 - \cos(xt))^2 (\langle R_b \rangle^2 - 1) \\
 & + \frac{N}{x^2} \left\{ (xt)^2 - i xt \langle R_b \rangle (1 - \cos(xt)) \langle \{F_p, \Upsilon_1\} \rangle + xt (\sin(xt) - xt) \langle \{F_p, \Upsilon_2\} \rangle \right. \\
 & \quad - \langle \Upsilon_1^2 \rangle (1 - \cos(xt))^2 + \langle \Upsilon_2^2 \rangle (\sin(xt) - xt)^2 - i \langle \{ \Upsilon_1, \Upsilon_2 \} \rangle \langle R_b \rangle \\
 & \quad \times (1 - \cos(xt)) (\sin(xt) - xt) - \langle F_p \rangle^2 (xt)^2 + 2i xt \langle F_p \rangle \langle \Upsilon_1 \rangle \langle R_b \rangle (1 - \cos(xt)) \\
 & \quad - 2 \langle F_p \rangle \langle \Upsilon_2 \rangle (\sin(xt) - xt) xt + \langle \Upsilon_1 \rangle^2 (1 - \cos(xt))^2 - \langle \Upsilon_2 \rangle^2 (\sin(xt) - xt)^2 \\
 & \quad \left. + 2i \langle \Upsilon_1 \rangle \langle \Upsilon_2 \rangle \langle R_b \rangle (\sin(xt) - xt) (1 - \cos(xt)) \right\}, \quad (7.3)
 \end{aligned}$$

where we introduced the commutators

$$\Upsilon_1 = [F_p, S_p] / 2 = i (\mathbf{f}_p \times \mathbf{s}_p) \cdot \boldsymbol{\sigma}, \quad \Upsilon_2 = [\Upsilon_1, S_p] / 2 = (\mathbf{s}_p - (\mathbf{f}_p \cdot \mathbf{s}_p) \mathbf{f}_p) \cdot \boldsymbol{\sigma}. \quad (7.4)$$

Other than in the ZZZZ model where we never reach HL scaling for  $\omega_p$ , with a general Hamiltonian it is possible to reach this scaling. We see that if  $\Upsilon_1$  is equal to zero, then the HL-like term disappears. As suspected we conclude that *a necessary condition for observing an HL scaling for  $\omega_p$  is that  $F_p$  and  $S_p$  do not commute.*

The situation is different for  $x$ , since setting  $\Upsilon_1$  to zero does not lead to the extinction of the HL scaling part of the QFI, as demonstrated by the ZZZZ model.

Again the mechanism at the base of enhancement in the coherent averaging protocol appears: It is the interaction part of the Hamiltonian that allows to reach the HL scaling. Since  $x$  is linked to the interaction we do not need any condition on the commutator with the free evolution part, but for  $\omega_p$  we need a mixed dynamics in order to beat the SQL scaling.

## 7.2 ZZXX Hamiltonian and methods

### 7.2.1 The ZZXX Hamiltonian

As we want to use numerical methods we need to introduce a specific Hamiltonian. To fulfil the non-commuting condition we take the interaction to be " $X \otimes X$ ":

$$H(x, \omega_p, \omega_b) = \delta \left( \frac{\omega_p}{2} \sum_{i=1}^N Z^{(i)} + \frac{\omega_b}{2} Z^{(0)} \right) + \varepsilon \left( \frac{x}{2} \sum_i^N X^{(i)} X^{(0)} \right), \quad (7.5)$$

We refer to this Hamiltonian as the ZZXX Hamiltonian and the corresponding coherent averaging model as the ZZXX model. Using the spin- $N/2$  notation the Hamiltonian is written as

$$H(x, \omega_p, \omega_b) = \delta \left( \frac{\omega_p}{2} J_z \otimes \mathcal{I} + \frac{\omega_b}{2} Z^{(0)} \right) + \varepsilon \frac{x}{2} (J_x \otimes \mathcal{I}) X^{(0)}. \quad (7.6)$$

The initial state is the same as we used for the ZZZZ model:

$$|\psi_0\rangle = \bigotimes_{i=1}^N (\cos(\alpha) |0\rangle_i + \sin(\alpha) e^{i\phi} |1\rangle_i) \otimes (\cos(\beta) |0\rangle_0 + \sin(\beta) e^{i\gamma} |1\rangle_0). \quad (7.7)$$

For the ZZZZ Hamiltonian, the dynamics due to the free Hamiltonian and the dynamics due to the interaction Hamiltonian were decoupled. The Hamiltonians were commuting, and thus also the evolution operator. The two dynamics sum up. In this new configuration the two dynamics get intricate, and we shall see how this enrichment in the dynamics of the system will be useful in terms of estimation. The price to pay is that now we do not have the possibility to solve exactly the dynamics of the system. We will still use perturbation theory, and instead of exact analytical solution we will use numerics. Notice that we will use the term "exact" in opposition to "perturbative" and not to "numerical", which is itself opposed to "analytical".

### 7.2.2 Numerics

Since it is difficult to diagonalize the Hamiltonian, we can do it numerically. Diagonalization of matrices is a common task in physics and there exist very efficient routines to do it. The computational cost is high and there is no parallel algorithm to compute the task. From this point of view, we see that it is extremely useful to work with symmetric initial states (for the probes), since the matrix that we diagonalize has size  $4N \times 4N$  instead of  $2^{N+1} \times 2^{N+1}$ .

After finding the eigenvalues and eigenvectors of the Hamiltonian we can calculate the evolution operator and thus compute the final state. But in order to compute the QFI we need the derivative with respect to the parameter to be estimated of the state. In order to do this numerically we should consider a second Hamiltonian with a small shift  $h$  in the parameter, calculate the corresponding final state and then we compute a numerical approximation of the derivative of the state:

$$|\partial_\theta \psi_\theta\rangle = \frac{|\psi_\theta\rangle - |\psi_{\theta+h}\rangle}{h} + \mathcal{O}(h). \quad (7.8)$$

The value of  $h$  has to be chosen carefully, and will depend on the numerical precision we are working. In practice we first tested the stability of the numerical derivation for small  $N$  with different values of  $h$ . Finally we did the calculations with  $h = 10^{-6}$  and  $h = 10^{-8}$ .

Once that we have the evolved state and its derivative we can calculate all our functions used to quantify the efficiency of the scheme:

- the full-system QFI: With the state and its derivative the calculation is direct:  $I_\theta = 4(\langle \partial_\theta \psi_\theta | \partial_\theta \psi_\theta \rangle - |\langle \psi_\theta | \partial_\theta \psi_\theta \rangle|^2)$ .
- the local QFI: We should calculate the reduced density matrix, and then diagonalize it. We use the basis independent formula (3.49) and solve the integral analytically. This calculation requires to diagonalize the density matrix in order to calculate the exponential of the matrix  $e^{-\rho_{\theta,0}}$ . The reduced density matrix of the bus  $\rho_{\theta,0}$  being a two by two matrix, this task can be done analytically.
- the SNR for a local observable: We have to calculate the variance of an observable  $A^{(0)}$  and the derivative of its expectation value. This can be done using the derivative of the state:  $\partial_\theta \langle \psi_\theta | A | \psi_\theta \rangle = \langle \partial_\theta \psi_\theta | A | \psi_\theta \rangle + \langle \psi_\theta | A | \partial_\theta \psi_\theta \rangle$ .

The complete algorithm takes roughly the following form:

1. creation of the initial state  $|\psi_0\rangle$
2. creation of two Hamiltonians with a small difference  $h$  in the parameter to be estimated:  $H(\theta)$  and  $H(\theta + h)$

3. diagonalization of  $H(\theta)$  and  $H(\theta + h)$
4. calculation of  $|\psi_\theta\rangle$  and  $|\psi_{\theta+h}\rangle$
5. calculation of  $\partial_\theta|\psi_\theta\rangle$  using  $|\psi_\theta\rangle$  and  $|\psi_{\theta+h}\rangle$  (finite difference)
6. calculation of the QFI with  $|\psi_\theta\rangle$  and  $\partial_\theta|\psi_\theta\rangle$
7. partial trace of  $|\psi_\theta\rangle\langle\psi_\theta|$
8. calculation of the local QFI
9. calculation of the SNR

### 7.3 Full-system estimation on the ZZXX model

We start by the full-system estimation. As the perturbative results in PT I are usually cumbersome we do not write explicitly the formulas and rely on the study of the plots for discussing the efficiency. All figures shown have the parameters  $\omega_p = 1$ ,  $\omega_b = 1$ ,  $x = 1$ ,  $t = 1$ . The initial state corresponds to (7.7) with  $\alpha = \pi/3$ ,  $\beta = \pi/6$ ,  $\phi = 3\pi/8$ ,  $\gamma = 5\pi/8$  unless otherwise indicated. We choose these values as they give a typical result, in the sense that they do not correspond to a particular optimal or sub-optimal state for this Hamiltonian.

#### Full-system estimation of $x$

In Fig. 7.1 we plotted the full-system QFI for  $x$  in the regime of weak, medium and strong interactions as a function of the number of probes  $N$ . The plots are in a log-log scale and we can check easily the scaling of the QFI, since polynomial functions  $f(N) = N^a$  are represented on it by straight lines with slope  $a$ . The value of  $\delta$  is set to one for the three cases.

**Weak interaction** The top plot corresponds to the weak regime of interaction, with  $\varepsilon = 0.001$ . We plot the QFI with blue X-symbols, and the solution in PT I with magenta circles. The green line represents the SQL scaling — *i.e.* a linear scaling — shifted to the origin of the QFI (we plotted  $f(N) = I_x(N=1)N$ ), while the orange line represents the HL scaling — *i.e.* a quadratic scaling — shifted to the origin of the QFI (we plotted  $f(N) = I_x(N=1)N^2$ ).

We notice that there is a very good agreement between the exact numerical QFI and the perturbative one, although for the last point ( $N = 4000$ ) the difference between them starts to be noticeable. This is again a hint that the perturbative result may not hold for very large  $N$  although it holds for smaller  $N$ .

In terms of scaling, up to  $N = 100$  the behavior follows closely an SQL scaling, the points are very close to the green line. After  $N = 100$  we see that the QFI deviates from the SQL scaling gets closer to the HL scaling. Nevertheless there are not enough points to asses that we observe a HL scaling. We know that the perturbation theory predicts an HL component for the QFI. If we plot the solution in perturbation theory for larger values of  $N$  we will see the HL scaling dominating, but due to computational limitations we cannot compare it with the exact result.

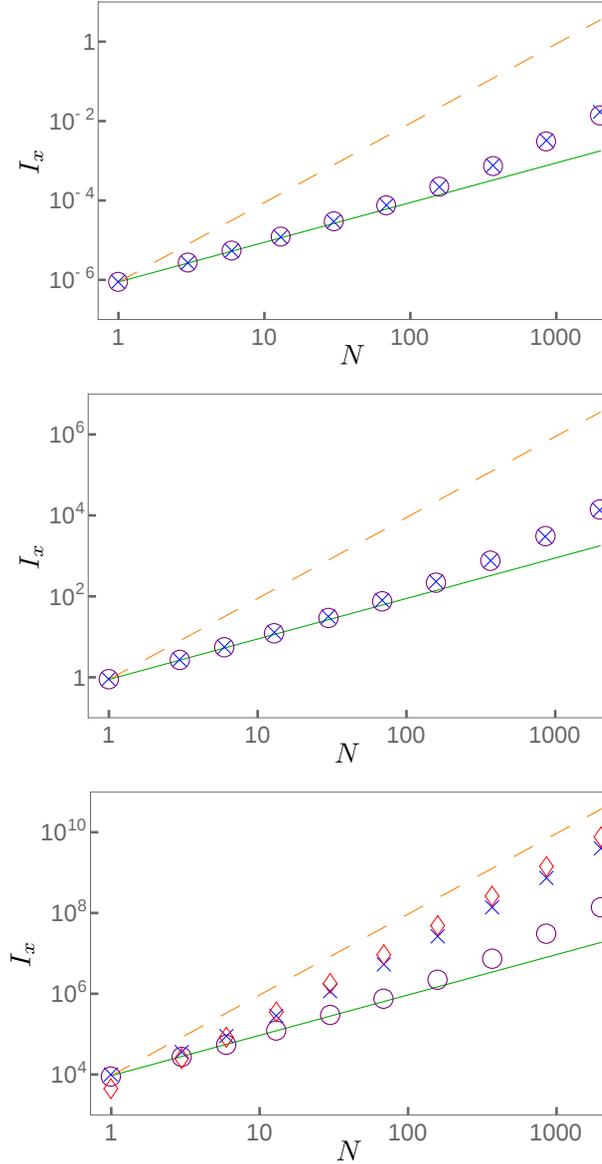


Figure 7.1: Full-system QFI for the ZZXX model for  $x$  for weak, medium, and strong interactions (from top to down:  $\varepsilon = 0.001$ ,  $1 \leq \varepsilon = 100$ , and  $\delta = 1$ ). Blue X-symbols: exact numerical results. Purple circles: perturbative result (PT I). Red diamonds: zeroth order (unperturbed) term in PT II. The dashed orange (resp. green continuous) lines represent  $f(N) \propto N^2$  (resp.  $N$ ).

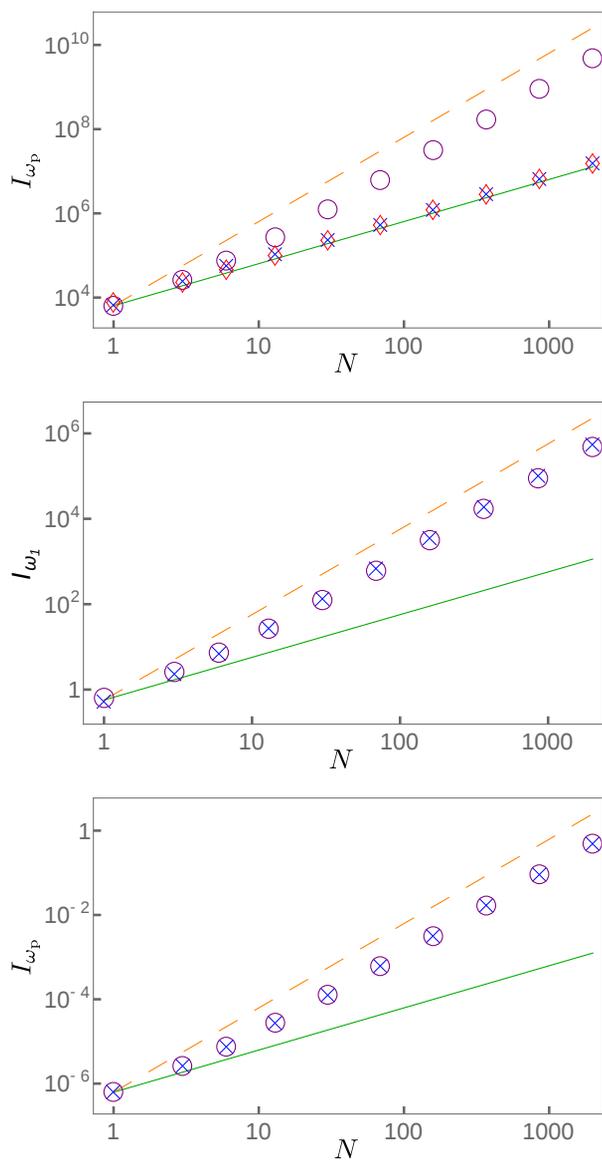


Figure 7.2: Full-system QFI for the ZZXX model for  $\omega_p$  for weak, medium, and strong interactions (from top to down:  $\delta = 100, 1, \delta = 0.001$ , and  $\varepsilon = 1$ ). Blue X-symbols: exact numerical results. Purple circles: perturbative result (PT1). Red diamonds: zeroth order (unperturbed) term in PT2. The dashed orange (resp. green continuous) lines represent  $f(N) \propto N^2$  (resp.  $N$ ).

**Medium interaction** The middle plot corresponds to the medium interaction regime with  $\varepsilon = 1$ .

Again the blue X-symbols represent the exact QFI calculated numerically. In this regime, we do not have any results from perturbation theory, since the interaction part of the Hamiltonian is of the same order of magnitude as the free evolution part. Nevertheless we also plotted the solution in PT I, *the one suited for weak interaction*. Surprisingly we observe that this solution is in very good agreement with the exact numerical solution.

The behavior of the QFI is pretty similar to the behaviour in the weak regime: it starts with an SQL scaling and ends up with an intermediary scaling between SQL and HL.

**Strong interaction** The bottom plot corresponds to the strong interaction regime, with  $\varepsilon = 100$ . The blue X-symbols still represent the exact numerical QFI. The red diamonds correspond to the zeroth order term in the framework of PT II. It means that this is the QFI in the case where there is no free evolution. We see that this term gives the good scaling of the QFI. Like before the purple circles corresponds to PT I, and as expected it does not reproduce the correct behavior for the QFI.

**Metrology, scaling and QFI** In the three interaction regimes we exclusively discussed the efficiency of the scheme in terms of scaling. Nevertheless *in fine* what matters is the value of the QFI. From this point of view we see that the QFI in the weak interaction regime starts at  $10^{-6}$ , in the medium regime at 1 and in the strong regime at  $10^4$ . Obviously the higher the interaction the higher the QFI.

### 7.3.1 Full-system estimation of $\omega_p$

We look at the estimation of the parameter of the free evolution of the probes,  $\omega_p$ . The QFI in PT I is given by Eq. (7.3). In order to explore the different regimes we set  $\varepsilon$  to one and we let  $\delta$  vary.

We plotted in Fig. 7.2 the exact numerical solution and the solution found with perturbation theory for the three regimes of weak, medium and strong interactions.

**Weak interaction** In the top plot, the scale parameters are  $\varepsilon = 1$  and  $\delta = 100$  corresponding to the regime of weak interaction (the free evolution is dominating). The exact numerical QFI is represented by the blue X-symbols. The red diamonds correspond to the zeroth order term in PT II, *i.e.* the QFI obtained when we completely neglect the interaction Hamiltonian. It is thus not surprising to see that the scaling is an SQL scaling.

We see that the agreement between the exact solution and the zeroth order term on PT II is very good. Even for large  $N$  the exact QFI follows completely the SQL scaling. We also plotted the solution in PT I, which is not suited for this regime, and indeed fails completely to reproduce the scaling of the QFI (purple circles).

**Medium interaction** The middle row plot corresponds to the medium regime where both scaling parameters are set to one:  $\varepsilon = \delta = 1$ . Although PT I is not suppose to work here, we see that it succeeds (purple circles) to describe the behavior of the exact QFI (blue X-symbol). The agreement is not perfect, but gives the good scaling and the errors are relatively small.

In terms of scaling we start by a scaling that looks like SQL (up to  $N \sim 10$ ) and then goes to an HL scaling in a clear way. This is interesting as it shows that even outside the range of validity of perturbation theory we can expect to estimate  $\omega_p$  with an HL scaling.

**Strong interaction** In the regime of strong interaction (bottom plot) we set  $\delta = 0.001$ . We can notice in the plot that as expected PT I (purple circles) is an excellent approximation of the exact QFI (blue X-symbol). We cannot notice the difference in the plot.

The behavior is very similar to the one for medium interaction. Starting by SQL for small value of  $N$  and thus being dominated by a HL scaling. Nevertheless, even if the scaling is similar, the values of the QFI are weaker in this case.

### 7.3.2 Full-system estimation of $\omega_b$

Due to the form of the free evolution Hamiltonian of the bus in the interaction picture we are not able to carry out the calculation in PT I. We can only rely on numerics. The numerical results for the three different regimes are depicted in the Fig. 7.3.

Setting  $\varepsilon = 1$ , the weak interaction regime corresponds to  $\delta = 100$  (top plot), the medium interaction regime to  $\delta = 1$  (middle plot) and the strong interaction regime to  $\delta = 0.001$  (bottom plot). In all plots we see that the coherent averaging scheme *proves to be counter productive*. Indeed we see that the QFI is a decreasing function of the number of probes. The more probes we add the less efficient is the estimation procedure.

Since the local estimation can only be worse than the full-system estimation, we will not investigate the local estimation of  $\omega_b$ .

## 7.4 Local estimation on the ZZXX model

As we did for the ZZZZ Hamiltonian, we investigate the metrological properties of the scheme when one has only access to the quantum bus. We calculate the local QFI and the SNR with a local observable.

The local QFI will be calculated only exactly with numerics, while the SNR will be calculated both in perturbation theory and exactly with numerics. For the choice of the observable we take  $A^{(0)} = (X^{(0)} + Z^{(0)})/2$ . In the same way as for the choice of the state or of the Hamiltonian, this specific case can be considered as a generic example. What is important is that the observable does not commute with the Hamiltonian. Notice that with this choice we are not in the simpler case described in [Braun and Martin, 2011] and we should use the most general form of the perturbation theory given in Appendix C.3.

### 7.4.1 Local estimation of $x$

Both figures of merit for the local estimation for  $x$  in the three different regimes are plotted in Fig. 7.4. As we did for the full-estimation, we set  $\delta$  to one and we vary  $\varepsilon$  in order to reach the weak, medium and strong interaction regimes.

**Weak interaction** This is the top plot and it corresponds to  $\varepsilon = 0.001$ . The blue X-symbols represent the local QFI, while the red crosses represent the exact numerical results for the SNR. The magenta circles stand for the perturbative SNR.

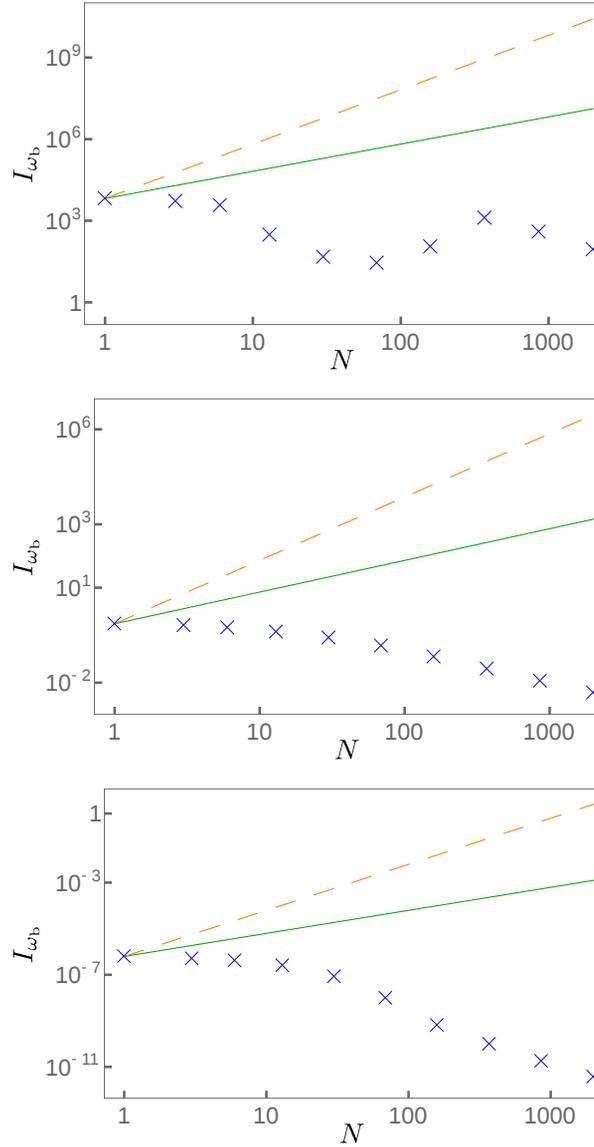


Figure 7.3: Global QFI for the ZZXX model for  $\omega_b$  for weak, medium, and strong interactions (from top to down:  $\delta = 100$ ,  $\delta = 1$ ,  $\delta = 0.001$  and  $\varepsilon = 1$ ). Blue X-symbols: exact numerical results for the QFI for  $\omega_b$ . The dashed orange (resp. green continuous) lines represent  $f(N) \propto N^2$  (resp.  $N$ ).

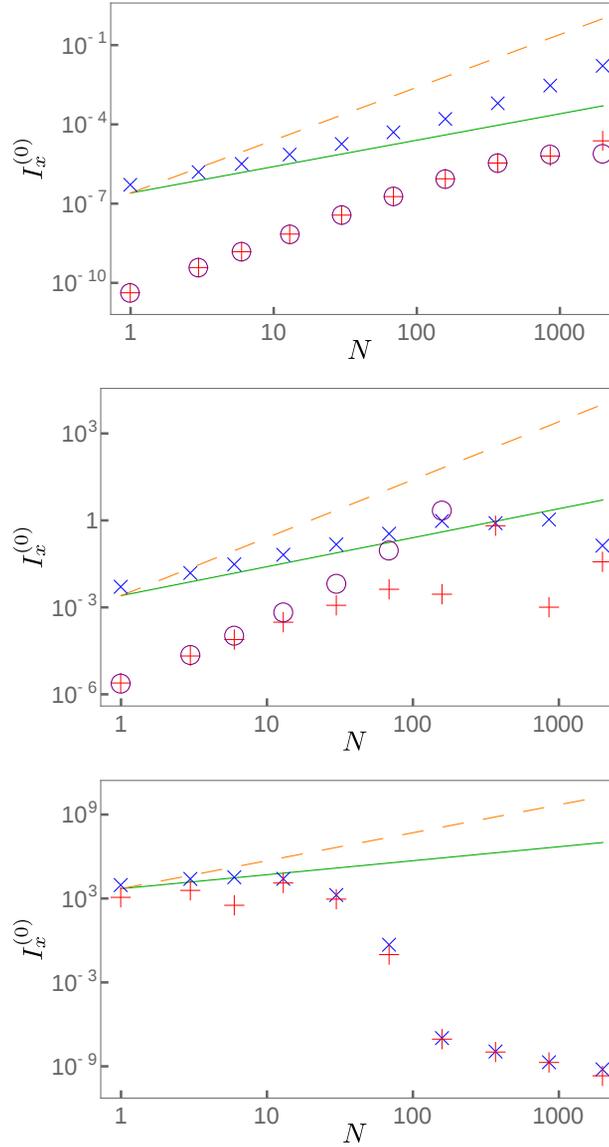


Figure 7.4: Local QFI and inverse squared uncertainties of  $x$  based on the local observable  $A^{(0)} = (X^{(0)} + Z^{(0)})/2$  for the ZZXX model for weak, medium, and strong interactions,  $\varepsilon = 0.001, 0.1, \varepsilon = 100$ , and  $\delta = 1$  from top to bottom ; Blue X-symbols: exact numerical result for  $I_x^{(0)}$ . Purple circles: perturbative solution for  $(\delta_x^{A^{(0)}})^{-2}$ . Red crosses: exact solution for  $(\delta_x^{A^{(0)}})^{-2}$ . The dashed orange (resp. green continuous) lines represent  $f(N) \propto N^2$  (resp.  $N$ ).

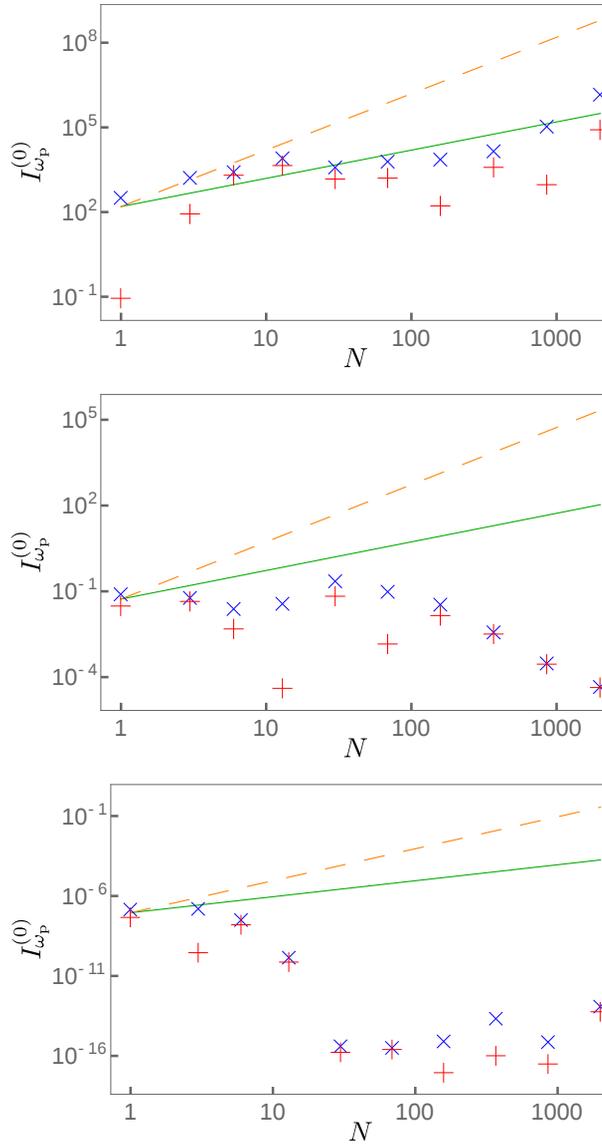


Figure 7.5: Local QFI and inverse squared uncertainties of  $\omega_p$  based on the local observable  $A^{(0)} = (X^{(0)} + Z^{(0)})/2$  for the ZZXX model for weak, medium, and strong interactions,  $\delta = 100$ ,  $1$ ,  $\delta = 0.001$ , and  $\varepsilon = 1$  from top to bottom. Blue X-symbols: exact numerical result for  $I_x^{(0)}$ . Purple circles: perturbative result for  $(\delta_x^{A^{(0)}})^{-2}$ . Red crosses: exact result for  $(\delta_x^{A^{(0)}})^{-2}$ . The dashed orange (resp. green continuous) lines represent  $f(N) \propto N^2$  (resp.  $N$ ). The saturation at  $10^{-16}$  reached in the bottom plot corresponds to the numerical precision.

First we check that the SNR of the observable is lower than the QFI, as it should be according to the chain of inequalities (5.6). We see that the local QFI starts clearly with an SQL scaling and slowly goes to an HL one. This is similar to the behaviour of the full-estimation QFI (Fig. 7.1). It is an encouraging result as it suggests that we could reach an HL scaling for  $x$  by measuring only the bus. The SNR for  $A^{(0)}$  starts with an HL scaling, but then seems to converge to an SQL scaling. We anyway see that this specific measurement does not saturates, nor approach the local QFI.

The solution given by perturbation theory is in good agreement with the exact solution for small values of  $N$ , but for the last two points we see that the perturbative solution starts to differ from the exact solution ( $N = 4000$  especially).

**Medium interaction** Here we set  $\varepsilon = \delta = 1$ , the plot is in the middle row of the figure. The exact local QFI, still the blue X-symbol, starts with an SQL scaling, like in the weak interaction regime, but instead of going slowly into a HL scaling with increasing  $N$ 's, it suddenly decreases after  $N = 200$ .

This is in great contrast to the case of full-system estimation where with increasing  $N$  the scaling was going into an HL one. It means that in this case we cannot achieve the best sensitivity with a local measurement, but moreover the local measurement becomes counter productive by increasing too much  $N$ .

The red crosses show the SNR by measuring  $A^{(0)}$ . As it is supposed the perturbation solution fails in reproducing the behaviour of the exact solution in this regime.

**Strong interaction** The bottom plot corresponds to the strong interaction regime ( $\varepsilon = 100$ ). In the same way as for the medium regime the exact local QFI decreases after a certain threshold on  $N$ . Here the threshold is even lower since after  $N \sim 10$  the QFI decreases extremely fast. The SNR for  $A^{(0)}$  is this time almost optimal, in the sense that it saturates the QCRB. Nevertheless this is not very useful since the scheme does not seem to be very suited for metrological tasks.

The situation looks even worse when comparing it with the full-system estimation cases. In the strong interaction regime, the estimation of  $x$  was done with an HL scaling and large values of the QFI. Restricting ourselves to measurement on the quantum bus only seems to kill the metrological interest of the method. We should still remember that these results hold just for the interval of  $N$  represented in the figure. We do not know if after a certain value of  $N$  the QFI will not increase again and which scaling it will have.

## 7.4.2 Local estimation of $\omega_p$

The last point to investigate is the local estimation of  $\omega_p$ . As we did for the full-system estimation we set  $\varepsilon$  to one and vary  $\delta$ . The results are presented in Fig. 7.5.

**Weak interaction** The top plot corresponds to  $\delta = 100$ . We see that the exact local QFI (blue X-symbol) scales for small  $N$  with the SQL while for large  $N$  it seems to follow an HL scaling. Nevertheless the behaviour is far too erratic to really say anything about the scaling. Moreover, the full-system QFI in this regime follows an SQL scaling and thus limits also the local QFI to an SQL scaling at most.

In terms of SNR we see again that the chain of inequalities (5.6) is respected, the red crosses being always below the local QFI.

Table 7.1: Summary for the ZZZZ model

|            | Full-system estimation          | Local estimation  | Thermal states (probes)   |
|------------|---------------------------------|---|---|
| $x$        | HL scaling for almost any state | HL scaling but very sensitive to initial state. Seems to go to zero for large $N$ . | HL scaling for $x$ in full-system estimation and in local estimation. |
| $\omega_p$ | SQL scaling. No advantage       | Reduced state independent of $\omega_p$ : $I_{\omega_p}^{(0)} = 0$ .                | $\omega_p$ -dependence comes only from the initial state.             |

**Medium interaction** The medium interaction regime corresponds to  $\delta = \varepsilon = 1$  (middle row in the figure).

In this regime, we have seen that the full-system QFI reaches an HL scaling. Here we see that it is not the case at all. As it happens for  $x$ , the local QFI for  $\omega_p$  decreases with increasing  $N$ . We do not reach the HL scaling, but we even do not reach the SQL scaling, giving again hints of the difficulty of finding an efficient local estimation outside PT I.

**Strong interaction** The bottom plot corresponds to the strong regime of interaction ( $\delta = 0.001$ ) and shows a similar behavior of the local QFI as the one in the medium regime of interaction. The local QFI goes to zero extremely fast with  $N$ . We see on the plot a saturation of the QFI for  $N \sim 50$  at a value of  $10^{-16}$  which corresponds to the numerical precision.

## 7.5 General summary of the results for the qubit-based model

In Table 7.1 we summarize the results obtained with the ZZZZ model in local and full-system estimation. In Table 7.2 (resp. Table 7.3) we summarize the results obtained with the ZZXX model in full-system (resp. local) estimation.

Eventually the most important points that we should keep in mind are:

- We can reach the HL scaling for  $\omega_p$ , *starting with a separable state* if the free evolution does not commute with the interaction. This is shown to be true using perturbation theory for strong interaction and numerical results show that it can also work with middle interaction.
- Exact calculations show that we can reach the HL for  $x$  by measuring only the bus with the ZZZZ Hamiltonian, and numerical results tend to show that it is also possible with the ZZXX model.
- In general, both exact analytical results and exact numerical results suggest that the local estimation is either not possible or a difficult task very sensitive to a change of initial state. Still we can reach the HL for  $x$  by measuring only the bus with the ZZZZ Hamiltonian, and numerical results tend to show that it is also possible with the ZZXX model.
- The regime of validity of the perturbation theory plays an important role that was not discussed before. It is important to keep in mind that the results in perturbation theory may not always hold for very large  $N$ .

Table 7.2: Summary for the the ZZXX model with full-system estimation.

|            | Weak interactions                 | Medium interactions                             | Strong interactions                           |
|------------|-----------------------------------|---|---|
| $x$        | HL scaling<br>Predicted by PT I   | SQL $\rightarrow$ HL<br>PT I gives good results | SQL $\rightarrow$ HL<br>HL predicted by PT II |
| $\omega_p$ | SQL scaling<br>Predicted by PT II | HL scaling<br>PT I gives good results           | HL scaling<br>Predicted by PT I               |

The symbol " SQL  $\rightarrow$  HL " means that the QFI starts with an SQL scaling for low  $N$  and goes to an HL scaling for larger  $N$ .

Table 7.3: Summary for the the ZZXX model with local estimation

|            | Weak interactions | Medium interactions       | Strong interactions                               |
|------------|-------------------|---------------------------|---|
| $x$        | Goes to zero      | SQL and then goes to zero | SQL $\rightarrow$ HL ; Good agreement SNR in PT I |
| $\omega_p$ | Goes to zero      | Goes to zero              | SQL ?   |

## Summary Chapter 7

- **General qubit-based model:** By introducing a general Hamiltonian for the qubit-base coherent averaging model we show that a necessary condition for obtaining an HL scaling for  $\omega_p$  is to have a free evolution of the probes not commuting with the interaction.
- **ZZXX model:** We consider a specific case for which the non-commuting condition is fulfilled: The ZZXX Hamiltonian is given by 
$$\delta \left( \frac{\omega_p}{2} \sum_{i=1}^N Z^{(i)} + \frac{\omega_b}{2} Z^{(0)} \right) + \varepsilon \left( \frac{x}{2} \sum_{i=1}^N X^{(i)} X^{(0)} \right).$$
- **Methods:** Since we cannot solve the exact dynamics we rely on numerics to have exact solutions.
- **Full-system estimation of  $x$ :** Reach an HL scaling for weak, medium and strong interactions.
- **Full-system estimation of  $\omega_p$ :** Reach an HL scaling for strong and medium interactions. Clearly an advantage in comparison to standard metrology as here we start with a separable state.
- **Local estimation of  $x$ :** As predicted by PT I, we obtain an HL scaling for weak interactions, but numerical results show that for strong interactions the local QFI goes to zero.
- **Local estimation of  $\omega_p$ :** Numerical results suggest an HL scaling of the local QFI for weak interactions. For medium and strong interactions the local QFI goes to zero.



## Part III

# Extensions for metrology



# Chapter 8

## Channel extension for qubit errors: robustness of W and GHZ states<sup>1</sup>

### 8.1 Depolarizing and phase-flip channel estimation

We have discussed in the context of quantum parameter estimation theory — from a pretty formal point of view — and of quantum metrology — from a more practical point of view — how ancillas can be useful for enhancing the QFI or the channel QFI. Importantly, in the last paragraph of Sec. 4.4.2, we show how in general it is enough to consider only one ancilla with Hilbert space dimension equal to the dimension of the Hilbert space of the probe. While this is true in the optimal case, in non-ideal situations adding more ancillas may prove to be useful.

In this chapter we study the estimation of the depolarizing and the phase-flip channel, extending the channels to many ancilla qubits, and eventually taking into account the loss of some qubits.

#### 8.1.1 The depolarizing and phase-flip channels

As we already introduced the channel in the second chapter (see Sec. 2.6.3), we now review their effects on states in some more detail.

##### Depolarizing channel

The depolarizing channel  $\mathcal{E}_{\text{dep}}$  consists of a loss of purity of the qubit. It corresponds to the situation where the qubit is replaced with a probability  $p$  by the totally mixed state:

$$\mathcal{E}_{\text{dep}}(\rho) = p\frac{\mathcal{I}}{2} + (1-p)\rho, \quad (8.1)$$

---

<sup>1</sup>This chapter is based on: "Quantum channel-estimation with particle loss: GHZ versus W states", Fraïsse, J. M. E. and Braun, D. (2017), *Quantum Measurements and Quantum Metrology*, 3(1) available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. All the figures and parts of the discussion are reproduced from there.

where  $p \in [0, 1]$  is the depolarizing (or depolarization) strength. The Kraus decomposition for the depolarizing channel acting on a qubit is

$$\mathcal{E}_{\text{dep}}(\rho) = \sum_{i=1}^4 E_i \rho E_i^\dagger, \quad (8.2)$$

with the four Kraus operators

$$E_1 = \sqrt{1 - \frac{3}{4}p} \mathcal{I}, \quad E_2 = \sqrt{\frac{p}{4}} X, \quad E_3 = \sqrt{\frac{p}{4}} Y, \quad E_4 = \sqrt{\frac{p}{4}} Z. \quad (8.3)$$

Often the depolarizing channel is parametrized as

$$\mathcal{E}_{\text{dep}}(\rho) = (1 - q)\rho + \frac{q}{3}X\rho X + \frac{q}{3}Y\rho Y + \frac{q}{3}Z\rho Z,$$

where  $q \in [0, 1]$  [Nielsen and Chuang, 2011]. Using this parametrization the depolarizing channel has another interpretation: The state  $\rho$  is conserved with a probability  $1 - q$  and each basic error is applied with a probability  $q/3$ . In this chapter we will use the first parametrization.

### Phase-flip channel

The phase-flip channel corresponds to the case where with probability  $p$  the qubit has its phase flipped. The decomposition in Kraus operators reads

$$\mathcal{E}_{\text{ph}}(\rho) = \sum_{i=1}^2 F_i \rho F_i^\dagger, \quad (8.4)$$

with the Kraus operators:

$$F_1 = \sqrt{1 - p} \mathcal{I}; \quad F_2 = \sqrt{p} Z. \quad (8.5)$$

As a result starting with a state  $\rho := (\mathcal{I} + r_x X + r_y Y + r_z Z)/2$  we obtain the state  $\mathcal{E}_{\text{ph}}(\rho) = (\mathcal{I} + (1 - 2p)r_x X + (1 - 2p)r_y Y + r_z Z)/2$ . In terms of Bloch sphere this channel consists of a contraction of the sphere in the  $xy$ -plane. If we start from a pure state  $|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\varphi} \sin(\theta/2)|1\rangle$  then the state after evolution is  $\mathcal{E}_{\text{ph}}(|\psi\rangle\langle\psi|) = (1 - p)|\psi\rangle\langle\psi| + p|\psi_{\text{ph}}\rangle\langle\psi_{\text{ph}}|$ , where  $|\psi_{\text{ph}}\rangle = \cos(\theta/2)|0\rangle + e^{i(\varphi+\pi)} \sin(\theta/2)|1\rangle$ . From there the name phase-flip, as the resulting state is a mixture between the original state and the state with its phase flipped (from  $\varphi$  to  $\varphi + \pi$ ).

### 8.1.2 Optimal estimation of a Pauli channel

Pauli channels are the unital channels constructed by using the Pauli matrices as Kraus operators:

$$\mathcal{E}(\rho) = \sum_{i=0}^3 d_i \sigma_i \rho \sigma_i,$$

with  $\sum_{i=0}^3 d_i = 1$ . Fujiwara and Imai [2003], using the tools of geometry of information developed by Amari and Nagaoka in the 80's in Japan (see [Amari and Nagaoka, 2007]), investigated the problem of estimating generalized Pauli channels. These channels act on  $d$ -dimensional systems — also known as qudits — and are built using a generalization of the Pauli matrices to higher dimensions as Kraus operators. The estimation of generalized Pauli channels is in general a

multi-parameter estimation problem, where one tries to estimate the set of  $d^2 - 1$  parameters  $\{d_i\}_{1 \leq i \leq d^2 - 1}$ .

The authors were interested in the optimal protocol when one can use  $m$  times the channel and when ancillas are available. In our discussion on quantum metrology we have seen how entanglement can play two different roles in view of enhancing the sensitivity: Either by entangling the probes between them, or by entangling a probe and an ancilla. Here they consider the possibility to use both kinds of enhancement.

It turns out that the optimal protocol (in terms of QFI matrix) consists in making  $m$  independent estimations of the channel extended to another ancilla qudit and to use as input a maximally entangled state:  $(\mathcal{E} \otimes \mathcal{I})^{\otimes m}(\rho_{\max. \text{ ent}}^{\otimes m})$ . The maximally entangled state is defined as

$$\rho_{\max. \text{ ent}} := \frac{1}{\sqrt{d}} \sum_i^d |e_i\rangle \otimes |t_i\rangle \quad (8.6)$$

where  $\{|e_i\rangle\}$  and  $\{|t_i\rangle\}$  are orthonormal bases for the probe and the ancilla. While this shows that entanglement between probes is not useful it does not directly imply that entanglement between probes and ancilla is necessary or even useful. The authors showed that the channel QFI of the extended channel is saturated by a maximally entangled state but it may happen that a separable state performs as well.

This result holds for any dimension and channel. In the case of the qubit,  $d = 2$ , and the depolarizing channel is obtained by taking  $d_1 = d_2 = d_3 = p/4$  while the phase-flip channel is obtained by taking  $d_1 = d_2 = 0$  and  $d_3 = p$ . Then the optimal strategy for estimating the depolarizing and the phase-flip channels is to use a maximally entangled state (*e.g.* a Bell state) on the channel extended to an ancilla qubit.

### 8.1.3 Non-optimal estimation of the depolarizing channel

Knowing the optimal case is useful as it provides a benchmark for further comparison. Still what we are interested in is to consider a non-ideal situation where systems can be lost. A different way of considering non-ideal situations is to look at the QFI using sub-optimal initial states. For example one can consider mixed initial states as we know that mixed states are never the ones that maximize the QFI. We can also look at some pure states that are not maximally entangled, or at some sub-optimal extensions (acting with a non-unitary channel on the ancilla).

**Pure states and sub-optimal cases for the depolarizing channel** Frey et al. [2011] studied the estimation of a depolarizing channel  $\mathcal{E}_{\text{dep},q}(p)$  acting on qudits, where  $p$  is the probability that the state is replaced by a completely mixed state<sup>2</sup>.

The authors studied and compared different protocols:

- (a) non-extended original channel  $\mathcal{E}_{\text{dep},q}(p)$  (QFI per channel use:  $I_{\text{sim}}$ ).
- (b) channel extended to the identity by a second qudit:  $\mathcal{E}_{\text{dep},q}(p) \otimes \mathcal{I}_q$  (QFI per channel use:  $I_{\text{ext.id}}$ ).
- (c) channel applied in parallel to two qudits:  $\mathcal{E}_{\text{dep},q}(p) \otimes \mathcal{E}_{\text{dep},q}(p)$  (QFI per channel use:  $I_{\text{dbl}}$ ).
- (d) channel extended to a depolarizing channel with known depolarizing strength equal to  $\eta$ :  $\mathcal{E}_{\text{dep},q}(p) \otimes \mathcal{E}_{\text{dep},q}(\eta)$  (QFI per channel use:  $I_{\text{dep},\eta}$ ).

---

<sup>2</sup>Notice that Frey et al. [2011] study the estimation of  $\theta = (1 - p)$  in their paper.

(e) the  $m$  times iterated use of the protocols (a), (b) and (c):

(e-i)  $\mathcal{E}_{\text{dep},q}^{\circ m}(p)$  (QFI per channel use:  $I_{\text{sim},m}$ ).

(e-ii)  $\mathcal{E}_{\text{dep},q}^{\circ m}(p) \otimes \mathcal{I}_q$  (QFI per channel use:  $I_{\text{ext.id},m}$ ).

(e-iii)  $\mathcal{E}_{\text{dep},q}^{\circ m}(p) \otimes \mathcal{E}_{\text{dep},q}^{\circ m}(p)$  (QFI per channel use:  $I_{\text{dbl},m}$ ).

In all the protocols the input state are pure state. In the protocols (c),(d),(e-ii) and (e-iii) the state is a maximally entangled state. In the protocol (b) two kinds of states are considered, maximally entangled state and partially entangled states. The figure of merit considered for the comparison between the protocols is the *QFI per channel invocation*.

In terms of results we already know from Fujiwara's work that the best estimation scenario is to consider the extended channel fed with a maximally entangled state (protocol (b)). It also turns out that the re-circulation of the probes is useless in the sense that  $I_{\text{sim},m} \leq I_{\text{sim}}$ ,  $I_{\text{ext.id},m} \leq I_{\text{ext.id}}$  and  $I_{\text{dbl},m} \leq I_{\text{dbl}}$ . The comparison between  $I_{\text{sim}}$  and  $I_{\text{dbl}}$  gives a result depending on  $d$  and  $p$ , meaning that depending on the dimension and on the depolarizing strength one protocol is better than the other.

They also emphasized that depending on the strength  $\eta$  of the additional depolarizing channel it may be better to use the protocol (c) or (a). The study of this additional channel is motivated by the fact that it may look unrealistic to extend the channel to the identity exactly. But the analysis is done assuming that the value of  $\eta$  is perfectly known, which may also look unrealistic. In this case one could think of making a multi-parametric investigation. From this they conclude that "*realizing the theoretical advantage of entanglement depends on precise control of the experimental apparatus*".

**Use of initial mixed states with the depolarizing channel** Collins and Stephens [2015] considered initial mixed states for the estimation of the depolarizing channel for a qubit. They investigated what is the best protocol given a collection of  $m$  qubits with purity  $r$ . As a benchmark they used the QFI for a single use of the channel and without ancillas. Then they looked at a pure sequential protocol where the channel is applied  $m$  times to the same qubit. They also investigated a more sophisticated protocol in which a joint state of  $n$  qubits is prepared and followed by the application of the depolarizing channel to the first  $m$  qubits. This is thus a correlated protocol with both probe-probe and probe-ancilla correlations. The figure of merit was again the QFI per channel invocation and they defined the gain of a protocol as the ratio of its QFI per channel invocation over the QFI obtain with the single use of the channel.

For the sequential protocol the authors were able to identify a gain for certain depolarization strengths, when the initial state is mixed. When it is pure the sequential protocol never leads to any gain.

More interesting for our discussion is the result with the correlated protocol. Due to the complexity of the formula for the QFI the authors relied on a numerical analysis. They found that for  $m = 1$ , meaning that the channel is applied only to the first qubit, the case with  $n = 5$  can lead to a better QFI than for  $n = 2$ . By considering initial qubits with very low purity ( $r \ll 1$ ) they could show analytically that the gain with the correlated protocol is equal to  $mn(1-p)^{2m-2}$ . For  $m = 1$  there is an improvement by a factor  $n$  of the QFI. This clearly demonstrates that for non-ideal situations using many ancillas can prove useful.

To conclude we should still say a word about the state used in the correlated protocol. The state preparation consists in applying a control-Z gate to each but the first qubit, using the first qubit as control. Then a Hadamard gate is applied to each qubit. Interestingly it turns out that

the state after preparation is separable for low values of purity. Separable but discordant, giving again hints on the role of discord in quantum metrology with mixed states.

## 8.2 Estimation of the ideal quantum channels

Because we need a benchmark for further comparisons we must look first at the estimation with ideal quantum channels. We will do this with channels extended to  $n$  ancillas fed with either the GHZ or the W state.

### 8.2.1 Extension of the channels

We start by defining the channel extension to  $n$  ancillas of the depolarizing and the phase-flip channel.

#### Extension with $n$ ancillas of the depolarizing channel

The depolarizing channel extended to  $n$  ancilla qubits  $\mathcal{E}_{\text{dep}}^{(n)}$  can be written with the help of new Kraus operators as:

$$\mathcal{E}_{\text{dep}}^{(n)}(\rho) = (\mathcal{E}_{\text{dep}} \otimes \cdots \otimes \mathcal{I})(\rho) = \sum_{i=1}^4 \Gamma_i \rho \Gamma_i^\dagger, \quad (8.7)$$

the  $\Gamma_i$  being the Kraus operators of the new channel:

$$\Gamma_1 = \sqrt{1 - \frac{3}{4}p} \mathcal{I} \otimes \mathcal{I}^{\otimes n}, \quad \Gamma_2 = \sqrt{\frac{p}{4}} X \otimes \mathcal{I}^{\otimes n}, \quad \Gamma_3 = \sqrt{\frac{p}{4}} Y \otimes \mathcal{I}^{\otimes n}, \quad \Gamma_4 = \sqrt{\frac{p}{4}} Z \otimes \mathcal{I}^{\otimes n}. \quad (8.8)$$

#### Extension with $n$ ancillas of the phase-flip channel

In the same way as for the depolarizing channel, we extend the phase-flip channel to  $n$  ancilla qubits:

$$\mathcal{E}_{\text{ph}}^{(n)}(\rho) = (\mathcal{E}_{\text{ph}} \otimes \cdots \otimes \mathcal{I})(\rho) = \sum_{i=1}^2 \Lambda_i \rho \Lambda_i^\dagger, \quad (8.9)$$

where the new Kraus operators are defined as

$$\Lambda_1 = \sqrt{1 - p} \mathcal{I} \otimes \mathcal{I}^{\otimes n}, \quad \Lambda_2 = \sqrt{p} Z \otimes \mathcal{I}^{\otimes n}. \quad (8.10)$$

### 8.2.2 States considered

For feeding our extended channels we consider two kinds of entangled states, GHZ states and W states. We have already encountered GHZ states in Chapter 4 when discussing the effect of noise. These states are the optimal states in the framework of Ramsey interferometry and are the equivalent of the ON states used in quantum optics.

### GHZ states

Formally, the GHZ state — named after Greenberger, Horne, and Zeilinger — for  $n + 1$  qubits is defined as

$$|\psi^{\text{GHZ-}n}\rangle = \frac{1}{\sqrt{2}}(|0, 0_n\rangle + |1, 1_n\rangle), \quad (8.11)$$

with  $|0, 0_n\rangle = |0\rangle_1 \otimes |0\rangle_2 \otimes \cdots \otimes |0\rangle_{n+1}$ ,  $|0, 1_n\rangle = |0\rangle_1 \otimes |1\rangle_2 \otimes \cdots \otimes |1\rangle_{n+1}$ ,  $|1, 0_n\rangle = |1\rangle_1 \otimes |0\rangle_2 \otimes \cdots \otimes |0\rangle_{n+1}$  and  $|1, 1_n\rangle = |1\rangle_1 \otimes |1\rangle_2 \otimes \cdots \otimes |1\rangle_{n+1}$ . Here and in the following, the first Hilbert space is the one for the probe and all the others are for ancillas. For our notation to be consistent we consider  $n \geq 1$ . When there is only one ancilla,  $n = 1$ , the GHZ state  $|\psi^{\text{GHZ-}1}\rangle$  is equal to the Bell state  $|\phi_+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ . GHZ states are very prone to decoherence, in the sense that if even a single qubit is lost (traced out), we end up with a mixed non-entangled state (see Eq. (8.39) below). We define the density matrix  $\rho^{\text{GHZ-}n} = |\psi^{\text{GHZ-}n}\rangle\langle\psi^{\text{GHZ-}n}|$ .

### W states

The W state for  $n + 1$  qubits, W- $n$  for short, is defined as

$$|\psi^{\text{W-}n}\rangle = \frac{1}{\sqrt{n+1}} \sum_{i=1}^{n+1} |1_i\rangle, \quad (8.12)$$

with  $|1_i\rangle = |0\rangle_1 \otimes \cdots \otimes |0\rangle_{i-1} \otimes |1\rangle_i \otimes |0\rangle_{i+1} \otimes \cdots \otimes |0\rangle_{n+1}$ ,  $\forall i \in \{1, \dots, n+1\}$ , *i.e.* it corresponds to a single excitation distributed evenly over all qubits. The case  $n = 1$  gives also a Bell state:  $|\psi^{\text{W-}1}\rangle = |\varphi_+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ . We also define  $\rho^{\text{W-}n} = |\psi^{\text{W-}n}\rangle\langle\psi^{\text{W-}n}|$ . Contrarily to the GHZ states, W states are more resilient against the loss of a qubit in terms of entanglement. Typically if we trace out a qubit from a W state we end up with a state still entangled. As a matter of fact W states and GHZ states belong to two different classes of entangled states<sup>3</sup>, meaning that we cannot transform one into the other by acting with local operations on each qubit.

## 8.2.3 Ideal estimation of depolarizing channels

### Optimal protocol and non-extended channels

In line with the results from Fujiwara the optimal protocol corresponds to using the channel extended to one qubit and to feed it with a Bell state. The calculation is straightforward and leads to

$$I_{\text{dep}}^{\text{opt}} = \frac{3}{p(4-3p)}. \quad (8.13)$$

To check to what extent this constitutes an improvement we look at the QFI obtained without using ancillas. We refer to this case as the non-extended case. Interestingly the QFI for the depolarizing channel does not depend on the specific form of the state but only on its purity. Starting with a pure state we obtain

$$I_{\text{dep}}^{\text{n.e.}} = \frac{1}{p(2-p)}. \quad (8.14)$$

where the superscript "n.e." stands for "non-extended".

---

<sup>3</sup>This is true only for states with at least three qubits. Indeed for two qubits states, W and GHZ states are maximally entangled states.

### Using GHZ states

Starting with a GHZ state and applying the depolarizing channel we obtain

$$\rho_{\text{dep}}^{\text{GHZ-}n} \equiv \mathcal{E}_{\text{dep}}^{(n)}(\rho^{\text{GHZ-}n}) \quad (8.15)$$

$$\begin{aligned} &= \frac{2-p}{4} (|0, 0_n\rangle\langle 0, 0_n| + |1, 1_n\rangle\langle 1, 1_n|) + \frac{1-p}{2} (|1, 1_n\rangle\langle 0, 0_n| \\ &\quad + |0, 0_n\rangle\langle 1, 1_n|) + \frac{p}{4} (|1, 0_n\rangle\langle 1, 0_n| + |0, 1_n\rangle\langle 0, 1_n|) . \end{aligned} \quad (8.16)$$

In view of calculating the QFI we can calculate the eigenvalues  $\sigma_i^{\text{dep}}$  of this state:

$$\sigma_1^{\text{dep}} = \frac{p}{4} \quad , \quad \sigma_2^{\text{dep}} = \frac{p}{4} \quad , \quad \sigma_3^{\text{dep}} = 1 - \frac{3p}{4} \quad , \quad \sigma_4^{\text{dep}} = \frac{p}{4} \quad , \quad (8.17)$$

as well as its eigenvectors  $|s_i^{\text{dep}}\rangle$ :

$$|s_1^{\text{dep}}\rangle = |0, 1_n\rangle \quad , \quad |s_3^{\text{dep}}\rangle = \frac{1}{\sqrt{2}} (|0, 0_n\rangle + |1, 1_n\rangle) \quad (8.18)$$

$$|s_2^{\text{dep}}\rangle = |1, 0_n\rangle \quad , \quad |s_4^{\text{dep}}\rangle = \frac{1}{\sqrt{2}} (|0, 0_n\rangle - |1, 1_n\rangle) . \quad (8.19)$$

The eigenvectors of  $\rho_{\text{dep}}^{\text{GHZ-}n}$  are independent of  $p$  which shows that all the contributions to the QFI are due to the classical mixing resulting from the application of the channel. Using Eq. (3.46) we obtain the QFI

$$I_{\text{dep}}^{\text{GHZ-}n} = \frac{3}{p(4-3p)} . \quad (8.20)$$

Importantly this QFI does not depend on the number of ancillas and is also equal to the optimal QFI:  $I_{\text{dep}}^{\text{GHZ-}n} = I_{\text{dep}}^{\text{opt}}$ . Since the GHZ state with  $n = 1$  is an optimal state it shows that all the GHZ states lead to the same QFI.

### W states

Starting with an initial W state with  $n$  ancillas the state after evolution through the depolarizing channel is

$$\rho_{\text{dep}}^{\text{W-}n} := \mathcal{E}_{\text{dep}}^{(n)}(\rho^{\text{W-}n}) \quad (8.21)$$

$$\begin{aligned} &= \frac{p}{2(n+1)} \left( |0, 0_n\rangle\langle 0, 0_n| + \sum_{i=2}^{n+1} |1, 1_i\rangle\langle 1, 1_i| + \sum_{\substack{i,j=2 \\ i \neq j}}^{n+1} |1, 1_i\rangle\langle 1, 1_j| \right) \\ &\quad + \frac{2-p}{2(n+1)} \left( \sum_{i=1}^{n+1} |1_i\rangle\langle 1_i| + \sum_{\substack{i,j=2 \\ i \neq j}}^{n+1} |1_i\rangle\langle 1_j| \right) + \frac{1-p}{n+1} \sum_{i=2}^{n+1} (|1_i\rangle\langle 1_1| + |1_1\rangle\langle 1_i|) , \end{aligned} \quad (8.22)$$

with  $|1, 1_i\rangle = |1\rangle_1 \otimes |0\rangle_2 \otimes \cdots \otimes |0\rangle_{i-1} \otimes |1\rangle_i \otimes |0\rangle_{i+1} \otimes \cdots \otimes |0\rangle_{n+1}$ ,  $\forall i \in \{2, \dots, n+1\}$ .

In the basis that we are using here — the computational basis — the matrix representation of  $\rho_{\text{dep}}^{\text{W-}n}$  is block diagonal (direct sum structure of the operator). More specifically it is composed of

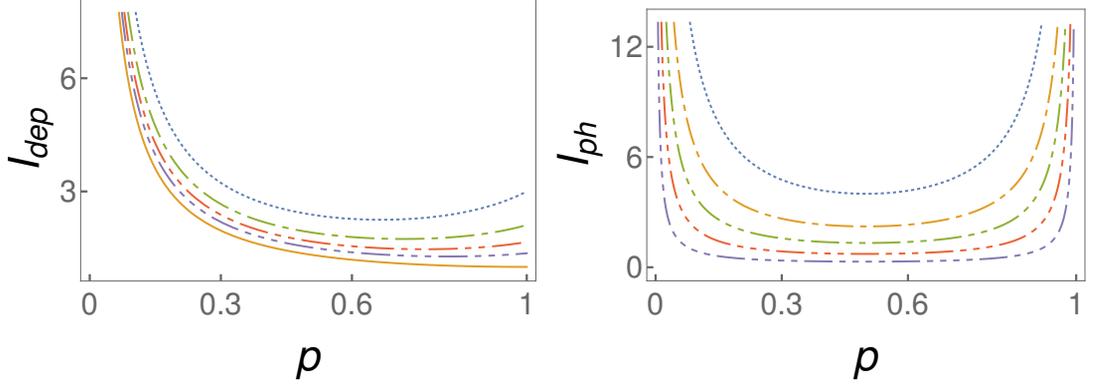


Figure 8.1: QFI with no loss of qubits. Left figure (depolarizing channel): dotted line: GHZ (optimal strategy); 1-dash line: W-5; 2-dash line: W-10; 3-dash line: W-20; full line: separable state. Right figure (phase-flip channel): dotted line: GHZ (optimal separable protocol); 1-dash line: W-5; 2-dash line: W-10; 3-dash line: W-20; 4-dash line: W-50.

- a first trivial  $1 \times 1$  block composed of the eigenvalue  $\frac{p}{2(n+1)}$ .
- a second block  $G^{(n)}(a)$  with  $a = \frac{p}{2(n+1)}$ .
- a third block  $K^{(n+1)}(a, b, a)$  with  $a = \frac{2-p}{2(n+1)}$  and  $b = \frac{1-p}{n+1}$ .

The matrices  $G^{(n)}(x)$  and  $K^{(n+1)}(x, y, z)$  are given in Appendix D as well as their eigenvalues and eigenvectors. This allows us to calculate the QFI which reads

$$I_{\text{dep}}^{W-n} = \frac{1}{p(2-p)} \frac{(3p-4(1+n(n+4)))/(1+n)^2}{(3p-4)}. \quad (8.23)$$

Even if this analysis is restricted to  $n \geq 1$ , Eq. (8.23) for  $n = 0$  gives the correct QFI.

Contrary to the QFI for the GHZ state, the QFI for the W state *does* depend on the number of ancillas added:  $I_{\text{dep}}^{W-n}$  decreases as a function of  $n$ . The more ancillas we add, the lower the QFI is.

In the left plot in Fig. 8.1 we represented the QFI for the depolarizing channel for various states as a function of  $p$ . The dotted line corresponds to the optimal state and to the GHZ states. The plain line corresponds to the non-extended protocol and serves as a lower bound. The 1-, 2-, and 3-dashed lines correspond to W states with five, ten and twenty ancillas. When we go to an infinite number of ancillas in the W state the QFI is equal to

$$I_{\text{dep}}^{W-n} \xrightarrow{n \rightarrow \infty} \frac{1}{p(2-p)} = I_{\text{dep}}^{\text{n.e.}}, \quad (8.24)$$

which is nothing but the QFI for the non-extended protocol.

### 8.2.4 Phase-flip channel

#### Optimal protocol and non-extended channel

Starting with a maximally mixed state of two qubits we obtain

$$I_{\text{ph}}^{\text{opt}} = \frac{1}{p(1-p)}, \quad (8.25)$$

which is thus the optimal QFI achievable using the phase-flip channel. Contrarily to the depolarizing channel here in the non-extended case the QFI depends strongly on the initial state. For example the states  $|0\rangle$  or  $|1\rangle$  are stationary states for this channel and therefore nothing can be inferred when using them. Intuitively we may expect that it is the state that lies in the  $xy$ -plane that is the most sensitive to a change of its phase, and therefore that leads to the highest QFI. Introducing  $|\psi_{xy}\rangle = (|0\rangle + e^{-i\varphi}|1\rangle)/\sqrt{2}$  we obtain the QFI for these states:

$$I_{\text{ph}}^{\text{n.e.}} = \frac{1}{p(1-p)}, \quad (8.26)$$

which is equal to the optimal QFI. This shows that the states in the  $xy$ -plane are indeed the optimal states in the non-extended protocol, but it also shows that for phase-flip channels using ancillas is not required to reach the maximal QFI. This is a known result. In [Fujiwara and Imai, 2003] the authors emphasized that when estimating the parameter of a Pauli channel lying on the boundaries of the tetrahedron of the simplex representing the different Pauli channels, non-maximally entangled states may be as efficient as maximally entangled ones.

Since ancillas are not useful here there is no physical interest in studying the loss of ancillas as the best strategy is to not use any ancillas anyway. We will still carry out the calculations with the phase-flip channel for the mathematical interest and to compare it with the depolarizing channel.

#### GHZ states

Consider  $\rho_{\text{ph}}^{\text{GHZ-}n} = \mathcal{E}_{\text{ph}}^{(n)}(\rho^{\text{GHZ-}n})$ . Applying the Kraus operators, one obtains immediately

$$\rho_{\text{ph}}^{\text{GHZ-}n} = \frac{1}{2} (|0, 0_n\rangle\langle 0, 0_n| + |1, 1_n\rangle\langle 1, 1_n|) + \frac{1-2p}{2} (|1, 1_n\rangle\langle 0, 0_n| + |0, 0_n\rangle\langle 1, 1_n|). \quad (8.27)$$

This state has rank two, so we can easily diagonalize it. Its eigenvalues are  $\sigma_1^{\text{ph}} = p$  and  $\sigma_2^{\text{ph}} = 1-p$ . Its eigenvectors reads

$$|s_1^{\text{ph}}\rangle = \frac{1}{\sqrt{2}} (|0, 0_n\rangle - |1, 1_n\rangle), \quad |s_2^{\text{ph}}\rangle = \frac{1}{\sqrt{2}} (|0, 0_n\rangle + |1, 1_n\rangle).$$

Once more the eigenvectors are independent of  $p$ , the QFI reduces to its classical component and is given by

$$I_{\text{ph}}^{\text{GHZ-}n} = \frac{1}{p(1-p)} = I_{\text{ph}}^{\text{opt}}. \quad (8.28)$$

As with the depolarizing channel, the QFI for GHZ states does not depend on the number of ancillas — the case with zero ancillas notwithstanding — and therefore is always optimal.

### W states

The state after acting with the phase-flip channel on the W states,  $\rho_{\text{ph}}^{\text{W-}n} := \mathcal{E}_{\text{ph}}^{(n)}(\rho^{\text{W-}n})$ , is given by

$$\rho_{\text{ph}}^{\text{W-}n} = \frac{1}{n+1} \left( \sum_{i=1}^{n+1} |1_i\rangle\langle 1_i| + \sum_{\substack{i,j=2 \\ i \neq j}}^{n+1} |1_i\rangle\langle 1_j| \right) + \frac{1-2p}{n+1} \left( \sum_{i=2}^{n+1} (|1_1\rangle\langle 1_i| + |1_i\rangle\langle 1_1|) \right). \quad (8.29)$$

The matrix representation of this state in the computational basis has a block structure with a single non-zero block of the form  $K^{(n+1)}(a, b, a)$ , where  $a = \frac{1}{n+1}$  and  $b = \frac{1-2p}{n+1}$ .

Using Eq. (D.2) and the normalized version of Eq. (D.3) both given in the appendix D, we can compute the QFI,

$$I_{\text{ph}}^{\text{W-}n} = \frac{4n}{(1+n)^2} \frac{1}{p(1-p)} = f(n) I_{\text{ph}}^{\text{opt}}, \quad (8.30)$$

where  $f(n) = \frac{4n}{(1+n)^2}$ . This function is smaller than one for  $n$  larger than one in agreement with the fact that we cannot obtain a better QFI than the optimal one. For large  $n$ , the function  $f(n)$  becomes small and the QFI goes to zero. Although we assumed  $n \geq 1$  the case  $n = 0$  is still correct as  $f(0) = 0$  and the W state with zero ancillas is equal to the state  $|1\rangle$ , which is a stationary state of the phase-flip channel.

In the right plot of Fig. 8.1 we represented the QFI for the phase-flip channel as a function of  $p$ . We observe that the QFI is symmetric in  $p$  and that W states have a performance that decreases with the number of ancillas.

## 8.3 Losing particles

In the ideal scenario adding more ancillas is at best useless — for GHZ states — and even sometimes damageable — for W states. We now turn to the non-ideal case where some qubits can be lost. We start by formalizing the idea of loss, then we look at the estimation when one ancilla is lost and we conclude with the general case where  $l \leq n$  ancillas are lost.

### 8.3.1 Modelling the loss

Consider an arbitrary extended quantum channel  $\mathcal{F}_{\text{ext}} = \mathcal{F}_P \otimes \text{Id}_A$  acting on  $\rho$  as

$$\mathcal{F}_{\text{ext}}(\rho) = \sum_k (F_k \otimes \mathcal{I}_A) \rho (F_k^\dagger \otimes \mathcal{I}_A). \quad (8.31)$$

We use subscripts  $P$  for the probe and  $A$  for the ancilla. We consider the situation where either the probe or the ancilla is lost after the evolution of the initial state through the channel  $\mathcal{F}_{\text{ext}}$ <sup>4</sup>.

---

<sup>4</sup>We have already studied a similar situation, although we did not present it as a study of noise. Indeed, this is close in spirit to how we define the local QFI (5.4) in the context of coherent averaging protocols: We let the system evolve and before measuring we trace out the probes. From our present point of view we conclude that when the local QFI is equal to the full-system QFI, then the protocol is resistant against noise affecting the probes.

### Losing the original probe

We denote the state which underwent the channel evolution  $\mathcal{F}_{\text{ext}}$  and then the loss of the probe as  $\rho_{\text{prob.}}^{\mathcal{F}_{\text{ext}}} := \text{tr}_P[\mathcal{F}_{\text{ext}}(\rho)]$  (the subscript *prob.* reminding that we *lost the probe* — usually the subscript refers to the system that remains but here we want to emphasize what is lost). By definition of the partial trace we have

$$\rho_{\text{prob.}}^{\mathcal{F}_{\text{ext}}} = \sum_{i,k} {}_P\langle i|(F_k \otimes \mathcal{I}_A) \rho (F_k^\dagger \otimes \mathcal{I}_A)|i\rangle_P. \quad (8.32)$$

Inserting two times the resolution of the identity on the form  $\mathcal{I}_P = \sum_i |i\rangle_P \langle i|$  we obtain

$$\rho_{\text{prob.}}^{\mathcal{F}_{\text{ext}}} = \sum_{\substack{i,k \\ \alpha,\beta}} {}_P\langle i|(F_k \otimes \mathcal{I}_A)|\alpha\rangle_P \langle \alpha|\rho|\beta\rangle_P \langle \beta|(F_k^\dagger \otimes \mathcal{I}_A)|i\rangle_P. \quad (8.33)$$

Using  ${}_P\langle i|(F_k \otimes \mathcal{I}_A)|i\rangle_P = {}_P\langle i|F_k|i\rangle_P \otimes \mathcal{I}_A$  and rearranging the terms we have

$$\rho_{\text{prob.}}^{\mathcal{F}_{\text{ext}}} = \sum_{\substack{i,k \\ \alpha,\beta}} {}_P\langle \beta|F_k^\dagger|i\rangle_P \langle i|F_k|\alpha\rangle_P \langle \alpha|\rho|\beta\rangle_P. \quad (8.34)$$

We can use again the resolution of the identity (with the sum over  $i$ ) plus the fact that  $\sum_k F_k^\dagger F_k = \mathcal{I}_P$  to finish the calculation:

$$\rho_{\text{prob.}}^{\mathcal{F}_{\text{ext}}} = \sum_{\alpha,\beta} {}_P\langle \beta|\alpha\rangle_P \langle \alpha|\rho|\beta\rangle_P = \sum_{\alpha} {}_P\langle \alpha|\rho|\alpha\rangle_P, \quad (8.35)$$

which is nothing but the reduced initial state of the ancilla:

$$\rho_{\text{prob.}}^{\mathcal{F}_{\text{ext}}} = \text{tr}_P[\rho]. \quad (8.36)$$

In this case there is nothing to estimate, we cannot get any information on the extended channel if the probe is lost.

### Losing the ancilla

This time we consider the case where the ancilla is lost after the application of the extended channel on the initial state:  $\rho_{\text{anc.}}^{\mathcal{F}_{\text{ext}}} = \text{tr}_A[\mathcal{F}_{\text{ext}}(\rho)]$ . In a similar way as for the loss of the probe we have

$$\begin{aligned} \rho_{\text{anc.}}^{\mathcal{F}_{\text{ext}}} &= \sum_{i,k} {}_A\langle i|(F_k \otimes \mathcal{I}_A) \rho (F_k^\dagger \otimes \mathcal{I}_A)|i\rangle_A \\ &= \sum_{\substack{i,k \\ \alpha,\beta}} {}_A\langle i|(F_k \otimes \mathcal{I}_A)|\alpha\rangle_A \langle \alpha|\rho|\beta\rangle_A \langle \beta|(F_k^\dagger \otimes \mathcal{I}_A)|i\rangle_A \\ &= \sum_{i,k} F_k {}_A\langle i|\rho|i\rangle_A F_k^\dagger, \end{aligned}$$

which is nothing else than the channel applied to the reduced state of the probe:

$$\rho_{\text{anc.}}^{\mathcal{F}_{\text{ext}}} = \mathcal{F}(\text{tr}_A[\rho]). \quad (8.37)$$

Losing the ancilla after evolution through the channel is equivalent to starting with the non-extended channel acting on the reduced state. Losing the ancilla after applying the channel, or starting with an initial state which already suffered the loss of the ancilla is equivalent. From this point of view, our study amounts in considering mixed initial states.

Here we did not specify the dimensions of the Hilbert spaces of the probe and of the ancilla. To go back to our protocol involving  $n$  ancilla qubits we can model the loss of  $l$  ancillas by considering that the original channel is  $\mathcal{F} = \mathcal{E}_{\text{dep/ph}} \otimes \text{Id} \otimes \cdots \otimes \text{Id}$  where there are  $n - l$  identity channels.

### 8.3.2 Losing one ancilla

Since the loss of the probe prevents to estimate the parameter we focus exclusively on the loss of ancillas. We start by considering the case where only one over the  $n$  ancillas is lost. Both the GHZ and the W state are invariant under permutation of any qubit and then it does not matter which ancilla exactly is lost (the application of the channel breaks the symmetry but there is still a symmetry among the ancillas).

#### Loss of one ancilla with the depolarizing channel

**GHZ states** Consider first the GHZ state. We are interested in the QFI  $I_{\text{dep},1}^{\text{GHZ}-n}$  of the state  $\rho_{\text{dep},1}^{\text{GHZ}-n} := \text{tr}_1[\mathcal{E}_{\text{dep}}^{(n)}(\rho^{\text{GHZ}-n})]$ , where the subscript "1" in the state, in the QFI, and *also in the trace* indicates that one ancilla got lost. In virtue of Eq. (8.37) we can write the state  $\rho_{\text{dep},1}^{\text{GHZ}-n}$  as

$$\rho_{\text{dep},1}^{\text{GHZ}-n} = \mathcal{E}_{\text{dep}}^{(n-1)}(\text{tr}_1[\rho^{\text{GHZ}-n}]) . \quad (8.38)$$

The GHZ state after the loss of one ancilla is equal to

$$\rho_1^{\text{GHZ}-n} := \text{tr}_1[\rho^{\text{GHZ}-n}] = (|0, \mathbb{0}_{n-1}\rangle\langle 0, \mathbb{0}_{n-1}| + |1, \mathbb{1}_{n-1}\rangle\langle 1, \mathbb{1}_{n-1}|) / 2 , \quad (8.39)$$

which is a separable state. After propagation through the extended depolarizing channel this state is equal to

$$\begin{aligned} \rho_{\text{dep},1}^{\text{GHZ}-n} &= \frac{2-p}{4} (|0, \mathbb{0}_{n-1}\rangle\langle 0, \mathbb{0}_{n-1}| + |1, \mathbb{1}_{n-1}\rangle\langle 1, \mathbb{1}_{n-1}|) \\ &\quad + \frac{p}{4} (|1, \mathbb{0}_{n-1}\rangle\langle 1, \mathbb{0}_{n-1}| + |0, \mathbb{1}_{n-1}\rangle\langle 0, \mathbb{1}_{n-1}|) . \end{aligned} \quad (8.40)$$

and the associated QFI reads

$$I_{\text{dep},1}^{\text{GHZ}-n} = \frac{1}{p(2-p)} . \quad (8.41)$$

This is exactly the same QFI than the one obtained with the non-extended channel:  $I_{\text{dep},1}^{\text{GHZ}-n} = I_{\text{dep}}^{\text{n.e.}}$ . This means that instead of starting with a pure state of a single qubit, we can also start with the mixed state (8.39) and use the extended channel.

When there is only one ancilla ( $n = 0$ ) the GHZ state is a Bell state, and if this ancilla is lost it becomes the completely mixed state of a single qubit which is the stationary state of the depolarizing channel. As a result the QFI equals zero:

$$I_{\text{dep},1}^{\text{GHZ}-1} = 0 . \quad (8.42)$$

**W states** After propagation of a W state through the extended depolarizing channel and the subsequent loss of an ancilla qubit, the state of the system

$$\begin{aligned}\rho_{\text{dep},1}^{\text{W}-n} &:= \text{tr}_1 \left[ \mathcal{E}_{\text{dep}}^{(n)}(\rho^{\text{W}-n}) \right] \\ &= \frac{1}{n+1} (|0, 0_{n-1}\rangle\langle 0, 0_{n-1}| + |1_1\rangle\langle 1_1|) + \frac{2-p}{2(n+1)} \sum_{i,j=2}^n |1_i\rangle\langle 1_j| \\ &\quad + \frac{1-p}{n+1} \sum_{i=2}^n (|1_1\rangle\langle 1_i| + |1_i\rangle\langle 1_1|) + \frac{p}{2(n+1)} \sum_{i,j=2}^n |1, 1_i\rangle\langle 1, 1_j|,\end{aligned}\quad (8.43)$$

has a block structure with three non-zero blocks:

- a first non-contributing  $1 \times 1$  block composed by the eigenvalue  $\frac{1}{n+1}$ .
- a second block  $G^{(n-1)}(a)$  with  $a = \frac{p}{2(n+1)}$ .
- a third block  $K^{(n)}(a, b, c)$  with  $a = \frac{2-p}{2(n+1)}$ ,  $b = \frac{1-p}{n+1}$  and  $c = \frac{1}{n+1}$ .

This leads to

$$I_{\text{dep},1}^{\text{W}-n} = \frac{n-1}{n+1} \frac{n(2p-3) - 9}{p(2p-3)(2n-p(n-1))}. \quad (8.44)$$

We show in Fig. 8.2 the effect of the loss of one ancilla for the estimation of the depolarizing channel. In the left plot we represent the QFI as a function of the depolarizing strength. For W states we see that W-2 with one ancilla lost leads to a lower QFI than the one obtained in the non-extended protocol for some value of  $p$ . We also see that while W-2 with no loss has a higher QFI than W-5 with no loss, when one ancilla is lost it is the opposite. This is a first hint that using many ancillas can have some beneficial effects. In the right plot we represent the QFI as a function of the number of ancillas. The GHZ state with no loss is optimal, while the GHZ state with one ancilla lost has a constant QFI equal to the one obtained with the non-extended protocol. We also see how the W states without loss have a QFI decreasing with  $n$  and more interestingly how when one ancilla is lost their QFI reaches a maximum.

### Loss of one ancilla with the phase-flip channel

For the phase-flip channel the GHZ state after evolution through the channel and subsequent loss of an ancilla is equal to  $\rho_{\text{ph},1}^{\text{GHZ}-n} := \text{tr}_1 \left[ \mathcal{E}_{\text{ph}}^{(n)}(\rho^{\text{GHZ}-n}) \right] = \mathcal{E}_{\text{ph}}^{(n-1)}(\rho_1^{\text{GHZ}-n})$ . The mixed state  $\rho_1^{\text{GHZ}-n}$  being a stationary state of  $\mathcal{E}_{\text{ph}}^{(n-1)}$ , the QFI is equal to zero for any number of ancillas used:  $I_{\text{ph},1}^{\text{GHZ}-n} = 0$ .

For the W state the state is  $\rho_{\text{ph},1}^{\text{W}-n} := \text{tr}_1 \left[ \mathcal{E}_{\text{ph}}^{(n)}(\rho^{\text{W}-n}) \right] = \mathcal{E}_{\text{ph}}^{(n-1)}(\text{tr}_1[\rho^{\text{W}-n}])$ . Direct calculation gives

$$\rho_{\text{ph},1}^{\text{W}-n} = \frac{1}{n+1} |0, 0_{n-1}\rangle\langle 0, 0_{n-1}| + \frac{n}{n+1} \rho_{\text{ph}}^{\text{W}-(n-1)}. \quad (8.45)$$

Note that  $\rho_{\text{ph},1}^{\text{W}-n}$  can be written as a direct sum. Since the first block does not depend on  $p$ , we can compute the QFI directly from the second block,

$$I_{\text{ph},1}^{\text{W}-n} = \frac{n}{n+1} I_{\text{ph}}^{\text{W}-(n-1)} = \frac{4(n-1)}{n(n+1)} \frac{1}{p(1-p)}. \quad (8.46)$$

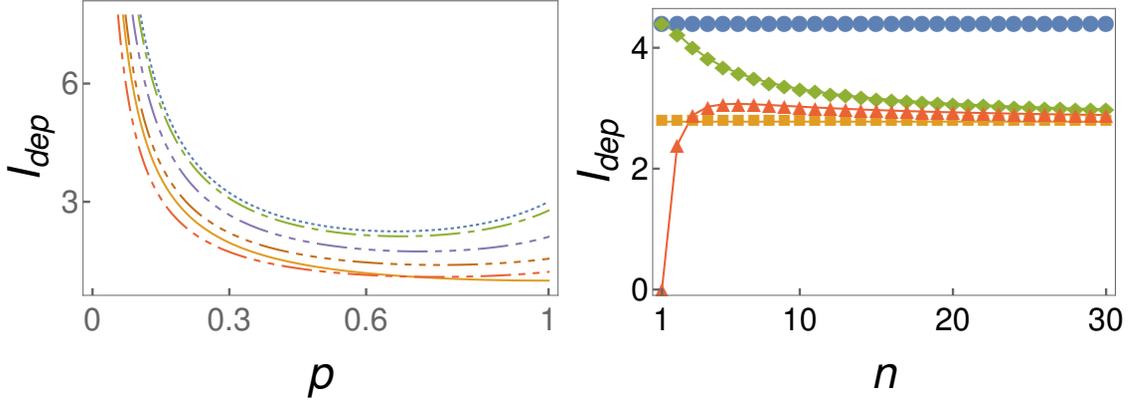


Figure 8.2: Effect of the loss of one ancilla qubit on the QFI for the depolarizing channel. Left figure: dotted line: optimal strategy / GHZ with no loss; 1-dash line: W-2 with no loss; 2-dash line: W-2 with one loss; 3-dash line: W-5 with no loss; 4-dash line: W-5 with one loss; full line: separable protocol / GHZ with one qubit lost. Right figure ( $p = 0.2$ ): full circles: GHZ with no loss; diamonds: W states with no loss; triangle up: W states with one ancilla lost; squares: separable protocol / GHZ with one ancilla lost.

We see in the left of Fig. 8.3 that this time W-2 has a larger QFI than W-5 with and without loss (for W-5 the QFI with and without loss are almost equal). In the right figure we observe the convergence of the QFI to zero for the W states in the ideal case and with one ancilla lost. We also see the maximal QFI when one ancilla is lost is reached for  $n = 2$  and  $n = 3$ . We should not conclude that adding ancillas is useful in this situation as for the phase-flip channel the optimal QFI is already reached without using any ancillas.

### 8.3.3 Generalization to the loss of $l$ ancillas

We now consider an arbitrary number  $l \leq n$  of lost ancillas.

### 8.3.4 Depolarizing channel

#### GHZ states

The GHZ state after  $l$  ancillas lost is equal to

$$\text{tr}_l[\rho^{\text{GHZ},n}] = (|0, \mathbb{0}_{n-l}\rangle\langle 0, \mathbb{0}_{n-l}| + |1, \mathbb{1}_{n-l}\rangle\langle 1, \mathbb{1}_{n-l}|)/2. \quad (8.47)$$

This state has exactly the same form than the state with only one ancilla lost. Indeed GHZ states are so sensitive to the loss of subsystems that losing one or more ancillas do not make a qualitative difference. Using the results of the last section we directly obtain for  $1 \leq l \leq n - 1$

$$I_{\text{dep},l}^{\text{GHZ}-n} = I_{\text{dep}}^{\text{sep}} = \frac{1}{p(2-p)}, \quad (8.48)$$

and for  $n = l$

$$I_{\text{dep},l}^{\text{GHZ}-l} = 0. \quad (8.49)$$

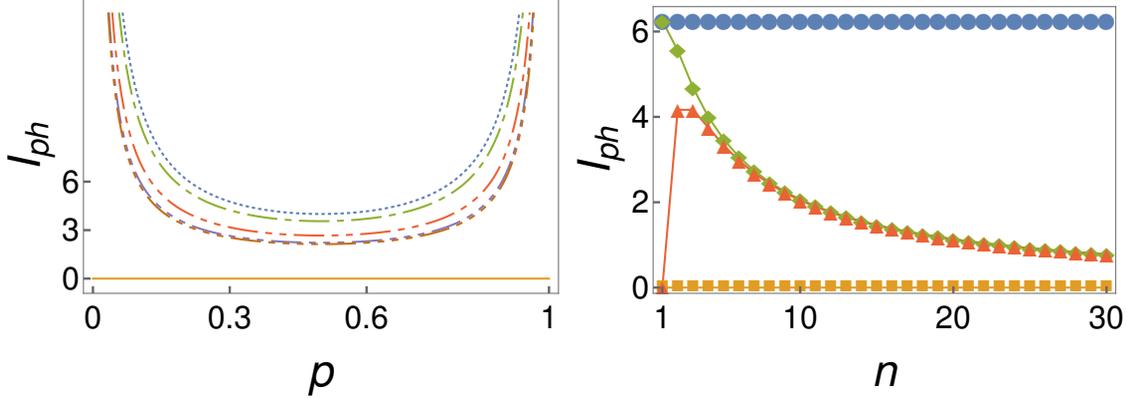


Figure 8.3: Effect of the loss of one ancilla qubit on the QFI for the phase-flip channel. Left figure: dotted line: GHZ with no loss / optimal separable protocol; 1-dash line: W-2 with no loss; 2-dash line: W-2 with one qubit lost; 3-dash line: W-5 with no loss; 4-dash line: W-5 with one qubit lost; full line: GHZ with one qubit lost. Right figure ( $p = 0.2$ ): full circles: GHZ with no loss / optimal separable protocol; diamonds: W states with no loss; triangle up: W states with one ancilla lost; squares: GHZ with one ancilla lost.

This definitely shows that GHZ states are not useful for the estimation of depolarizing channels when considering loss: Bell states are already optimal and using more than one ancilla does not help to make the state more robust to loss.

### W states

The W state after  $l$  ancillas lost is equal to

$$\rho_l^{W-n} := \text{tr}_l[\rho^{W-n}] = \frac{l}{n+1}|0, 0_{n-l}\rangle\langle 0, 0_{n-l}| + \frac{n+1-l}{n+1}\rho^{W-(n-l)}. \quad (8.50)$$

In clear contrast to the GHZ states we see that here the form of the states depends strongly on the number of lost ancillas.

After evolution through the depolarizing channel the state is equal to  $\text{tr}_l[\rho_{\text{dep}}^{W-n}] =: \rho_{\text{dep},l}^{W-n}$  and reads

$$\begin{aligned} \rho_{\text{dep},l}^{W-n} &= \frac{2l-p(l-1)}{2(n+1)}|0, 0_{n-1}\rangle\langle 0, 0_{n-1}| + \frac{2+p(l-1)}{2(n+1)}|1_1\rangle\langle 1_1| + \frac{2-p}{2(n+1)} \sum_{i,j=2}^{n+1-l} |1_i\rangle\langle 1_j| \\ &+ \frac{1-p}{n+1} \sum_{i=2}^{n+1-l} (|1_1\rangle\langle 1_i| + |1_i\rangle\langle 1_1|) + \frac{p}{2(n+1)} \sum_{i,j=2}^{n+1-l} |1, 1_i\rangle\langle 1, 1_j|. \end{aligned} \quad (8.51)$$

The representation matrix of this state has three non-zero blocks:

- a first  $1 \times 1$  block composed by the eigenvalue  $\frac{2l-p(l-1)}{2(n+1)}$ .
- a second block  $G^{(n-l)}(a)$  with  $a = \frac{p}{2(n+1)}$ .

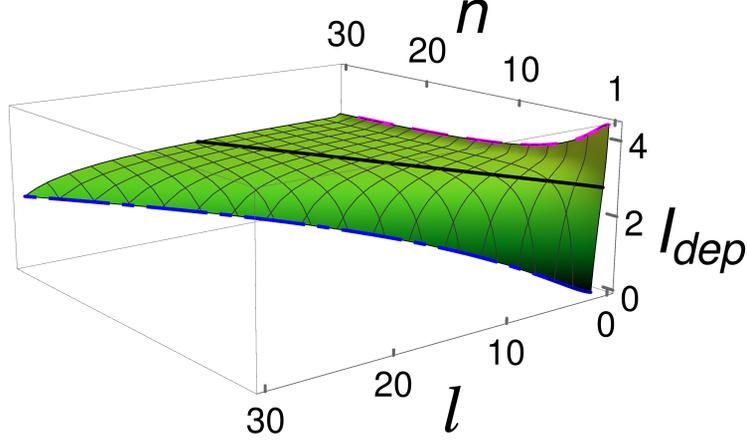


Figure 8.4: QFI for the depolarizing channel for a W state with  $n$  ancillas and  $l$  lost ones ( $p = 0.2$ ). The full line corresponds to the separable strategy or GHZ with loss, the dashed line to the case where no ancillas are lost ( $l = 0$ ) and the 1-dash line to the case where all ancillas are lost ( $l = n$ ).

- a third block  $K^{(n+1-l)}(a, b, c)$  with  $a = \frac{2-p}{2(n+1)}$ ,  $b = \frac{1-p}{n+1}$  and  $c = \frac{2+p(l-1)}{2(n+1)}$ .

Using this block decomposition we can calculate the QFI

$$\begin{aligned}
 I_{\text{dep},l}^{\text{W}-n} = & \left\{ -2p \left( -2(l(l+2) - 1)n^2 + l(l(3l+2) - 9)n \right. \right. \\
 & \left. \left. + l(l(3l+4) - 9) + 8n + 2 \right) + (l-1)(l+3)(n+1)p^2(2l-n-1) \right. \\
 & \left. + 4l(l+2)(n+3)(l-n) \right\} / \left\{ (n+1)p((l-1)p - 2l) \right. \\
 & \left. ((l+3)p - 2(l+2))(l(2-2p) + (n+1)(p-2)) \right\}. \quad (8.52)
 \end{aligned}$$

One checks that by setting  $l$  to zero or to one we recover our previous results (8.23) and (8.44), respectively.

In Fig. 8.5 we demonstrate the effect of the loss of several ancillas on the estimation of the depolarizing channel. In the left plot the QFI is represented as a function of  $p$ . The 1-, 2-, and 3-dashed lines represent the W-8 states with respectively no loss, two ancillas lost, and six ancillas lost. For six ancillas lost the QFI is always below the one obtained with the non-extended strategy. In the right plot we represent the QFI as a function of the number of ancillas lost. The dot, square and diamonds represent respectively W-15, W-20 and W-25. We see that the more initial ancillas are in the state the more resistant to loss the state is. But at the same time the more initial ancillas are in the state the lower is the initial QFI.

From a metrological point of view the interesting question is to compare the QFI for W states to the one obtained with the non-extended case. Indeed when  $I_{\text{dep},l}^{\text{W}-n}$  is higher than  $I_{\text{dep}}^{\text{n.e.}}$  it is worth

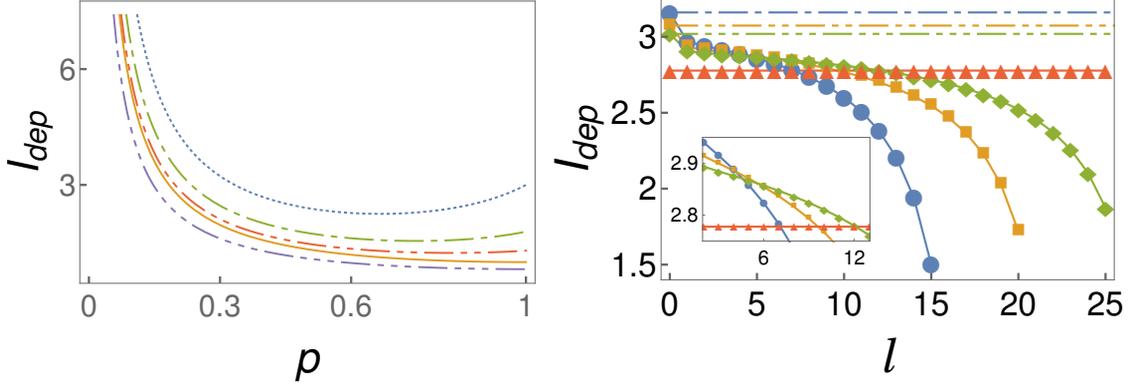


Figure 8.5: QFI for depolarizing channel for arbitrary loss. Left plot: dotted line: GHZ with no loss; 1-dash line: W-8 with no loss; 2-dash line: W-8 with 2 losses; 3-dash line: W-8 with 6 losses; full line: separable strategy / GHZ with one loss. Right plot ( $p = 0.2$ ): 1-dash line: W-15 with no loss; full circles: W-15 with loss; 2-dash line: W-20 with no loss; squares: W-20 with loss; 3-dash line: W-25 with no loss; diamonds: W-25 with loss; triangle up: separable strategy / GHZ with one loss.

to invest in ancillas. Otherwise it is better to start directly without ancillas. To compare them we can calculate the number  $l_{\text{lim}}(n)$  of lost ancillas as a function of the number of initial ancillas such that the protocol stays more efficient than the separable strategy. This function is cumbersome but actually behaves mainly linearly:  $l_{\text{lim}}(n) = 0.5n$ . This means that as long as less than half the ancillas are lost the extended protocol fairs better than the non-extended one. We illustrated this behaviour in Fig. 8.4. The QFI is plotted as a function of  $l$  and  $n$ . The black continuous line corresponds to the QFI for the non-extended protocol. The area over the black line corresponds to the situation where ancillas bring an advantage even when  $l$  of them are lost. The dashed pink line corresponds to the situation where no ancillas are lost. The area below the black line corresponds to the situation where introducing the ancillas and losing  $l$  of them deteriorates the QFI. The 1-dashed line correspond to the situation were all the ancillas are lost. Notice that even in this case the QFI does not equal zero, provided that  $n > 1$ : Setting  $l$  to  $n$  in Eq. (8.52), leads to

$$I_{\text{dep},n}^{\text{W-}n} = \frac{(n-1)^2}{(2+(n-1)p)(p+n(2-p))}, \quad (8.53)$$

which converges to  $I_{\text{dep}}^{\text{sep}}$  when  $n$  goes to infinity. For  $n = 1$ ,  $I_{\text{dep},1}^{\text{W-}1} = 0$ . We can also use the inset in the right plot of Fig. 8.5 to illustrate the behaviour of  $l_{\text{lim}}(n)$ : For W-15 we see that up to seven qubits the QFI is higher than the one for non-extended protocols. For W-20 we have  $l_{\text{lim}}(20) = 9$  and for W-25  $l_{\text{lim}}(25) = 12$ .

### 8.3.5 Phase-flip channel

For the sake of comparison we turn our attention to the phase-flip channel. As we noticed the GHZ states have the same form whatever number of ancillas is lost, and the resulting state is a stationary state of the phase-flip channel which implies that its QFI vanishes:

$$I_{\text{ph},l}^{\text{GHZ-}n} = 0. \quad (8.54)$$

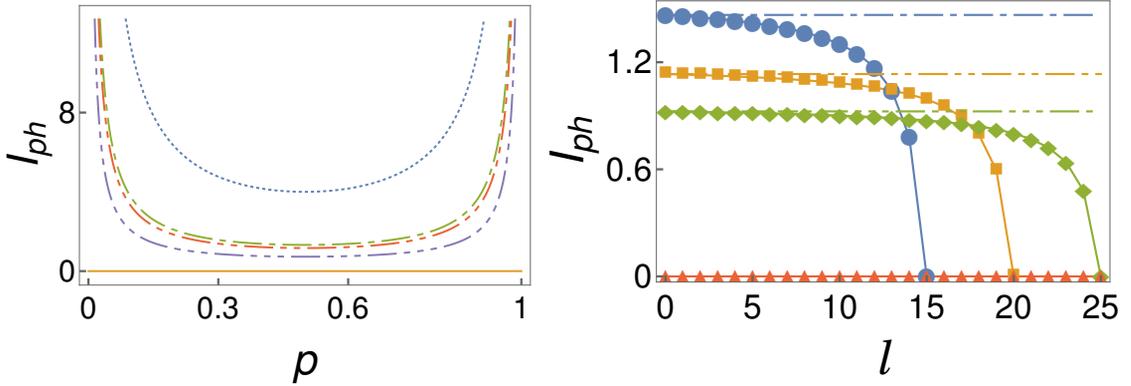


Figure 8.6: QFI for phase-flip channel for arbitrary loss. Left plot: dotted line: optimal strategy / GHZ with no loss; 1-dash line: W-10 with no loss; 2-dash line: W-10 with 6 lost; 3-dash line: W-10 with 9 losses; full line: GHZ with at least one ancilla lost. Right plot ( $p = 0.2$ ): 1-dash line: W-15 with no loss; full circles: W-15 with loss; 2-dash line: W-20 with no loss; squares: W-20 with loss; 3-dash line: W-25 with no loss; diamonds: W-25 with loss; triangle up: GHZ with loss (GHZ without loss is not represented).

Using the W state with  $l$  ancillas lost, the state after evolution through the extended channel has the form of a direct sum involving an already studied state, leading to the QFI

$$I_{\text{ph},l}^{\text{W-}n} = \frac{n+1-l}{n+1} I_{\text{ph}}^{\text{W-}(n-l)} = \frac{4(n-l)}{(n+1)(n+1-l)} \frac{1}{p(1-p)}. \quad (8.55)$$

As expected, the QFI decreases as a function of  $l$ : The more ancillas are lost the worse is the estimation. When all ancillas are lost the QFI vanishes, since the resulting state is insensitive to the phase-flip channel.

This is demonstrated in Fig. 8.6. The left plot shows the QFI as a function of  $p$ . In the right plot  $p = 0.2$ , and we plot the QFI as a function of the number of lost ancillas for W states. The more ancillas we add the smaller the initial QFI, but also the QFI decreases more slowly as a function of  $l$ . This leads to an optimal number of initial ancillas for a given number of ancillas lost, even though we have to remember that for the phase-flip channel the best strategy is to not use any ancillas at all.

### 8.3.6 Gain versus robustness

The study of the W states with  $l$  ancillas lost has shown that there are two competing behaviours regarding the number of ancillas: The more ancillas we have the higher is the QFI for a large number of ancillas lost. This is what we call robustness. But at the same time the more ancillas we have the lower is the QFI when no ancillas are lost. This was the gain (in comparison to the non-extended protocol) discussed in the ideal case. As a result, for a given number of lost ancillas there exists an optimal number of initial ancillas, as clearly demonstrated in Fig.8.7.

For the depolarizing channel, when looking at the left plot in Fig. 8.5, we see that while in the ideal case ( $l = 0$ ) W-15 is more efficient than W-25, this is already no longer true when six ancillas are lost as the inset clearly shows. We define  $n_{\text{opt,dep}}(l)$  as the optimal number of initial

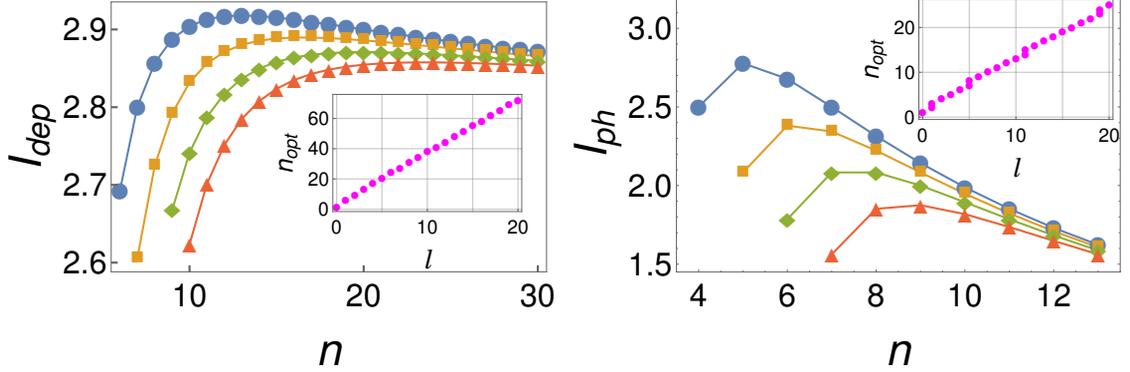


Figure 8.7: Main plots: QFI as a function of the number of initial ancillas in a W state for a fixed number of lost ancillas. The full circles correspond to three ancillas lost, the squares to four, the diamonds to five, and the triangle to six. In the insets we see the optimal number of initial ancillas in a W state as a function of the number of lost ancillas. The left plot corresponds to the depolarizing channel, the right one to the phase-flip channel ( $p = 0.2$ ).

ancillas in a W state provided that  $l$  ancillas are lost. The function  $n_{\text{opt,dep}}(l)$  has a complicated form, but its leading term is given by

$$n_{\text{opt,dep}}(l) \simeq \left(2 + \frac{2}{\sqrt{2-p}}\right) l, \quad (8.56)$$

which for  $p = 0.2$  gives roughly  $3.5l$ . We see that this is in good agreement with the inset of the left plot of Fig. 8.7.

Nevertheless, when increasing the number of ancillas in the W state we get a QFI closer to the one of the separable strategy, and thus the small gain in QFI may not justify the use of so many ancillas. As an example, when losing fifteen ancillas, the best W state is W-55 (the leading term in this case will give  $n_{\text{opt,dep}} = 52$  or  $53$ ), but its QFI equals 2.81 and the QFI for the separable strategy equals 2.77.

A similar behaviour is observed for the phase-flip channel. Although there the optimal strategy is to not add any ancilla, the study of the QFI for a fixed number of lost ancillas leads also to a maximum as represented in the right plot in Fig. 8.7. We can calculate the optimal number of initial ancillas as a function of lost ancillas in a W state,

$$n_{\text{opt,ph}}(l) = \begin{cases} \lfloor l + \sqrt{1+l} \rfloor \equiv l_f & \text{if } I_{\text{ph},l}^{\text{W-}l_f} > I_{\text{ph},l}^{\text{W-}l_c}, \\ \lceil l + \sqrt{1+l} \rceil \equiv l_c & \text{if } I_{\text{ph},l}^{\text{W-}l_f} < I_{\text{ph},l}^{\text{W-}l_c}, \\ \{l_f, l_c\} & \text{if } I_{\text{ph},l}^{\text{W-}l_f} = I_{\text{ph},l}^{\text{W-}l_c}, \end{cases} \quad (8.57)$$

with  $\lfloor \cdot \rfloor$  the floor function and  $\lceil \cdot \rceil$  the ceiling function. Thus  $n_{\text{opt,ph}}$  scales roughly linearly with  $l$ .

## Summary Chapter 8

- **Ideal estimation for Pauli channels:** For a generalized Pauli channel acting on a qudit, the optimal strategy consists to use a second qudit as an ancilla and to input a maximally entangled state. If we can use the channel  $M$  times in parallel the best strategy is an *i.i.d.* one: entanglement between probes is useless.
- **Phase-flip channel:** The non-extended channel performs as well as the extended: no need of ancillas.
- **Depolarizing channel:** An ancilla is needed to reach the maximal QFI. The optimal state is a Bell state.
- **Non-ideal estimation:** Under non-ideal situation (mixed initial state, concatenation with another non-unitary channel) using many ancillas can be useful. We investigate the situation where we start with  $n$  ancillas and lose  $l$  of them. We use the channel QFI as figure of merit.
- **GHZ states:** Become completely mixed when one particle is lost. Do not offer any advantage.
- **W states:** Stay entangled after the loss of some particles. Robust against loss for the estimation of the depolarizing channel: Even after  $l$  ancillas are lost the channel QFI can be higher than the one obtained with the non-extended channel.
- **Loss of the probe:** In full generality if the probe is lost then the QFI is equal to zero.

# Chapter 9

## Hamiltonian engineering<sup>1</sup>

### 9.1 From channel extension to Hamiltonian extension

In the last chapter we saw how the channel extension allows sometimes an enhancement in the channel QFI. This is the case for the depolarizing channel, while for the phase-flip channel ancillas do not bring any advantage in terms of channel QFI. In this chapter we will come back to the study of Hamiltonian parameters. Let us remind the main results for Hamiltonian parameter estimation (see Sec. 3.6.2): We are dealing with a Hamiltonian  $H(\theta)$  and its corresponding evolution operator  $U_{H(\theta)} = e^{-itH(\theta)}$ . The corresponding channel will be denoted  $\mathcal{U}_{H(\theta)}$  and its effect on the state  $\rho_0$  is given by  $\mathcal{U}_{H(\theta)}(\rho_0) = U_{H(\theta)}\rho_0U_{H(\theta)}^\dagger$ . Introducing the local generator

$$\mathcal{H} = iU_{H(\theta)}^\dagger\dot{U}_{H(\theta)}, \quad (9.1)$$

the QFI starting with the pure state  $|\psi_0\rangle$  is given by

$$I(\mathcal{U}_{H(\theta)}(|\psi_0\rangle); \theta) = 4 \text{Var}[\mathcal{H}, |\psi_0\rangle]. \quad (9.2)$$

In our discussion of the role of entanglement between probes and ancillas (see Sec. 4.5.3) we made clear that due to the convexity of the function  $\text{tr}[\rho\mathcal{H}^2]$  with respect to  $\rho$ , unitary channels do not benefit from channel extensions. If we want to do better, we have to come up with a different way to extend unitary channels than we did so far. The key point here is to go back to the essence of what we call an extension. Behind the concept of channel extension is the idea that the basic object, the one that we are given, is the channel  $\mathcal{E}_\theta$ . We have seen in the presentation of quantum metrology how the question of comparison is crucial, and how this question is related to the definition of resources. In our present case, it is because we consider the channel  $\mathcal{E}_\theta$  as the resource that we find fair the comparison of  $\mathcal{E}_\theta$  to  $\mathcal{E}_\theta \otimes \text{Id}$ .

Choosing the dynamics as a resource is in general justified. Still, we know that not only quantum channels are used to represent the dynamics. In the standard presentation of quantum mechanics the dynamics is attached to the Hamiltonian and the corresponding evolution operator.

---

<sup>1</sup>This chapter is based on: "Enhancing sensitivity in quantum metrology by hamiltonian extensions", Fraïsse, J. M. E. and Braun, D. (2017), *Physical Review A*, 95:062342. ©2017 Physical Review A ; and: "Hamiltonian extensions in quantum metrology", Fraïsse, J. M. E. and Braun, D. (2017), *Quantum Measurements and Quantum Metrology*, 4(1):8–16 available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. All the figures and parts of the discussion are reproduced from these two references.

In this context the resource would be the Hamiltonian  $H(\theta)$ . In a similar way as we did with channels, we can try to extend the Hamiltonian. We introduce a second system, the ancilla, and look at a more general dynamics for the total system:

$$H(\theta) \rightarrow H(\theta) \otimes \mathcal{I} + (H_{\text{int}} + \mathcal{I} \otimes H_{\text{anc}} + H_{\text{pro}} \otimes \mathcal{I}) ,$$

where  $H_{\text{int}}$  is an interaction Hamiltonian between the probe and the ancilla,  $H_{\text{anc}}$  the free Hamiltonian of the ancilla, and  $H_{\text{pro}}$  an extra free Hamiltonian for the probe. The crucial condition is that all these Hamiltonians should be  $\theta$ -independent:

$$\frac{\partial H_{\text{int}}}{\partial \theta} = \frac{\partial H_{\text{anc}}}{\partial \theta} = \frac{\partial H_{\text{pro}}}{\partial \theta} = 0 \quad \forall \theta . \quad (9.3)$$

For the sake of concision we will denote all the added terms in the Hamiltonian extension as  $H_1$  and we will not make explicit anymore the presence of the second ancilla. We formally define our Hamiltonian extension as

$$H_{\text{ext}} = H(\theta) + H_1 . \quad (9.4)$$

The corresponding evolution operator reads  $U_{H_{\text{ext}}} = e^{-itH_{\text{ext}}}$ , and  $\mathcal{H}_{\text{ext}}$  is the corresponding generator.

When studying channel extensions we showed (Eq. 3.100) how the monotonicity property of the QFI enforces the fact that the best strategy is to act on the ancilla with a unitary transformation, chosen to be the identity by convenience:  $I((\mathcal{E}_\theta \otimes \mathcal{A})(\rho); \theta) \leq I((\mathcal{E}_\theta \otimes \text{Id})(\rho); \theta)$ . For Hamiltonian extensions the situation is rather different. Notice first that if we simply added an ancilla and only considered the Hamiltonian  $H(\theta) \otimes \mathcal{I}$ , we would be in the situation of channel extensions. Indeed the corresponding evolution operator would be  $U_{H(\theta)} \otimes \mathcal{I}$  and the corresponding channel  $\mathcal{U}_{H(\theta)} \otimes \text{Id}$ , which cannot lead to an increased channel QFI.

We can see why in general we need the extra term  $H_1$  and why monotonicity does not prevent us to reach an increased channel QFI. The Zassenhaus formula states that  $e^{A+B} = e^A e^B \times e^{-[A,B]/2} \times e^{([B,[A,B]]+[A,[A,B]])/6} \times \dots$ . Using this formula we can write the evolution operator of the extended Hamiltonian as

$$U_{\text{ext}} = e^{-it(H(\theta)+H_1)} = e^{-itH(\theta)} e^{-itH_1} e^{-it[H(\theta),H_1]/2} + \dots = U_{H(\theta)} U_{H_1} U_{C_1(\theta)} U_{C_2(\theta)} \dots \quad (9.5)$$

The corresponding channel is  $\mathcal{U}_{H_{\text{ext}}} = \mathcal{U}_{H(\theta)} \circ \mathcal{U}_{H_1} \circ \mathcal{U}_{C_1(\theta)} \circ \mathcal{U}_{C_2(\theta)} \circ \dots$ . In general the  $\mathcal{U}_{C_i(\theta)}$  depends on  $\theta$  and we cannot use the results on the monotonicity. Notice that if  $H_\theta$  commutes with  $H_1$ , then all the commutators in the expansion are null and again we are left with a channel extension, and no enhancement is to be expected.

## 9.2 Upper bound to the channel QFI

We have seen in the last section how Hamiltonian extension goes beyond channel extension and how, in certain situations, Hamiltonian extension cannot lead to any improvement. This is interesting on its own right, but obviously what we are interested in is to investigate when and in which way Hamiltonian extension can lead to an enhancement for the channel QFI. In this section we will show that there exists an upper bound to the channel QFI which is in general not saturated.

### 9.2.1 Channel QFI as a semi-norm

#### Semi-norm

Our figure of merit throughout this chapter will be the channel QFI. We want to focus on the Hamiltonians and on the sensitivity that they allow. By considering the QFI we would have to also consider the initial states. With the channel QFI we can directly compare a Hamiltonian and a Hamiltonian extension.

As we only deal with Hamiltonians, we simplify the notation of the channel QFI by referring to the channel QFI of the Hamiltonian while strictly speaking we should talk of the channel QFI of the corresponding channel:  $C(\mathcal{U}_H(\theta); \theta) \equiv C(H(\theta); \theta)$ . With this notation and using the definition of the channel QFI, we have

$$C(H(\theta); \theta) = 4 \max_{\rho_0 \in \mathcal{S}(\mathcal{H})} I(\mathcal{U}_H(\theta)(\rho_0); \theta) . \quad (9.6)$$

Since it is enough to carry out the maximization over the set of pure states, using Eq. (9.2) we obtain for the channel QFI of the unitary channel

$$C(H(\theta); \theta) = 4 \max_{|\psi_0\rangle \in \mathcal{H}} \text{Var}[\mathcal{H}, |\psi_0\rangle] , \quad (9.7)$$

We have already seen when deriving the SQL for bounded Hamiltonians (see Sec. 4.2.2) how we can maximize the variance over the pure states using the Popoviciu's inequality. In the present case we obtain for the channel QFI

$$C(H(\theta); \theta) = (h_M - h_m)^2 , \quad (9.8)$$

where  $h_M$  and  $h_m$  are respectively the maximal and minimal eigenvalue of  $\mathcal{H}$ . The state for which the variance is saturated is  $(|h_M\rangle + e^{i\varphi}|h_m\rangle)/\sqrt{2}$ , where  $|h_M\rangle$  and  $|h_m\rangle$  are respectively an eigenvector with maximal associated eigenvalue and an eigenvector with minimal associated eigenvalue. Because we will often refer to these eigenvectors we call a *maximal eigenvector* (respectively *minimal eigenvector*) an eigenvector whose associated eigenvalue is maximal (respectively minimal). Both of them are referred to as *extremal eigenvectors*.

Following [Boixo et al., 2007], we introduce the semi-norm of an operator  $A$  as the difference between its extremal eigenvalues:

$$\|A\|_{sn} := a_M - a_m , \quad (9.9)$$

where  $a_M$  and  $a_m$  are the maximal and minimal eigenvalues of  $A$ . Using the semi-norm we can write the channel QFI in a simple and elegant form:

$$C(H(\theta); \theta) = \|\mathcal{H}\|_{sn}^2 . \quad (9.10)$$

#### Use and properties of the semi-norm

**Triangle inequality** The triangle inequality will play a fundamental role in the following. We thus show how we can derive it from the definition of the semi norm. Consider two operators  $B$  and  $C$ . We denote by  $|b_M\rangle$  and  $|b_m\rangle$  a maximal and a minimal eigenvector of  $B$ , with corresponding eigenvalues  $b_M$  and  $b_m$ . In the same way we introduce the extremal eigenvectors and eigenvalues of  $C$  as  $|c_M\rangle$ ,  $|c_m\rangle$ ,  $c_M$  and  $c_m$ . We are interested in the semi-norm of the sum of  $B$  and  $C$ , denoted as  $A := B + C$ . If  $|a_M\rangle$  and  $|a_m\rangle$  are a maximal and a minimal eigenvector of  $A$  and if  $a_M$  and  $a_m$  are its maximal and minimal eigenvalues we can write the semi-norm of  $A$  as

$$\|A\|_{sn} = a_M - a_m = \langle a_M | A | a_M \rangle - \langle a_m | A | a_m \rangle . \quad (9.11)$$

Using the definition of  $A$  we obtain

$$\|B + C\|_{sn} = \langle a_M | B | a_M \rangle + \langle a_M | C | a_M \rangle - \langle a_m | B | a_m \rangle - \langle a_m | C | a_m \rangle . \quad (9.12)$$

Since  $b_M$  is the maximal eigenvalue of  $B$  we have by definition that  $b_M \geq \langle a_M | B | a_M \rangle$ . The same argument with  $C$  reads  $c_M \geq \langle a_M | C | a_M \rangle$ . Since  $b_m$  is the minimal eigenvalue of  $B$  we have by definition that  $\langle a_m | B | a_m \rangle \geq b_m$  which is equivalent to  $-b_m \geq -\langle a_m | B | a_m \rangle$ . Using the same argument with  $C$ , which reads  $-c_m \geq -\langle a_m | C | a_m \rangle$ , we obtain the inequality

$$\|B + C\|_{sn} \leq b_M + c_M - b_m - c_m . \quad (9.13)$$

By definition  $\|B\|_{sn} = b_M - b_m$  and  $\|C\|_{sn} = c_M - c_m$  and we are left with the triangle inequality for the semi-norm:

$$\|B + C\|_{sn} \leq \|B\|_{sn} + \|C\|_{sn} . \quad (9.14)$$

**Saturation of the triangle inequality** The triangle equality turns into an inequality when

$$\begin{aligned} b_M &= \langle a_M | B | a_M \rangle, \\ c_M &= \langle a_M | C | a_M \rangle, \\ b_m &= \langle a_m | B | a_m \rangle, \\ c_m &= \langle a_m | C | a_m \rangle. \end{aligned}$$

The first equality implies that  $|a_M\rangle$  is a maximal eigenvector of  $B$ . In the same way the second equality implies that  $|a_M\rangle$  is also a maximal eigenvector of  $C$ . Taking in consideration that these operators can have degenerate extremal eigenvalues, the first two equalities are equivalent to the fact that they share a common maximal eigenvector. The same argument using the two last equalities shows that the second condition is that  $B$  and  $C$  should also share a common minimal eigenvector.

We call  $\mathcal{P}_B$  (respectively  $\mathcal{P}_C$ ) the invariant subspace<sup>2</sup> associated to the maximal eigenvalue of  $B$  (respectively  $C$ ). By  $\mathcal{D}_B$  (respectively  $\mathcal{D}_C$ ) we refer to the invariant subspace attached to the maximal eigenvalue of  $B$  (respectively  $C$ ). The condition of saturation of the triangle inequality is that the intersection between  $\mathcal{P}_B$  and  $\mathcal{P}_C$  as well as the intersection between  $\mathcal{D}_B$  and  $\mathcal{D}_C$  are not empty:  $\mathcal{P}_B \cap \mathcal{P}_C \neq \emptyset$  and  $\mathcal{D}_B \cap \mathcal{D}_C \neq \emptyset$ .

**Unitary invariance of the norm** Trivially, the semi-norm depending only on the eigenvalues of the operator, any operation that conserves the eigenvalues (but not necessarily the eigenvectors) conserves the semi-norm. This is especially the case with similarity transformations and among them with the unitary transformation

$$\|UAU^{-1}\|_{sn} = \|A\|_{sn} , \quad (9.15)$$

where  $U$  is any unitary operator.

---

<sup>2</sup>Say  $A$  is an operator acting on  $E$ . Then a subspace  $M \subset E$  is an invariant subspace of  $A$  if and only if  $AM \subset M$ . We also say that  $M$  is stable by  $A$ .

### Channel extensions

We can use the semi-norm to show in a few lines that unitary channels do not benefit from channel extensions in terms of channel QFI. Starting with the unitary channel  $\mathcal{U}_{H(\theta)}$ , its channel extension is given by  $\mathcal{U}_{H(\theta)} \otimes \text{Id}$ , which can also be written as  $\mathcal{U}_{H(\theta) \otimes \mathcal{I}}$ . To obtain the channel QFI we only have to look at the semi-norm of the local generator of  $H(\theta) \otimes \mathcal{I}$ . If  $\mathcal{H}$  is the local generator of  $H$  then the local generator of  $H(\theta) \otimes \mathcal{I}$  is simply  $\mathcal{H} \otimes \mathcal{I}$ . As the tensor product with the identity does not change the eigenvalues but only their multiplicity we have that  $\|\mathcal{H} \otimes \mathcal{I}\|_{sn}^2 = \|\mathcal{H}\|_{sn}^2$ , which shows that

$$C(\mathcal{U}_{H(\theta)} \otimes \text{Id}; \theta) = C(\mathcal{U}_{H(\theta)}; \theta). \quad (9.16)$$

### 9.2.2 Upper bound to the channel QFI

Equipped with this semi-norm, we can write in a simple way the upper bound of the channel QFI of any unitary channel.

**Theorem 9.1 — UPPER BOUND FOR CHANNEL QFI [BOIXO ET AL., 2007].**

*Consider a Hamiltonian  $H(\theta)$  parametrized by  $\theta$ . The channel QFI of  $\mathcal{U}_{H(\theta)}$  is upper bounded as follows:*

$$C(H(\theta); \theta) \leq t^2 \left\| \dot{H}(\theta) \right\|_{sn}^2. \quad (9.17)$$

#### Proof

The derivation of the upper bound is not difficult since we use the semi-norm and thus can apply the triangle inequality. The first step is to write the local generator in a more appealing form. To do so we should write down the derivative of the evolution operator. In general calculating the derivative of the exponential of an operator is a difficult task. The following result [Wilcox, 1967; Snider, 1964] is precious in this context: Consider a bounded operator  $M(x)$  parametrized by  $x$ . The derivative of its exponential with respect to  $x$  is given by

$$\frac{d e^{M(x)}}{dx} = \int_0^1 e^{\alpha M(x)} \frac{dM(x)}{dx} e^{(1-\alpha)M(x)} d\alpha. \quad (9.18)$$

Using this equation the local generator is written as

$$\mathcal{H} = i U_{H(\theta)}^\dagger \int_0^1 d\alpha e^{-i\alpha t H(\theta)} \left( i t \frac{\partial H(\theta)}{\partial \theta} \right) e^{-i(1-\alpha)t H(\theta)}. \quad (9.19)$$

Introducing the operators  $V(\alpha) = e^{-i\alpha t H(\theta)}$  and  $W(\alpha, \theta) = V(\alpha) \dot{H}(\theta) V(\alpha)^\dagger$ , we obtain

$$\mathcal{H} = t \int_{-1}^0 W(\alpha, \theta) d\alpha. \quad (9.20)$$

Using the triangle inequality in the form  $\left\| \int dx A(x) \right\|_{sn} \leq \int dx \|A(x)\|_{sn}$  leads to

$$\|\mathcal{H}\|_{sn} \leq t \int_{-1}^0 d\alpha \|W(\alpha, \theta)\|_{sn}. \quad (9.21)$$

In virtue of the unitary invariance of the semi-norm we have  $\|W(\alpha, \theta)\|_{sn} = \left\| \dot{H}(\theta) \right\|_{sn}$  and consequently

$$\|\mathcal{H}\|_{sn} \leq t \int_{-1}^0 \|W(\alpha, \theta)\|_{sn} d\alpha = t \left\| \dot{H}(\theta) \right\|_{sn}. \quad (9.22)$$

### 9.2.3 Condition of saturation

**Theorem 9.2 — SATURATION OF THE BOUND.**

Consider the Hamiltonian  $H(\theta)$  and its derivative  $\dot{H}(\theta) := \frac{\partial H(\theta)}{\partial \theta}$  written in its eigenbasis as

$$\dot{H}(\theta) = \sum_{i=1}^d e_i |i\rangle\langle i|, \quad (9.23)$$

with  $e_M := e_1 = \dots = e_a > e_i > e_b = \dots = e_d =: e_m$  for all  $a < i < b$ , where  $a$  is the dimension of the invariant subspace of maximal eigenvalue, noted  $\mathcal{P}$ , and  $d - b + 1$  is the dimension of the invariant subspace of minimal eigenvalue, noted  $\mathcal{D}$ .

In the case where  $\dot{H}(\theta)$  has no degenerate extremal eigenvalues, equality in Eq. (9.17) is reached if and only if the extremal eigenvalues of  $\dot{H}(\theta)$  are also eigenvectors of  $H(\theta)$ .

In the general case, equality in Eq. (9.17) is reached if and only if there exist  $|\psi\rangle \in \mathcal{P}$  and  $|\phi\rangle \in \mathcal{D}$  such that  $V(\alpha)|\psi\rangle \in \mathcal{P}$  and  $V(\alpha)|\phi\rangle \in \mathcal{D}$  for all  $\alpha \in [0, 1]$ , and where  $V(\alpha) = e^{-i\alpha t H(\theta)}$ . A sufficient condition is that there exists an eigenvector of  $H(\theta)$  in  $\mathcal{P}$  and another one in  $\mathcal{D}$ .

The inequality is saturated by  $|\psi_{\text{opt}}\rangle = (|M\rangle + |m\rangle)/\sqrt{2}$  with  $|M\rangle \in \mathcal{P}$  and  $|m\rangle \in \mathcal{D}$ , i.e. a balanced superposition of a maximal eigenvector and a minimal eigenvector of  $\dot{H}(\theta)$ .

#### Proof

We start directly by the general case, taking into account the possibility that  $\dot{H}(\theta)$  has some degenerate eigenvalues. The inequality in Eq. (9.22) is based on the triangle inequality used in its integral form:  $\|\int dx A(x)\|_{sn} \leq \int dx \|A(x)\|_{sn}$ . The condition for equality is that there exists a fixed vector belonging to the invariant subspace with maximal eigenvalue of  $A(x)$  for all values of  $x$  as well as a fixed vector belonging to the invariant subspace with minimal eigenvalue of  $A(x)$  for all values of  $x$ . In other words  $A(x)$  should have at least one maximal and one minimal eigenvector independent of  $x$ .

Coming back to our case we have to look at  $W(\alpha, \theta)$ . Working in the eigenbasis of  $\dot{H}(\theta)$  we have

$$W(\alpha, \theta) = \sum_{i=1}^d e_i |\alpha_i\rangle\langle \alpha_i| \quad (9.24)$$

with  $|\alpha_i\rangle = V(\alpha)|i\rangle$  the eigenvectors of  $W(\alpha, \theta)$ . The condition for equality is the existence of both a vector  $|\psi_M\rangle$  eigenvector of  $W(\alpha, \theta)$  with associated eigenvalue  $e_M$  for  $\alpha \in [-1, 0]$  and a vector  $|\psi_m\rangle$  eigenvector of  $W(\alpha, \theta)$  with associated eigenvalue  $e_m$  for  $\alpha \in [-1, 0]$ .

The invariant subspace with maximal eigenvalues of  $W(\alpha, \theta)$  is denoted  $\mathcal{P}_\alpha$  and equal to  $\mathcal{P}_\alpha = \text{span}\{V(\alpha)|1\rangle, \dots, V(\alpha)|a\rangle\}$ . The vector  $|\psi_M\rangle$  should be an element of  $\mathcal{P}_\alpha$  for  $\alpha \in [-1, 0]$ . This means that we can write it as

$$|\psi_M\rangle = \sum_{i=1}^a \psi_i^{(M)}(\alpha) V(\alpha)|i\rangle = V(\alpha)|\varphi(\alpha)\rangle, \quad (9.25)$$

where  $|\varphi(\alpha)\rangle := \sum_{i=1}^a \psi_i^{(M)}(\alpha) |i\rangle$  is an element of  $\mathcal{P}$ . We can thus rephrase the first condition as the existence of a fixed vector  $|\psi_M\rangle$  such that  $V(\alpha)|\psi_M\rangle$  belongs to  $\mathcal{P}$  for  $\alpha \in [0, 1]$  (where we made the transformation from  $\alpha$  to  $-\alpha$ ). In a similar way we found that the second part of the condition is that there exists a fixed vector  $|\psi_m\rangle$  such that  $V(\alpha)|\psi_m\rangle$  belongs to  $\mathcal{D}$  for  $\alpha \in [0, 1]$ . Notice that this is true especially for  $\alpha = 0$ , which shows that  $|\psi_M\rangle \in \mathcal{P}$  and  $|\psi_m\rangle \in \mathcal{D}$ .

If  $H(\theta)$  has one eigenvector in  $\mathcal{P}$ , say  $|h_{\mathcal{P}}\rangle$ , and one eigenvector in  $\mathcal{D}$ , say  $|h_{\mathcal{D}}\rangle$ , then these eigenvectors are also eigenvectors of  $V(\alpha)$  and the condition given above is fulfilled (we have  $|\psi_{\mathcal{M}}\rangle = |h_{\mathcal{P}}\rangle$  and  $|\psi_{\mathcal{m}}\rangle = |h_{\mathcal{D}}\rangle$ ). We see that in the general case we do not need to ask  $|\psi_{\mathcal{M}}\rangle$  to be an eigenvector of  $V(\alpha)$ . All what we want is that the orbit of  $|\psi_{\mathcal{M}}\rangle$  under  $V(\alpha)$  stays in  $\mathcal{P}$  for  $\alpha \in [0, 1]$  (and the similar condition for  $|\psi_{\mathcal{m}}\rangle$ ).

In the non-degenerate case we have  $\mathcal{P} = \mathbf{span}\{|1\rangle\}$  and  $\mathcal{D} = \mathbf{span}\{|d\rangle\}$ . The condition for equality is thus that  $V(\alpha)|1\rangle \propto |1\rangle$  and that  $V(\alpha)|d\rangle \propto |d\rangle$ , which means that the extremal eigenvectors of  $\dot{H}(\theta)$  should also be eigenvectors of  $V(\alpha)$ . Because it is more natural to discuss in terms of  $H(\theta)$  rather than in terms of  $V(\alpha)$  we should translate this condition. In one direction it is trivial: If a vector is an eigenvector of  $H(\theta)$  it is also an eigenvector of  $V(\alpha)$ . The other direction is more subtle. We write  $H(\theta)$  in its eigenbasis:  $H(\theta) = \sum_i h_i |h_i\rangle\langle h_i|$ . Consider two eigenvalues  $h_i$  and  $h_j$  that are non-degenerate. It is still possible that  $e^{-i\alpha h_i}$  is equal to  $e^{-i\alpha h_j}$  while  $h_i \neq h_j$ . If this is the case we are able to build a superposition of the corresponding eigenvectors which is also an eigenvector of  $V(\alpha)$  but not of  $H(\theta)$ , showing that in general  $A$  and  $e^{xA}$  do not share necessarily the same eigenvectors. But this is without taking into account the fact that we need the condition of being an eigenvector to hold for all values of  $\alpha$  in  $[0, 1]$ . And the accidental degeneracies that may cause a problem can happen only for a countable number of  $\alpha$ -values. As a consequence the condition  $e^{-i\alpha h_i} = e^{-i\alpha h_j}$  cannot be satisfied for all  $\alpha \in [-1, 0]$  if  $h_i \neq h_j$ . This shows that the condition for equality in Eq. (9.22) when  $\dot{H}(\theta)$  has no degenerate extremal eigenvalues is that the extremal eigenvectors of  $\dot{H}(\theta)$  are eigenvectors of  $H(\theta)$ .

### 9.2.4 History of the bound

In the review of quantum enhanced measurement in Chapter 4 we saw how Boixo et al. [2007] proposed to use  $k$ -body Hamiltonians to obtain a more favourable scaling of the QFI. In our presentation we just study the  $k$ -body interaction term, neglecting the interaction terms of lower order as well as the free Hamiltonian. This was not the case in the original paper. Rather they considered a Hamiltonian of the form  $\theta H_k^{\text{nl}} + \tilde{H}$ . Due to the presence of  $\tilde{H}$  it is not possible to calculate in all generality the channel QFI for this Hamiltonian. To circumvent the problem they introduced (9.17) for the specific case of the phase shift<sup>3</sup>. They also noticed that if  $\tilde{H} = 0$  then the upper bound is equal to the true value of the QFI.

Independently we derived this upper bound for a general Hamiltonian [Fraïsse and Braun, 2017b]. We were inspired by the work done on channel extension from Fujiwara mainly. Our first derivation of the bound was done without the use of the semi-norm, which makes the calculation lengthy and less elegant (need to introduce a shift in the Hamiltonian such that the extremal eigenvalues have the same absolute value but opposite signs).

## 9.3 Reaching maximal sensitivity with Hamiltonian extensions

The use of the upper bound is especially interesting in the context of Hamiltonian extensions. Indeed the upper bound involves only the first derivative of the Hamiltonian, a quantity which is by definition conserved in Hamiltonian extensions. For the original Hamiltonian and its Hamiltonian

---

<sup>3</sup>They also mention that this upper bound was present in essence but in a quite different way in [Giovannetti et al., 2006].

extension the application of theorem 9.1 reads

$$\|\mathcal{H}\|_{sn}^2 \leq t^2 \left\| \dot{H}(\theta) \right\|_{sn}^2, \quad (9.26)$$

$$\|\mathcal{H}_{\text{ext}}\|_{sn}^2 \leq t^2 \left\| \dot{H}(\theta) \right\|_{sn}^2. \quad (9.27)$$

We see that in general there is no *a priori* order between the channel QFI of the original Hamiltonian and its Hamiltonian extension. This opens the possibility to design useful Hamiltonian extensions that help to improve the sensitivity. Ultimately we would be interested in Hamiltonian extensions that saturate the bound, providing the best sensitivity achievable using the Hamiltonian  $H(\theta)$ . Our goal is to find situations were  $\|\mathcal{H}_{\text{ext}}\|_{sn}^2 > \|\mathcal{H}\|_{sn}^2$ .

Notice that this implies to start with a Hamiltonian that does not fulfil the requirement for equality given in Theorem 9.2. This excludes *de facto* the use of phase-shift Hamiltonians. Since they commute with their first derivative we can find a common set of eigenvectors to both of them, and then the inequality (9.17) is saturated, leaving no space for an enhancement using Hamiltonian extension.

### 9.3.1 Case study: The broken phase-shift

#### Repairing the phase-shift

To go beyond the phase shift we consider what we call a broken phase shift:  $K(\theta) = \theta G + F$ . We also introduce its evolution operator  $U_{K(\theta)} = e^{-iK(\theta)t}$  and its local generator  $\mathcal{K} = iU_{K(\theta)}^\dagger \dot{U}_{K(\theta)}$ . We assume furthermore that the condition for equality in (9.17) is not fulfilled (especially this implies that  $G$  and  $F$  do not commute) and we are thus in the situation were  $\|\mathcal{K}\|_{sn}^2 \leq t^2 \|G\|_{sn}^2$ . We want to find a Hamiltonian extension  $K_{\text{ext}}(\theta, \beta) := K(\theta) + K_1$  which saturates the upper bound:

$$C(K_{\text{ext}}(\theta); \theta) = t^2 \|G\|_{sn}^2. \quad (9.28)$$

There is a first solution for saturating the bound that looks trivial but works: As we know that the phase shift saturates the bound we can repair the broken phase shift by subtracting the annoying part,  $F$ . We are thus left with  $\theta G$  which saturates the upper bound. It is not necessary to subtract the entire operator  $F$  from the Hamiltonian. Theorem 9.2 teaches us that it would be sufficient that the extremal eigenvectors of  $G$  are also eigenvectors of  $K_{\text{ext}}$  to saturate the upper bound. If we use the eigenbasis of  $G$  — that we assume non-degenerate for simplicity — the Hamiltonian extension is written as

$$K_{\text{ext}}(\theta) = \theta \sum_i^d e_i |i\rangle \langle i| + \sum_{i,j=1}^d (F_{ij} + K_{1,ij}) |i\rangle \langle j|. \quad (9.29)$$

where  $F_{ij} := \langle i|F|j\rangle$  and  $K_{1,ij} := \langle i|K_1|j\rangle$ . If now we design a Hamiltonian extension such that  $K_{1,mn} = -F_{mn}$  for  $n = 1$  with  $m \in \{2, \dots, d\}$  and for  $n = d$  with  $m \in \{1, \dots, d-1\}$ , then  $|1\rangle$  and  $|d\rangle$  are eigenvectors of  $K_{\text{ext}}(\theta, \beta)$  and the extended Hamiltonian saturates the bound.

#### Shifting the parameter

In the method that we just presented the idea is to exactly cancel the part that spoils the Hamiltonian without touching the part that we are interested in. The second technique is rather the opposite: We do not try to cancel the annoying part but rather to enhance the part that produces

### 9.3 Reaching maximal sensitivity with Hamiltonian extensions

the signal. Obviously we cannot do this by adding a  $\theta$ -dependent part, as this will violate our definition of a Hamiltonian extension. It is still possible to increase the useful part without adding extra  $\theta$  dependence. The idea is to engineer the Hamiltonian in such a way that we work at a more favourable *effective* value of the parameter. For this purpose we introduce the Hamiltonian extension

$$K_{\text{ext}}(\theta, \beta) := \theta G + F + \beta G = K(\theta + \beta). \quad (9.30)$$

We thus see that with this Hamiltonian extension we shifted the parameter from  $\theta$  to  $\theta + \beta$ . Crucially this is not a reparametrization: We are still interested in  $\theta$  and not in  $\theta + \beta$ . We will investigate this point in the general case by going back to the Fisher information (FI).

**Going back to the FI** Consider the probability distribution  $\mu_\theta$  and its associated FI  $J(\mu_\theta; \theta)$  defined as

$$J(\mu_\theta; \theta) := \int dx \mu_\theta(x) \left( \frac{\partial \ln(\mu_\theta(x))}{\partial \theta} \right)^2. \quad (9.31)$$

In Sec. 1.3.2 we studied the effect of a change of parametrization in the FI. If we want to estimate the parameter  $g := g(\theta)$  then Theorem 1.2 states that the FI for  $g$  is given by

$$J(\mu_\theta; g)|_{\theta=\theta_0} = \frac{J(\mu_\theta; \theta)|_{\theta=\theta_0}}{\left( \frac{dg(\theta)}{d\theta} \Big|_{\theta=\theta_0} \right)^2}. \quad (9.32)$$

This reparametrization does not correspond to a change in the random process, the distribution is still the same. What changes is just the external point of view. In some way it corresponds to a passive transformation. If we think as a physicist we can say that it is only the data analysing that changes, but not the experiment.

*A contrario*, an active transformation changes the distribution itself. This is the case when we keep the same mathematical form of the distribution but replace the parameter. We go from  $\mu_\theta$  to  $\tilde{\mu}_\theta = \mu_{f(\theta)}$ . Using the notation  $f_0 := f(\theta_0)$  we can calculate the new FI

$$\begin{aligned} J(\mu_{f(\theta)}; \theta)|_{\theta=\theta_0} &:= \int dx \mu_{f_0}(x)^{-1} \left( \frac{\partial \mu_{f(\theta)}(x)}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 \\ &= \int dx \mu_{f_0}(x)^{-1} \left( \frac{\partial \mu_y(x)}{\partial y} \Big|_{y=f(\theta_0)} \frac{\partial f(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 \\ &= \left( \frac{\partial f(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 \int dx \mu_y(x)^{-1} \left( \frac{\partial \mu_y(x)}{\partial y} \right)^2 \Big|_{f=f_0} \\ &= \left( \frac{\partial f(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 J(\mu_y; y)|_{y=f_0}. \end{aligned}$$

By replacing  $y$  by  $\theta$  we obtain

$$J(\mu_{f(\theta)}; \theta)|_{\theta=\theta_0} = \left( \frac{\partial f(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right)^2 J(\mu_\theta; \theta)|_{\theta=f_0}. \quad (9.33)$$

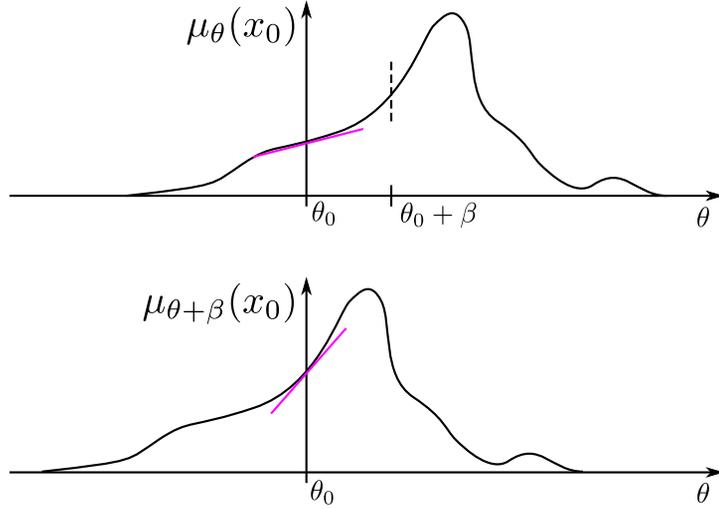


Figure 9.1: Change of probability distribution which amounts to a shift of the parameter in the FI.

There is a subtle point in this formula that makes it particularly interesting: In the left hand side we made appearing the FI of the original distribution ( $\mu_\theta$ ) for the original parameter ( $\theta$ ) *but taken at the value  $\theta = f(\theta_0)$*  (see Fig. 9.1).

In particular, for a shift in the parameter  $f(\theta, \beta) = \theta + \beta$  we obtain

$$J(\mu_{\theta+\beta}(x); \theta)|_{\theta=\theta_0} = J(\mu_\theta(x); \theta)|_{\theta=\theta_0+\beta} . \quad (9.34)$$

**Implementation in quantum metrology** This analysis shows how the Hamiltonian extension (9.30) can lead to an increased channel QFI. By going from  $\theta$  to  $\theta + \beta$  we end up with the channel QFI of the original Hamiltonian, *but at the value  $\theta_0 + \beta$  instead of  $\theta_0$* . Then we should think for which value of  $\theta$  the channel QFI is the highest. For the broken phase shift this is the case when the first part, that generates the useful signal, completely dominates, which means at very large  $\theta$ . Then we can take  $\beta$  arbitrarily large to get as close as we want to the upper bound. Without trying to formalize it too much we can write  $K_{\text{ext}}(\theta, \beta) = \beta[(1 + \varepsilon\theta)G + \varepsilon F]$  where  $\varepsilon = 1/\beta$ . In the limit  $\beta \rightarrow \infty$  the Hamiltonian becomes equal to  $\beta G$ , which of course has the same eigenvectors as  $G$  itself. The conditions of Theorem 9.2 are then fulfilled and the Hamiltonian extension saturates its upper bound.

### 9.3.2 Hamiltonian subtraction

We now turn to the study of the general case, inspiring us from the two methods proposed to saturate the upper bound when dealing with a broken phase shift.

The idea of subtracting the annoying part can be generalized without much effort. For the broken phase shift what we subtracted can be written as  $K(\theta) - \dot{K}(\theta)$ . We can do the same for a general Hamiltonian with two slight modifications. First we will have to work at a fix  $\theta_0$ . In the broken phase shift case we did not have to take care of this problem as  $K(\theta) - \dot{K}(\theta)$  turns to be  $\theta$ -independent. The second point is to notice that it is not necessary to subtract also " $-\dot{K}(\theta)$ ".

## Box 15: From active to passive transformation

If we introduce  $f := f(\theta)$  we have  $\theta = f^{-1}(f)$  and we can rewrite the terms appearing in Eq. (9.33) as

- $J(\mu_{f(\theta)}; \theta)|_{\theta=\theta_0} = J(\mu_f; f^{-1}(f))|_{f=f_0}$ ,
- $\frac{\partial f(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} = \frac{\partial f}{\partial(f^{-1}(f))} \Big|_{f=f_0} = \left( \frac{\partial(f^{-1}(f))}{\partial f} \Big|_{f=f_0} \right)^{-1}$ ,
- $J(\mu_\theta; \theta)|_{\theta=f_0} = J(\mu_f; f)|_{f=f_0}$ .

The formula for the active transformation now reads

$$J(\mu_f; f^{-1}(f))|_{f=f_0} = \left( \frac{\partial(f^{-1}(f))}{\partial f} \Big|_{f=f_0} \right)^{-2} J(\mu_f; f)|_{f=f_0}. \quad (9.35)$$

If we denote the inverse function of  $f$  as  $g$  (meaning that  $f^{-1}(x) = g(x)$ ) we eventually find

$$J(\mu_f; g(f))|_{f=f_0} = \frac{J(\mu_f; f)|_{f=f_0}}{\left( \frac{\partial(g(f))}{\partial f} \Big|_{f=f_0} \right)^2}, \quad (9.36)$$

which is nothing else than the formula for the reparametrization (Eq. (9.32)).

This calculation shows that the concepts of "passive" and "active" transformations make sense only from a physical point of view. From a pure mathematical point of view Eq. (9.32) and (9.33) are equivalent. It is the interpretation that we have of each one that corresponds to different physical situations.

Finally Hamiltonian subtraction amounts in subtracting the entire Hamiltonian at  $\theta_0$  from the original Hamiltonian, leading to the extension denoted with a subscript "sub":

$$H_{\text{sub}}(\theta) := H(\theta) - H(\theta_0). \quad (9.37)$$

This is a proper Hamiltonian extension as we add a  $\theta$ -independent parameter (we will come back to the dependence on  $\theta_0$  later on). As a matter of fact  $\dot{H}_{\text{sub}}(\theta) = \dot{H}(\theta)$ . The crucial point to show that this Hamiltonian extension saturates the upper bound is to look at its form at  $\theta = \theta_0$ :

$$H_{\text{sub}}(\theta_0) = 0. \quad (9.38)$$

The extended Hamiltonian vanishes at  $\theta = \theta_0$ . One could wonder in which sense this is an advantage? Rather we could think that there is nothing to estimate and that no information can be imprinted in any state. But this would be forgetting that in the local estimation theory we are interested in the neighbourhood of the distribution around  $\theta_0$ . Here the fact that the Hamiltonian vanishes results in the fact that it also commutes with every other operator, and especially its own derivative

$$\left[ H_{\text{sub}}(\theta_0), \dot{H}_{\text{sub}}(\theta_0) \right] = 0. \quad (9.39)$$

In virtue of Theorem 9.2 we have

$$C(H_{\text{sub}}(\theta); \theta)|_{\theta=\theta_0} = t^2 \left\| \dot{H}(\theta_0) \right\|_{sn}^2, \quad (9.40)$$

which shows that using Hamiltonian subtraction we saturate the upper bound.

As we did with the broken phase shift we can actually subtract the Hamiltonian only on the subspace we are interested in. We write  $\dot{H}(\theta) = \sum_i^d e_i(\theta) |\psi_i(\theta)\rangle \langle \psi_i(\theta)|$  where we made explicit that the eigenvalues and eigenvectors may depend on  $\theta$ . Furthermore we assume that  $e_1(\theta)$  and  $e_d(\theta)$  are not degenerate. To saturate the upper bound is enough to subtract  $R(\theta_0)$  such that  $\langle \psi_m(\theta_0) | R(\theta_0) | \psi_n(\theta_0) \rangle = -\langle \psi_m(\theta_0) | H(\theta_0) | \psi_n(\theta_0) \rangle$  for  $n = 1$  with  $m \in \{2, \dots, d\}$  and for  $n = d$  with  $m \in \{1, \dots, d-1\}$ . Indeed with such subtraction we enforce  $|\psi_1(\theta_0)\rangle$  and  $|\psi_d(\theta_0)\rangle$  to be eigenvectors of the extended Hamiltonian, a sufficient and necessary condition for the saturation of the bound in the non-degenerate case.

### Perturbation theory

One may argue that subtracting exactly the Hamiltonian is a difficult task — actually it is — especially as this should be done at the value  $\theta_0$ . We now will look at the effect of a small shift in the value of  $\theta_0$ . Instead of subtracting  $H(\theta_0)$  we subtract  $H(\theta_0 + \varepsilon)$ . We define the approximate Hamiltonian subtraction as

$$H_{\text{sub},\varepsilon}(\theta) := H(\theta) - H(\theta_0 + \varepsilon). \quad (9.41)$$

We consider  $\varepsilon$  to be a perturbative parameter and we thus obtain the expansion of the extended Hamiltonian

$$H_{\text{sub},\varepsilon}(\theta) = H(\theta) - H(\theta_0) - \varepsilon \dot{H}(\theta_0) - \varepsilon^2 \ddot{H}(\theta_0)/2 + \mathcal{O}(\varepsilon^3). \quad (9.42)$$

We are interested in the value of this Hamiltonian at  $\theta = \theta_0$  which reads

$$H_{\text{sub},\varepsilon}(\theta_0) = \varepsilon \left( -\dot{H}(\theta_0) - \varepsilon \ddot{H}(\theta_0)/2 + \mathcal{O}(\varepsilon^2) \right). \quad (9.43)$$

To calculate the channel QFI we need to express the local generator as

$$\mathcal{H}_{\text{sub},\varepsilon}|_{\theta=\theta_0} := i e^{i t H_{\text{sub},\varepsilon}(\theta_0)} \left( \frac{\partial}{\partial \theta} e^{-i t H_{\text{sub},\varepsilon}(\theta)} \right) \Big|_{\theta=\theta_0}. \quad (9.44)$$

Using Eq. (9.20) and the following expansion

$$e^X Y e^{-X} = Y + [X, Y] + \frac{1}{2!} [X, [X, Y]] + \dots, \quad (9.45)$$

we obtain

$$\mathcal{H}_{\text{sub},\varepsilon} = t(\dot{H}(\theta_0) - i \frac{\varepsilon^2 t}{2} \Gamma + \mathcal{O}(\varepsilon^3)), \quad (9.46)$$

where  $\Gamma = [\ddot{H}(\theta_0), \dot{H}(\theta_0)]/2$  and where we used

$$e^{-i t \alpha H_{\text{sub},\varepsilon}(\theta_0)} = \mathcal{I} - i t \alpha \varepsilon \left( -\dot{H}(\theta_0) - \varepsilon \ddot{H}(\theta_0)/2 + \mathcal{O}(\varepsilon^2) \right). \quad (9.47)$$

We now turn to the calculation of the semi-norm of the local generator  $\mathcal{H}_{\text{sub},\varepsilon}$ . In its eigenbasis  $\dot{H}(\theta_0)$  is written as  $\dot{H}(\theta_0) = \sum_{i=1}^d e_i |i\rangle \langle i|$ . Using perturbation theory at the first order we can calculate the modified eigenvalues  $e_i^{(\varepsilon)}$ . Assuming non-degenerate  $e_i$  we obtain

$$e_i^{(\varepsilon)} = t e_i - i \frac{\varepsilon^2 t^2}{2} \langle i | \Gamma | i \rangle + \mathcal{O}(\varepsilon^3). \quad (9.48)$$

Since the eigenvalues are not degenerate there is no crossing between eigenvalues with the perturbation and the extremal eigenvalues remains extremal. We are thus left with the channel QFI

$$C(H_{\text{sub},\varepsilon}(\theta); \theta) \simeq t^2 \left\| \dot{H}(\theta_0) \right\|_{sn}^2 - i\varepsilon^2 t^3 (e_1 - e_d) (\langle 1|\Gamma|1 \rangle - \langle d|\Gamma|d \rangle). \quad (9.49)$$

This shows that when we introduce a deviation of the order  $\varepsilon$  in the value of  $\theta_0$  we can still, up to corrections of the order  $\varepsilon^2$ , saturate the upper bound. Notice that the calculation was made simple by introducing the perturbation in  $\theta_0$ . We could have also considered a perturbation of the form  $H(\theta_0) + \varepsilon V$  with  $V$  arbitrary, but then the perturbation theory would have been more involved. A way to work out this situation would have been to use a similar formalism as the one used for the coherent averaging scheme.

### 9.3.3 Signal flooding

The second method used to saturate the upper bound for the broken phase shift was to somehow increase the useful part of the Hamiltonian. In the general case this useful part is its first derivative and we thus design the so-called *signal flooding* Hamiltonian

$$H_{\text{fl}}(\theta, \beta) = H(\theta) + \beta \dot{H}(\theta_0). \quad (9.50)$$

Again this is a valid Hamiltonian extension as the extra part  $\beta \dot{H}(\theta_0)$  is  $\theta$ -independent. To show that we can saturate the upper bound with this extension we will first show a more general result using the QFI and we will eventually move to the channel QFI.

The QFI for the signal flooding Hamiltonian is given by

$$I(U_{\text{fl}}|\psi_0; \theta) = \text{Var}[\mathcal{H}_{\text{fl}}, |\psi_0\rangle], \quad (9.51)$$

with the local generator  $\mathcal{H}_{\text{fl}} = iU_{\text{fl}}^\dagger \dot{U}_{\text{fl}}$  and the evolution operator  $U_{\text{fl}} = e^{-i t H_{\text{fl}}(\theta, \beta)}$ . As usual the crucial part is the derivative of the evolution operator. Using Eq. (9.18) we can write it as

$$\dot{U}_{\text{fl}}|_{\theta=\theta_0} = \left. \frac{\partial U_{\text{fl}}}{\partial \theta} \right|_{\theta=\theta_0} = -i t \int_0^1 e^{-i \alpha t H_{\text{fl}}(\theta_0, \beta)} \left. \frac{\partial H(\theta)}{\partial \theta} \right|_{\theta=\theta_0} e^{-i(1-\alpha)t H_{\text{fl}}(\theta_0, \beta)} d\alpha. \quad (9.52)$$

We write explicitly some of the derivation with respect to  $\theta$  because we will also use the derivative with respect to  $\beta$  — still the dot always means that we differentiate with respect to  $\theta$ . In particular the derivative of the evolution operator with respect to  $\beta$  reads

$$\left. \frac{\partial U_{\text{fl}}}{\partial \beta} \right|_{\theta=\theta_0} = -i t \int_0^1 e^{-i \alpha t H_{\text{fl}}(\theta_0, \beta)} \left. \frac{\partial H_{\text{fl}}(\theta_0, \beta)}{\partial \beta} \right|_{\theta=\theta_0} e^{-i(1-\alpha)t H_{\text{fl}}(\theta_0, \beta)} d\alpha. \quad (9.53)$$

Notice that the derivative of the signal flooding Hamiltonian with respect to  $\beta$  is independent of the value of  $\beta$ . This is just the translation from the fact that  $H_{\text{fl}}$  is a broken phase shift for  $\beta$ . Since at  $\theta = \theta_0$ , the derivative of the signal flooding Hamiltonian with respect to  $\beta$  is equal to its derivative with respect to  $\theta$ ,

$$\left. \frac{\partial H_{\text{fl}}(\theta_0, \beta)}{\partial \beta} \right|_{\theta=\theta_0} = \left. \frac{\partial H_{\text{fl}}(\theta, \beta)}{\partial \theta} \right|_{\theta=\theta_0}, \quad (9.54)$$

the same holds for the derivatives of the evolution operator

$$\left. \frac{\partial U_{\text{fl}}}{\partial \theta} \right|_{\theta=\theta_0} = \left. \frac{\partial U_{\text{fl}}}{\partial \beta} \right|_{\theta=\theta_0}. \quad (9.55)$$

As a consequence the local generator with respect to  $\theta$  is equal to the local generator with respect to  $\beta$ , which shows that the QFI for  $\theta$  at  $\theta_0$  is equal to the QFI for  $\beta$  at  $\theta = \theta_0$ :

$$I(U_{\mathbb{R}}|\psi_0\rangle; \theta)|_{\theta=\theta_0} = I(U_{\mathbb{R}}|\psi_0\rangle; \beta)|_{\theta=\theta_0} , \quad (9.56)$$

While this result is quite general we should now go back to the saturation of the bound. We already noticed that  $H_{\mathbb{R}}(\theta, \beta)$  is a broken Hamiltonian with respect to  $\beta$  with generator  $\dot{H}(\theta_0)$ . And we know that for estimating a broken phase shift we should try to work at large values of the parameter. Especially, in the limit of large  $\beta$  the Hamiltonian is completely dominated by  $\beta\dot{H}(\theta_0)$  and the broken phase shift saturates its upper bound:

$$C(H_{\mathbb{R}}(\theta, \beta); \beta) \xrightarrow{\beta \rightarrow \infty} t^2 \left\| \dot{H}(\theta_0) \right\|_{sn}^2 . \quad (9.57)$$

It is now easy to go back to the original question. Indeed Eq. (9.56) is true for all states  $|\psi_0\rangle$  and in particular for the state that maximizes the QFI. As a result we have

$$C(H_{\mathbb{R}}(\theta, \beta); \theta)|_{\theta=\theta_0} = C(H_{\mathbb{R}}(\theta, \beta); \beta)|_{\theta=\theta_0} . \quad (9.58)$$

This shows that signal flooding allows one to saturate the upper bound for the channel QFI,

$$C(H_{\mathbb{R}}(\theta, \beta)|_{\beta \gg 1}; \theta)|_{\theta=\theta_0} \simeq t^2 \left\| \dot{H}(\theta_0) \right\|_{sn}^2 . \quad (9.59)$$

### 9.3.4 Parameter dependence and ancillas

Before continuing we should make some comments about the two methods used for saturating the bound for arbitrary Hamiltonians, namely Hamiltonian subtraction and signal flooding.

The first comment is related to the use of  $\theta_0$ . In both extensions  $\theta_0$  appears. And this is the true value of the parameter that we are trying to estimate. If we know its value why shall we try to estimate it? It is actually the second time in this thesis that we ask this question. This is exactly the same problem that we faced when we looking at the optimal POVM that allows to maximize the FI (see Sec. 3.4.2). Without entering in the details again the solution comes from the fact that we are working in the formalism of local estimation theory. We suppose already known with a good precision the value of the parameter and we are interested in tracking some very small changes. Still what we want again to emphasize is that introducing terms depending on  $\theta_0$  is absolutely different that introducing terms that are  $\theta$ -dependent. The latter case implies to design a Hamiltonian which depends truly on the parameter. Typically this case excludes the use of an adaptive protocol to saturate the bound. This would be similar in spirit to the work done by Seveso et al. [2017] where they consider an extra parameter dependence through the POVM (see Box 7).

The second comment is about the use of ancillas. At the beginning of the chapter we wrote the extended Hamiltonian with three added terms: the interaction between probe and ancilla, the free evolution of the ancilla and the extra free evolution of the probe. Then we noted them as  $H_1$  and stopped writing explicitly the tensor product formally necessary to take into account the presence of the ancillas. We understand now the reason for this: We do not use ancillas neither with Hamiltonian subtraction nor with signal flooding. This is an important result from both a metrological point of view. It means that in full generality ancillas are not necessary to achieve the best sensitivity for a given Hamiltonian. While this was already known for the case of phase shift our methods show that it is completely general.

## 9.4 Time engineering with Hamiltonian extension

Hamiltonian extensions are not only useful to saturate the upper bound (9.17). There is a specific situation where Hamiltonian extensions lead to a real enhancement without being optimal. It concerns the time scaling of the channel QFI (and the QFI in general). In our discussion on parameter estimation with arbitrary Hamiltonians we saw that one can express the local generator as a function of the eigenvalues and eigenvectors of the Hamiltonian. Two contributions appear in the local generator, one depending on the derivative of the eigenvalues and coming along with a linear time dependence and a second one depending on the derivative of the eigenvectors coming along with a periodic time dependence. When the eigenvalues of the Hamiltonian are parameter independent the local generator loses its linear time dependence and conserves only the periodic time dependence. The QFI being only a function of this generator we end up with the situation where the channel QFI is periodic in time. This is especially damageable as it prevents to use the evolution time to obtain high QFI. All the more so that the linear time dependence in the local generator leads to a quadratic time scaling in the QFI in comparison to the linear time scaling obtained by classical averaging.

We will see how Hamiltonian extension can restore the quadratic time dependence. For the sake of simplicity we assume that the Hamiltonian has no degenerate eigenvalues:  $H(\theta) = \sum_{i=1}^d \lambda_i(\theta) |\psi_i(\theta)\rangle \langle \psi_i(\theta)|$ . Then Eq. (3.136) reads

$$\begin{aligned} \mathcal{H} = t \sum_{i=1}^d \frac{\partial \lambda_i(\theta)}{\partial \theta} |\psi_i(\theta)\rangle \langle \psi_i(\theta)| + 2 \sum_{k \neq l} e^{i t \frac{\lambda_k(\theta) - \lambda_l(\theta)}{2}} \sin\left(\frac{\lambda_k(\theta) - \lambda_l(\theta)}{2} t\right) \\ \times \langle \psi_l(\theta) | \frac{\partial \psi_k(\theta)}{\partial \theta} \rangle |\psi_k(\theta)\rangle \langle \psi_l(\theta)|, \quad (9.60) \end{aligned}$$

We furthermore assume that the eigenvalues of the Hamiltonian  $H(\theta)$  are  $\theta$ -independent:  $\lambda_i(\theta) \rightarrow \lambda_i$ . To restore the sought time scaling we have to design a Hamiltonian extension whose eigenvalues depend on the parameter. Using perturbation theory we can actually show that most of the Hamiltonian extensions can achieve this. We consider the Hamiltonian extension

$$H_{\varepsilon V}(\theta) = H(\theta) + \varepsilon V. \quad (9.61)$$

where  $V$  is an arbitrary operator. As we already did with Hamiltonian subtraction we use the first order time-independent perturbation theory to calculate the perturbed eigenvalues  $\lambda_i^{(\varepsilon V)}$ . Assuming non-degenerate  $\lambda_i$ , we obtain

$$\lambda_i^{(\varepsilon V)}(\theta) = \lambda_i + \varepsilon \langle \psi_i(\theta) | V | \psi_i(\theta) \rangle. \quad (9.62)$$

This simple analysis shows that as long as  $\langle \psi_i(\theta) | V | \psi_i(\theta) \rangle$  is not a constant function of  $\theta$  the introduction of  $\varepsilon V$  in the Hamiltonian leads to parameter dependent eigenvalues and *in fine* to the restoration of the quadratic time scaling in the channel QFI. In general this result shows that not only optimal Hamiltonian extensions can lead to an increase of the channel QFI, but rather one can check case-by-case if a given extension provides an advantage or not.

## 9.5 Known results

**Saturation of the upper bound with controls** The question of the time scaling was already discussed in [Yuan and Fung, 2015; Yuan, 2016], also based on the results from [Pang and Brun,

2014]. There the authors proposed a scheme using feedback controls to restore the quadratic time scaling. The idea is to break the evolution into many time steps and to intersperse controls in between. They start with a Hamiltonian  $H_\theta$  along with its evolution operator  $U(T, \theta) := e^{-iTH(\theta)}$ . Then they divide this evolution in  $m$  parts and use a control in between each part:

$$U_{\text{YF}} := U_m U(t, \theta) \cdots U_3 U(t, \theta) U_2 U(t, \theta) U_1 U(t, \theta) , \quad (9.63)$$

where  $t = T/m$  and the operators  $U_i$  are the controls. With an optimal choice of the controls and in the limit of large  $m$ , which implies small  $t$ , they are able to also saturate the upper bound (9.17).

**Broken phase shift and dithering** De Pasquale et al. [2013] studied a broken phase shift in the form  $K = \theta G + \eta F$ , where the parameter of interest is  $\theta$ . The authors studied the effect of  $\eta$  on the channel QFI. While it is clear that the optimal choice is  $\eta = 0$ , as then we recover a phase shift which saturates its upper bound, it is not clear given a value  $\eta_0$  if it is better to increase or decrease the value. One should not think that because the optimal value is zero then the channel QFI is a monotone function of  $\eta$ . It actually turns out that for some values of  $\eta_0$  it is better to increase the value of  $\eta$  rather than to decrease it, an effect that the authors called "dithering". Obviously this is only true locally. If one has the possibility to increase or decrease arbitrarily then bringing  $\eta$  to zero is always optimal.

## 9.6 Examples and applications

To conclude this chapter we will illustrate the different methods of Hamiltonian extension by studying two examples, both in magnetometry.

### 9.6.1 NV center magnetometry

#### Hamiltonian of the NV center

We first study the estimation of the magnetic field in a given direction using a nitrogen vacancy center. Nitrogen-vacancy centers, known as NV centers, correspond to defects in the lattice of a diamond: There is a nitrogen atom in substitution of a carbon atom adjacent to a vacancy, *i.e.* the lack of a carbon atom. These defects can be negatively charged, neutral or positively charged. The most interesting for magnetometry is the negatively charged one, and when we will refer to NV centers we will always refer to the negatively charged one. This defect involves six electrons. Two coming from the nitrogen, three from the dangling bond around the vacancy and the last one is given by the lattice. We are more particularly interested in a spin triplet of the NV center, which can be efficiently monitored through optical means and has a coherence time as long as a few microseconds. It is quite remarkable that these NV centers can have a quantum behaviour at room temperature. Due to the dependence of the spin triplet on external magnetic fields, these systems are good candidates to design extremely sensitive magnetometers or magnetometers with very high spatial resolution. A recent review can be found in [Schirhagl et al., 2014; Rondin et al., 2014].

If we neglect the interaction with the  $^{14}\text{N}$  nuclear spin as well as the bath of the  $^{13}\text{C}$  nuclear spins, the Hamiltonian  $H_{\text{NV}}$  for the triplet state of the NV center can be written as [Rondin et al., 2014]

$$H_{\text{NV}} = g\mu_{\text{B}}(B_x S_x + B_y S_y + B_z S_z)/\hbar + DS_z^2/\hbar + E(S_x^2 - S_y^2)/\hbar , \quad (9.64)$$

with  $g$  the Landé factor,  $\mu_B$  the Bohr magneton,  $D$  and  $E$  the zero field splitting parameters and  $S_x, S_y$  and  $S_z$  the spin-1 matrices, fulfilling  $[S_i, S_j] = i\hbar\varepsilon_{ijk}S_k \forall i, j, k \in \{x, y, z\}$ . Notice that in this section we come back to the usual value of  $\hbar$  ( $\hbar \simeq 6,626 \times 10^{-34}/(2\pi)$ ). The zero field splitting has two components, the axial one, with parameter  $D$  (taken as  $D = 2\pi \times 2.87$  GHz), and the off-axis one, with parameter  $E$  (taken as  $E = 2\pi \times 5$  MHz).

### Magnetometry with NV center

From the point of view of metrology we are interested in the estimation of  $B_z$ , the magnetic field in the direction of the axis between N and V. The Hamiltonian of the spin triplet is with respect to  $B_z$  a broken phase shift  $H_{NV} = B_z G + F$  with  $G = g\mu_B S_z/\hbar$ . Using Theorem 9.1 we find that the channel QFI is bounded by

$$C(H_{NV}; B_z) \leq (t/\hbar)^2 \|G\|_{sn}^2 = (tg\mu_B/\hbar)^2 \|S_z/\hbar\|_{sn}^2 = 4(tg\mu_B/\hbar)^2. \quad (9.65)$$

We plot in Fig. 9.2 (log-log plot) the channel QFI of the original Hamiltonian (dashed line). We see that for large values of  $B_z$  the channel QFI presents a plateau close to its upper bound (dotted line). When  $B_z$  decreases there exists a threshold, here roughly equal to  $10^{-1}$  T which is the value of  $B_x$ , after which the channel QFI starts to decrease. We can explain this as the fact that for these values the part that spoils the phase shift starts to dominate. Interestingly there is another threshold, here close to  $10^{-8}$  T, after which the channel QFI reaches another plateau.

One can actually show that the existence of this plateau for low values of the parameter is a quite general feature of broken phase shift  $K(\theta) = \theta G + F$ . We will work out the extreme case where the parameter of the phase shift vanishes and try to get some insights about  $C(K(\theta); \theta)|_{\theta=0}$ . To do so we should calculate the local generator  $\mathcal{K}$  at  $\theta = 0$ . Using Eq. (9.20) we find

$$\mathcal{K}|_{\theta=0} = t \int_{-1}^0 e^{-i\alpha t F} G e^{i\alpha t F}. \quad (9.66)$$

We furthermore assume that  $F$  does not have degenerate eigenvalues and we write it as  $F = \sum_i f_i |i\rangle\langle i|$ . In this basis  $G$  is written as  $G = \sum_{i,j} g_{ij} |i\rangle\langle j|$ , with  $g_{ij}^* = g_{ji}$  and the local generator at  $\theta = 0$  reads

$$\mathcal{K}|_{\theta=0} = t \sum_i g_{ii} |i\rangle\langle i| + i \sum_{i \neq j} g_{ij} \frac{1 - e^{i t(f_i - f_j)}}{f_i - f_j} |i\rangle\langle j|. \quad (9.67)$$

It is not easy from this formula to calculate the semi-norm of  $\mathcal{K}|_{\theta=0}$  as this requires to calculate the extremal eigenvalues. Still we can at least investigate if the semi-norm is finite or equal to zero. To be equal to zero it requires that its extremal eigenvalues are equal. This is equivalent to say that the semi-norm will be equal to zero only when the operator is proportional to the identity operator<sup>4</sup>. The identity operator is written in the same way in any basis, and thus especially in the eigenbasis of  $F$ . This means that in our case the channel QFI will be equal to zero if and only if  $\mathcal{K}|_{\theta=0}$  is proportional to the identity. In general this will be the case if  $t(f_i - f_j)$  is an integer multiple of  $2\pi$  and  $g_{ii} = g_{jj}$  for all  $i$  and  $j$ . Although this is not a necessary condition (some  $g_{ij}$  may already be equal to zero) it still shows that apart from a very particular case the channel QFI of a broken phase shift is not equal to zero even when the parameter is equal to zero.

---

<sup>4</sup>And not when the operator is equal to zero itself. This is why it is a semi-norm and not a norm.

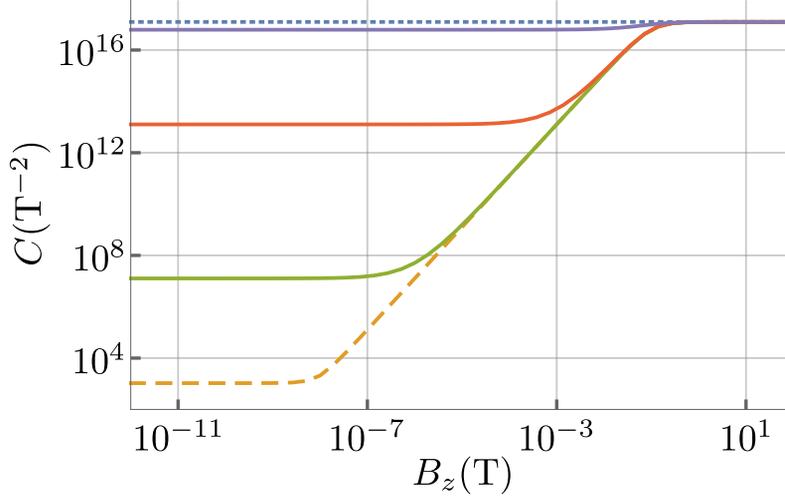


Figure 9.2: Channel QFI for the NV center ( $t = 10^{-3}$  s,  $B_x = 10^{-1}$  T and  $B_y = 0$  T), as a function of the parameter to be estimated,  $B_z$ . The dotted line represents the upper bound, the dashed line the channel QFI of the original Hamiltonian (9.64) and the continuous lines the channel QFI of the extended Hamiltonian (9.68) for different values of  $\beta$ : From bottom to top we have  $\beta = 10^{-6}$ ,  $\beta = 10^{-3}$  and  $\beta = 10^{-1}$ .

### Signal flooding with the NV center

We can use signal flooding to saturate the upper bound (9.65). We define the signal flooding Hamiltonian for the NV center as

$$H_{\text{NV},\text{fl}} = H_{\text{NV}} + \beta \dot{H}_{\text{NV}} = H_{\text{NV}} + \beta g \mu_B S_Z / \hbar. \quad (9.68)$$

In this situation signal flooding amounts in adding a magnetic field with known amplitude in the  $z$ -direction. Notice that this strategy is already heuristically known. When fields are very weak adding an external magnetic field allows one to reach a linear Zeeman effect [Rondin et al., 2014].

We represented in Fig. 9.2 the channel QFI for the signal flooding Hamiltonian for different values of  $\beta$  (plain lines). As we discussed, signal flooding for a broken phase shift corresponds to evaluating the QFI at shifted values of the parameter (at  $\theta + \beta$  instead of  $\theta$ ). We see in the plot that the larger the value of  $\beta$  the higher the channel QFI. In the case  $\beta = 10^{-1}$  T we are very close to the upper bound for all values of  $B_z$ , which shows that with this technique we can reach a very high sensitivity even if the magnetic field is very weak.

### 9.6.2 Vector magnetometry with spin-1

Broken phase shifts are not necessarily the most interesting examples — although they are very important examples as phase shift and broken phase shift appears often in metrology — since the effect of signal flooding is pretty trivial there. In this section we will look at an example that does not correspond to a broken phase shift. This will allow us to nicely illustrate the three different kinds of Hamiltonian extensions that we developed.

We will again study a task of magnetometry, but this time we are not interested in the value of the field but rather in its direction. To stay in a scalar parameter framework we will only estimate

the direction of the field in a given plane. We introduce the Hamiltonian for a spin-1 in a magnetic field,

$$H(B, \theta, \varphi) = g\mu_B \mathbf{B} \cdot \mathbf{S} / \hbar, \quad (9.69)$$

with  $\mathbf{B} = B(\sin(\theta) \cos(\varphi), \sin(\theta) \sin(\varphi), \cos(\theta))$  and  $\mathbf{S} = (S_x, S_y, S_z)$ . We want to estimate the spherical angle  $\theta$  assuming  $\varphi$  known.

The upper bound for the channel QFI is given by

$$C(H(B, \theta, \varphi); \theta) \leq (t/\hbar)^2 \left\| \frac{\partial H(B, \theta, \varphi)}{\partial \theta} \right\|_{sn}^2 = 4(g\mu_B B)^2, \quad (9.70)$$

since the eigenvalues of  $\partial H(B, \theta, \varphi) / \partial \theta$  are 0 and  $\pm g\mu_B B$ . The local generator  $\mathcal{H}$  of the translation in  $\theta$  can be computed exactly and its eigenvalues are 0 and  $\pm 2 \sin(g\mu_B B t / (2\hbar))$ . As a result the channel QFI is given by

$$C(H(B, \theta, \varphi); \theta) = 16 \sin^2(g\mu_B B t / (2\hbar)). \quad (9.71)$$

Since the eigenvalues of  $H(B, \theta, \varphi)$  are  $\theta$ -independent this channel QFI has a periodic time dependence. We represent the channel QFI as a function of time in Fig. 9.3 (dashed yellow line). We clearly see the periodic behaviour that prevents us to use the time as a resource. As a result we see that the discrepancy between the upper bound (dotted line) and the channel QFI for the original Hamiltonian increases quadratically as a function of time.

### Signal flooding

We start by applying signal flooding to our system. The signal flooding Hamiltonian is

$$H_{\text{fl}}(B, \theta, \varphi) = H(B, \theta, \varphi) + \beta g\mu_B \tilde{\mathbf{B}} \cdot \mathbf{S} / \hbar. \quad (9.72)$$

where  $\tilde{\mathbf{B}} = B(\cos(\theta_0) \cos(\varphi), \cos(\theta_0) \sin(\varphi), -\sin(\theta_0))$  which can also be written as  $\tilde{\mathbf{B}} = B(\sin(\theta_0 + \pi/2) \cos(\varphi), \sin(\theta_0 + \pi/2) \sin(\varphi), \cos(\theta_0 + \pi/2))$ . Signal flooding amounts here in adding a magnetic field with strength  $\beta B$  in the direction opposite to the direction of the original magnetic field. We can see in Fig. 9.3 the effect of signal flooding on the channel QFI (plain lines). For an additional field five times stronger than the original one ( $\beta = 5$ ), we see that the upper bound is already almost saturated.

### Time engineering

The original channel QFI being periodic in time we can apply the technique of time engineering to restore the quadratic time scaling. We consider the following channel extension:

$$H_{S_z}(B, \theta, \varphi) = H(B, \theta, \varphi) + \kappa B g\mu_B S_z / \hbar. \quad (9.73)$$

This corresponds to the addition of a magnetic field in the  $z$ -direction with a strength  $\kappa B$ . The eigenvalues of  $H_{S_z}(B, \theta, \varphi)$  are 0 and  $\pm g\mu_B \sqrt{B^2 + \kappa^2 + 2B\kappa \cos(\theta)}$ . We see in Fig. 9.3 (dotted-dashed line) that this extension restores the quadratic time scaling. We also see that the larger the value of  $\kappa$  the larger the channel QFI. Still if we look at the values of  $\kappa$  we used, the difference with signal flooding is clear. While for low values of  $\beta$  and  $\kappa$  both Hamiltonian extensions perform more or less equally, for large values the channel QFI of  $H_{S_z}(B, \theta, \varphi)$  saturates below the upper bound. Indeed we see that between  $\kappa = 10$  and  $\kappa = 10^9$  the channel QFI does almost not increase anymore. This is because after a certain threshold the added magnetic field dominates completely the dynamics, while at the same time its corresponding Hamiltonian does not fulfil the condition of Theorem 9.2.

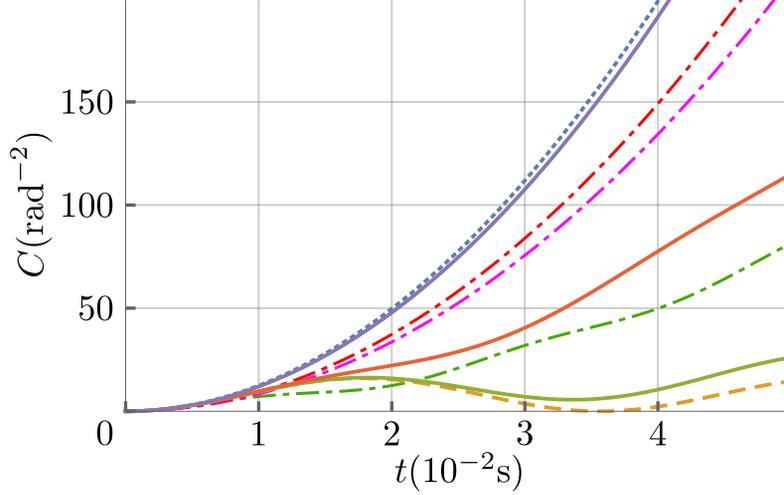


Figure 9.3: Channel QFI for a direction of the magnetic field ( $B = 10^{-9}$  T,  $\varphi = \pi/4$  and  $\theta = \pi/3$ ), as a function of the time. The dotted line represents the upper bound, the dashed line the channel QFI of the original Hamiltonian (9.69). The continuous lines represent the channel QFI for the "signal flooding" Hamiltonian (9.72) for different values of  $\beta$ : From bottom to top we have  $\beta = 0.2$ ,  $\beta = 0.75$  and  $\beta = 5$ . The dotted-dashed lines represent the channel QFI for the extended Hamiltonian (9.73), for different values of  $\kappa$ : From bottom to top we have  $\kappa = 1$ ,  $\kappa = 10$  and  $\kappa = 10^9$ .

### Hamiltonian subtraction

Finally we go to the last technique: Hamiltonian subtraction. We will focus on the effect of a slight deviation from the value of  $\theta_0$  in the added Hamiltonian. We thus introduce the perturbed Hamiltonian

$$H_{\text{sub},\varepsilon}(B, \theta, \varphi) = H(B, \theta, \varphi) - H(B, \theta_0 + \varepsilon, \varphi). \quad (9.74)$$

It is possible to compute exactly the local generator for this Hamiltonian and therefore its channel QFI, which reads

$$C(H_{\text{sub},\varepsilon}(B, \theta, \varphi); \theta)|_{\theta=\theta_0} = 4(g\mu_B B t / \hbar)^2 \cos^2(\varepsilon/2) + 4 \sin^2(g\mu_B B t \sin(\varepsilon/2) / \hbar). \quad (9.75)$$

As it should be, if we set  $\varepsilon = 0$  in this equation we recover the value of the upper bound  $4(tg\mu_B B / \hbar)^2$ . We can also verify that the perturbative results obtained in Sec. 9.3.2 agree with the results obtained here. This is the case as the correction of second order in  $\varepsilon$  in the channel QFI vanishes exactly, and the leading order corrections are of order  $\varepsilon^4$ , demonstrating the stability of the method to perturbations in the direction of the subtracted magnetic field, as represented in Fig. 9.4.

### 9.6.3 Application to the coherent averaging

The formalism developed until here can directly be applied to the coherent averaging protocol. We can formally identify two Hamiltonian extensions in the coherent averaging protocol:

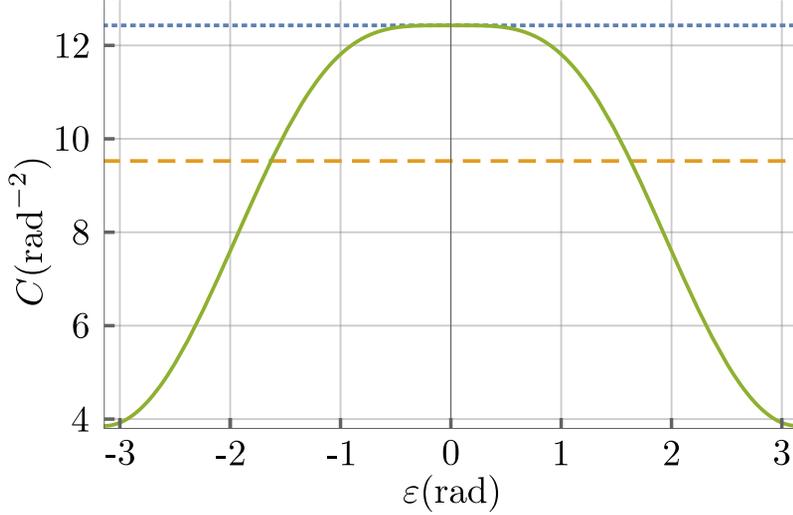


Figure 9.4: Effect of a perturbation in "Hamiltonian subtraction" for the estimation of a direction of a magnetic field. ( $B = 10^{-9}$  T,  $\varphi = \pi/4$ ,  $\theta_0 = \pi/3$  and  $t = 10^{-2}$ s). The dotted line represents the upper bound, the dashed line the channel QFI of the original Hamiltonian (9.69), and the continuous line the channel QFI of the Hamiltonian (9.74).

- (a) We can consider that the original Hamiltonian is the one of the  $N$  probes. Then the introduction of the bus corresponds to a Hamiltonian extension with ancillas.
- (b) We can consider that the original Hamiltonian is the one of the quantum bus. The  $N$  probes play the role of ancillas in the extension.

From the point of view (a) we conclude that by introducing the quantum bus we should not hope to surpass the HL scaling for  $\omega_p$  (parameter of the free evolution of the probe). Here lies a very crucial point that illustrates the limits of discussing only the channel QFI. Indeed if we rely only on the channel QFI we could conclude that coherent averaging is useless. But we know that the optimal state needed to reach the HL is a massively entangled state, difficult to produce for large number of probes and very fragile. The fact that the coherent averaging protocol allows to reach a HL scaling starting with a separable state consists in an actual enhancement.

From the point of view (b) we conclude that the coherent averaging protocol is of no use for the estimation of  $\omega_b$  (parameter of the free evolution of the bus). Indeed in contrast with  $\omega_p$ , the upper bound for the estimation of  $\omega_b$  does not depend on  $N$ : there is no HL or SQL scaling to try to reach. In particular the optimal state is not entangled and we cannot argue that the coherent averaging may help to reach the upper bound without entanglement<sup>5</sup>.

---

<sup>5</sup>Entanglement is not the only resource that one can consider. It would be interesting to consider the purity or the amount of coherences.

## Summary Chapter 9

- **Hamiltonian extension:** Adding an extra term to the Hamiltonian  $H(\theta)$ , which possibly includes interactions with an ancilla.
- **Upper bound:** The channel QFI, and then the QFI, for the parameter  $\theta$  in  $H(\theta)$  is upper bounded by the semi-norm of  $\dot{H}(\theta)$ :  $C(H(\theta); \theta) \leq t^2 \left\| \dot{H}(\theta) \right\|_{sn}^2$ , where the semi-norm is equal to the difference between the maximal and the minimal eigenvalue.
- **Condition of saturation:** A sufficient condition for saturating the bound is that a maximal and a minimal eigenvector of  $\dot{H}(\theta)$  are also eigenvectors of  $H(\theta)$ . Especially if  $H(\theta)$  commutes with  $\dot{H}(\theta)$  the bound is saturated.
- **Signal flooding:** We can saturate the upper bound by adding  $\beta \dot{H}(\theta_0)$  to  $H(\theta)$  and take  $\beta$  to be large.
- **Hamiltonian subtraction:** We can saturate the upper bound by subtracting  $H(\theta_0)$  from  $H(\theta)$ . This method is robust against perturbation in the value of  $\theta_0$ .
- **Use of ancillas:** Signal flooding and Hamiltonian subtraction do not use ancillas. This shows that ancillas are not necessary to reach the maximal sensitivity allowed by a Hamiltonian.
- **Time engineering:** When the eigenvalues of  $H(\theta)$  are  $\theta$ -independent, the QFI is periodic in time. It turns out that almost any Hamiltonian extension can restore the quadratic time scaling.

# Conclusion and outlook

*Metrology does not consist in predicting the weather*

Anonym

In this thesis we worked on several aspects of quantum metrology, showing how the tools of parameter estimation theory can serve to analyze very different problems.

We have studied the estimation of the depolarizing channel and the phase-flip channel in a non-optimal situation [Fraïsse and Braun, 2017c]. Our study was motivated by understanding the effect of considering many ancillas in quantum metrology. In an optimal context it is known that one ancilla is enough to obtain the maximal channel QFI. To go beyond this result we have taken into account the possible loss of subsystems. When losing the probe there is nothing to estimate anymore and adding ancillas does not help. But when losing some ancillas, we showed that using a W state with many ancillas can help to preserve a part of the gain provided by the ancillas, a property that we called robustness.

While this demonstration shows that using many ancillas can be useful, the method does not seem suited for applications. Producing entangled states of many particles is a difficult task, and the improvement offered by the use of the W states seems too small to justify the use of so many resources. Further research could focus on the use of entanglement between different probes and not only between the probe and the ancilla, in line with the studies of Collins and Stephens [2015] and Collins [2013]. Another direction would be to look at the optimal state in terms of robustness. Such study could be done using the channel QFI. What would be needed is to replace the original channel by the channel concatenated with the loss channel. The new channel to be estimated would then be a channel acting on  $l + 1$  qubits, which also shows that the optimal state only requires to use  $2l + 1$  ancillas.

The two other topics that we have studied concern Hamiltonian parameter estimation. One was the study of the coherent averaging protocol. In its original version [Braun and Martin, 2011] this protocol was used to estimate a parameter encoded in the interaction Hamiltonian between the bus and the probes. While this applies to some relevant physical situations it was not allowing a direct comparison with the standard metrological protocol where one uses  $N$  probes and estimates a parameter encoded in their free evolution Hamiltonian.

We have extended the original perturbative results to the estimation of the parameters of the free Hamiltonians, and especially of the one characterizing the free evolution of the probes [Fraïsse and Braun, 2015]. We have shown that in the strong interaction regime we reach the Heisenberg limit scaling for this parameter even though we start with a separable state. This participates in the effort of finding quantum enhanced measurements not primarily based on the use of initial

## Conclusion

entangled states [Braun et al., 2017]. We have also introduced two models of coherent averaging protocol where both the bus and all the probes are qubits. Both the numerical and analytical exact results (meaning non-perturbative) for these two models challenged the idea that we can always reach the Heisenberg limit scaling for the interaction parameter by measuring only the bus. It appears that outside the regime of validity of the perturbation the scaling often breaks down.

Since all the studies of the coherent averaging protocol have been oriented to the study of the scaling with the number of probes the question of the regime of validity should be more thoroughly studied. Indeed the motivation to focus on the scaling is that for a large enough number of probes, the QFI with a Heisenberg limited scaling will always be higher than the QFI with a SQL scaling, whatever are the numerical prefactors. To make one step further in direction of actual experiments, it would also be useful to take into consideration decoherence, which has proven to be crucial to assess the real enhancement offered by any quantum metrological protocol [Huelga et al., 1997; Kolodynski, 2014].

The last topic we have studied was inspired by our work on channel estimation. The original motivation was to understand how the estimation of Hamiltonian parameters benefits from ancillas. We thus introduce the concept of Hamiltonian extension, where we add a parameter-independent operator to the Hamiltonian, which possibly encompasses an interaction with an ancilla [Fraïsse and Braun, 2017b]. Ironically it turned out that ancillas do not really play a role in the formalism but that this formalism could be used to analyze the coherent averaging protocol. In general this research participates in the understanding of metrology with arbitrary Hamiltonians [Giovannetti et al., 2006; Pang and Brun, 2014], a growing field, as for years the research was mainly focused on phase-shift Hamiltonians.

We have used a bound on the channel QFI derived by [Boixo et al., 2007] and have specified the condition for reaching the bound. Moving from phase-shift Hamiltonians has allowed us to show that Hamiltonian extensions can lead to an increase of the channel QFI. Especially we have introduced two methods of Hamiltonian extension that saturate the upper bound [Fraïsse and Braun, 2017a]. These two methods do not require the use of any ancillas, demonstrating that ancillas are not necessary to obtain the maximal QFI that can be reached given a Hamiltonian. This is in agreement with the results of Yuan and Fung [2015], who proposed a method based on control feedback that also saturates the bound.

Several directions could be considered to go beyond our results. On one side we could include decoherence effects. By taking a master equation approach we could keep the formalism of Hamiltonian extensions and verify if in this noisy case ancillas are still not necessary. Notice that [Demkowicz-Dobrzański and Maccone, 2014] found a situation where adding passive ancillas gives an advantage in presence of noise while it does not in the noiseless case. Another approach would be to consider different figures of merit. While the channel QFI enjoys a very fundamental meaning, besides the fact that its calculation is relatively easy, it may also hide some aspects of the problem, as it completely avoids the question of the initial state. Going back to QFI could help to discover new applications of Hamiltonian extensions. For example when the Hamiltonian acts on several subsystems it would be interesting — in the spirit of coherent averaging — to try to design Hamiltonian extensions that allow to reach the maximal QFI starting with a separable state.

Eventually we also want to say a word about the Part I of this thesis. There we presented the field of quantum metrology and quantum parameter estimation. We tried to give an introduction to the field that could be useful beyond our own results. Indeed we believe that quantum metrology and quantum parameter estimation are now mature fields that are summoned to play

an increasingly important role in physics. For fundamental physics — as it is already the case — but also for engineering and applied physics. We hope that this contribution would participate in the effort to spread quantum metrology and quantum parameter estimation in the community.



# Appendix A

## Probability theory

### A.1 Foundation of probabilities

Probability theory is the modern way to handle randomness. It aims to give a mathematical framework to random processes, *i.e.* processes where the result cannot be predicted exactly. Although the concepts of "chance", "luck" or "hazard" have been present in the history of humanity since thousands of years, the proper formalization of randomness goes back only to the beginning of the last century.

The history of probability is closely related to the games of chance. To analyse them, mathematicians were led to try to formalize the concept of probability in the 16th and 17th centuries. While this analysis stays intuitive when dealing with games and processes involving a discrete number of results, the treatment of random processes with a continuous number of results leads to mathematical difficulties. The proper answer to handle such cases was given by Kolmogorov.

#### Box 16: Intuitive approach to probabilities and historic difficulties

The common way to express "chance" was for long done by the use of ratios. To express the degree of chance that something happens, people would use formulas as "there are three chances over ten that this event will occur", *i.e.* they divided the desired results over the number of possible outcomes. These calculations come from counting the number of cases that lead to the desired result over the total number of cases. With a dice with six faces we say intuitively that the chance that the dice shows a one is *one over six*.

So far so good. But how general is such rule? Imagine asking somebody to randomly give a real number between zero and one. Whatever his answer is, we would have attributed the probability zero to it following our empirical rule, since there was an infinite number of possible answers. This very simple example shows that when dealing with continuous random processes, the intuitive approach may lead to paradoxical conclusions.

## A.2 Probability space

Probability theory aims to formalize random processes. First one has to define the process itself. This is done by defining all the possible outcomes of the experiment. The space corresponding to all the outcomes is known as the sample space  $\Omega$ . A *realization* of the process leads to the observation of one specific outcome  $\omega_i \in \Omega$ . The process being random, we cannot predict which outcome is obtained, but we want to give the probability of observing each outcome. In general this is too restrictive, since we are not necessarily interested in single elements of  $\Omega$  but in sets of elements of  $\Omega$ , so called *events*. Therefore we define a set  $\mathcal{A}$  of events. This set  $\mathcal{A}$  is a  $\sigma$ -algebra, and obeys the three properties

- (a)  $\mathcal{A}$  is not empty
- (b)  $\mathcal{A}$  is closed under complementation: If  $a$  is in  $\mathcal{A}$  then its complement  $\Omega \setminus a$  is also in  $\mathcal{A}$
- (c)  $\mathcal{A}$  is closed under countable unions: If  $a_1, a_2, \dots, a_n$  are elements of  $\mathcal{A}$ , then the union of them  $a_1 \cup a_2 \cup \dots \cup a_n$  is also in  $\mathcal{A}$

We introduce the probability of obtaining an event through a probability measure  $\mu$ ,

$$\begin{aligned} \mu : \mathcal{A} &\rightarrow [0, 1] \\ a &\mapsto \mu(a), \end{aligned}$$

which fulfills  $\mu(\Omega) = 1$ . Therefore a random process is completely specified by a probability space, *i.e.* a triplet  $(\Omega, \mathcal{A}, \mu)$

### Box 17: Dice and random process

We want to model a random process involving a six-faces dice, the faces labelled from one to six. The first step is to define the experiment, which is tantamount to define the possible outcomes. Consider that the process consists of throwing the dice once and observing the result. The probability space is thus  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . This corresponds to the possible outcomes of the experiment. The next step is to define the events we are interested in, which mathematically translates to defining a  $\sigma$ -algebra. This could be the parity of the result. Therefore we define the  $\sigma$ -algebra  $\mathcal{A}_P = \{\Omega, \emptyset, \{2, 4, 6\}, \{1, 3, 5\}\}$ . Then the last step consists in assigning a probability to each event. We thus introduce a measure *which is a function of the  $\sigma$ -algebra and not a function of the outcomes*. In our example it does not make sense to define the probability of getting a five. We can just define the probability of getting an odd or an even number. For example a first dice could have the measure  $\mu_1$  defined by  $\mu_1(\{1, 3, 5\}) = 1/2 = \mu_1(\{2, 4, 6\})$ . Notice that this result does not tell us if the dice is fair or not. By definition, with  $\mathcal{A}_P$  we can just give a statement about the parity. A different dice could give us the measure  $\mu_2$  with  $\mu_2(\{1, 3, 5\}) = (1 + f(T))/2$  and  $\mu_2(\{2, 4, 6\}) = (1 - f(T))/2$  where  $f(T)$  is a known function of the temperature  $T$  and fulfilling  $-1 \leq f(T) \leq 1$ . This last case drives us to the question of parametric random processes, and thus to the question of parameter estimation.

### A.2.1 Interpretation of probabilities

Using the concept of measures, we give the proper tool to calculate probabilities, but we do not explain their interpretation. What is thus the meaning of a "probability"?

Actually this question is present since the beginning of the theory of probability and no agreement has been found up to day. Plenty of interpretations have been proposed, but at a first sight, it is enough to distinguish to main categories of interpretations.

#### Frequentist interpretation

The first approach, called *frequentist interpretation*, is based on the idea of repeating the random process. The probability  $\mu(a)$  of an event  $a$  is interpreted as the ratio of the number of occurrence  $\#_a$  to the total number of realizations  $\#_{\text{tot}}$  of the process in the limit of an infinite number of repetitions:

$$\mu(a) \sim \lim_{n \rightarrow \infty} \frac{\#_a}{\#_{\text{tot}}} . \quad (\text{A.1})$$

This interpretation of probability is very natural when one is interested in a random process easily reproducible under controlled conditions. But probabilities are not exclusively used in this framework. When scientist attributes a probability to a volcano to enter in eruption, we cannot exactly talk of a "process easily reproducible under controlled conditions".

#### Bayesian interpretation

In the Bayesian interpretation, there is no reference to an asymptotic property obtained by repetition. Instead probabilities are considered as *degree of beliefs* and are thus subjective, depending on the agent who enunciates the probability. In this framework, saying that there is a probability one half that a volcano enters in eruption in the next hundred years means that the we believe that there is one chance over two that the event will happen. At this point we encounter the problem that we use again the notion of chance to define the probability.

#### So what

So which interpretation do we adopt? Actually it is of utmost importance to note that the question of interpretation does not affect the theory from a mathematical point of view. The axiomatic approach of Kolmogorov is consistent with both interpretations. As a matter of fact, most of the mathematicians have adopt a strategy of "calculate and shut up". But then why do we hear people claiming to be frequentist or Bayesians? The point is that depending on the interpretation of probabilities one will have the tendency to prefer some methods when working on a specific statistical problem. This applies for example to parameter estimation theory, where there is the frequentist approach and the Bayesian approach.

## A.3 Real random variables

In our definition of the probability space we did not put restrictions on the elements of the sample space  $\Omega$ , and therefore on the events. This freedom is an important characteristic of the axiomatic approach of probabilities. But when going from probability theory to statistics it can become problematic. If a random process has two events, say "Blue" and "Yellow", with probability

one half each what is the average result when performing many time the experiment? Certainly not "Green".

In order to circumvent this problem we introduce the concept of *real random variables*. Historically real random variable have been introduced not to mix colors but to model the gain when playing games. A random variable aims to attribute numerical values to events.

Formally, a real random variable  $X$  between the probability space  $(\Omega, \mathcal{A}, \mu)$  and the measurable space  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  (where  $\mathcal{B}(\mathbf{R}^n)$  is the Borel  $\sigma$ -algebra of  $\mathbf{R}^n$  — see next section for a proper definition) is a mapping from  $\Omega$  to  $\mathbf{R}^n$ :

$$\begin{aligned} X : \Omega &\rightarrow \mathbf{R}^n \\ \omega_i &\mapsto X(\omega_i) := x_i . \end{aligned}$$

Since  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$  is measurable, we can define the *probability law*  $\mu_X$  of  $X$  as

$$\begin{aligned} \mu_X : \mathcal{B}(\mathbf{R}^n) &\rightarrow [0; 1] \\ b &\mapsto \mu_X(b) := \mu(X \in b) , \end{aligned}$$

where  $\mu(X \in b) = \mu(X^{-1}(b))$ . The triplet  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mu_X)$  forms a new probability space. When  $n = 1$  we say that  $X$  is a real random variable and when  $n > 1$  we say that  $X$  is a real random vector. For concision, we will work with real random variables and not vectors, apart when stating some results developed for vectors.

At this stage we can forget the origin of our model, in the sense that we do not consider anymore  $(\Omega, \mathcal{A}, \mu)$  but we just focus on  $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mu_X)$  and the central object that we study is the real random variable  $X$  (or vector — we will refer to both as just random variables for conciseness) with its probability distribution  $\mu_X$ . This description allows one to focus on the form of the probability distribution without referring to a specific experiment and its events. Thus the results derived from the study of a particular random variable can be applied to all experiments having the same random variable.

### A.3.1 Continuous random variables

The Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbf{R})$  is a central element in the theory of measures. It is defined as the  $\sigma$ -algebra generated<sup>1</sup> by all the open sets of  $\mathbf{R}$ . Intuitively, it is the set of all the open and closed intervals in  $\mathbf{R}$  and their countable unions. Apart from the probability measure we also define two useful quantities, the cumulative distribution function and the probability density function.

The cumulative distribution function  $f$  corresponds to the probability that the random variable is under a certain threshold

$$f : \mathbf{R} \rightarrow [0, 1] \tag{A.2}$$

$$x \mapsto f(x) := \mu_X(X \in [-\infty, x]) . \tag{A.3}$$

Since  $\mu_X$  is a probability measure we have  $f(-\infty) = 0$  and  $f(+\infty) = 1$ .

The probability density function  $p$  is defined implicitly by the relation

$$\mu_X(X \in [x_1, x_2]) =: \int_{x_1}^{x_2} p(x) \, dx . \tag{A.4}$$

---

<sup>1</sup>A  $\sigma$ -algebra generated by a set of parts  $S$  of  $\Omega$  is the smallest  $\sigma$ -algebra containing  $S$

Since  $\mu_X$  is a probability measure one obtains  $\int_{-\infty}^{\infty} p(x) dx = 1$ . We will assume that the function  $p$  has a "decent" behavior, in the sense that we expect from it to be continuous and derivable as many times as we need.

By setting  $x_1$  to  $-\infty$  in the definition of  $p$ , we go from probability densities to cumulative distribution function:

$$\int_{-\infty}^{x_2} p(x) dx = f(x_2) , \quad (\text{A.5})$$

From the fundamental theorem of calculus it follows the identity

$$p(x) = \frac{d}{dx} f(x) . \quad (\text{A.6})$$

The probability density can be viewed as the probability that  $X$  lies in a small interval around  $x$ :  $p(x) \sim \mu_X(X \in [x, x + dx])/dx$ .

### A.3.2 Discrete random variables

Let us consider a random variable  $X$  defined on a probability space  $(\Omega, \mathcal{A}, \mu)$ .  $X$  is said to be a *discrete random variable* if  $X(\Omega)$  is countable. In such case we do not need to define density probabilities and cumulative functions. Instead we can go back to the intuitive approach since the events are countable.

We often do not need intervals for discrete random variables. Therefore we can define them over the set  $\mathbf{N}$  or some subset of it.

#### Box 18: Tossing a coin

The typical example of tossing a coin can be described by the sample space  $\Omega = \{\text{Head}, \text{Tail}\}$ , while a  $\sigma$ -algebra can be  $\mathcal{A} = \{\Omega, \emptyset, \{\text{Head}\}, \{\text{Tail}\}\} = \{\{\text{Head}, \text{Tail}\}, \{\}, \{\text{Head}\}, \{\text{Tail}\}\}$  and the measure  $\mu$  defined by  $\mu(\Omega) = 1$ ,  $\mu(\emptyset) = 0$ ,  $\mu(\{\text{Head}\}) = p$  and  $\mu(\{\text{Tail}\}) = 1 - p$  with  $0 \leq p \leq 1$ .

A discrete random variable  $X$  could be  $X(\text{Head}) = 1$  and  $X(\text{Tails}) = 0$  giving rise to the probability distribution  $\mu_X$  defined by  $\mu_X(X = 1) = p$  and  $\mu_X(X = 0) = 1 - p$ , which is know as a Bernoulli distribution.

## A.4 Moments

### A.4.1 Expectation value

The first moment of a random variable is its expectation value. For a discrete random variable  $X$  with sample space  $\{x_i\}_{i \in \mathcal{S}}$  with  $\mathcal{S} \subset \mathbf{N}$  and associated probabilities  $p_i$  the expectation value  $E[X]$  is defined as

$$E[X] := \sum_{i \in \mathcal{S}} x_i p_i . \quad (\text{A.7})$$

When the random variable is continuous and has a probability density  $p(x)$  the expectation value is defined as

$$E[X] := \int dx p(x)x . \quad (\text{A.8})$$

In both cases the expectation values obey the following properties:

- (a) For two random variables  $X$  and  $Y$  and two real number  $\alpha$  and  $\beta$  we have  $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$ .
- (b) If  $X$  is constant and equal to  $\alpha$  then  $E[X] = \alpha$ . Especially we have  $E[E[X]] = E[X]$ .

### A.4.2 Variance

The centred second moment of a random variable is its variance. In all generality the variance of a random variable  $X$  is defined as

$$\text{Var}[X] := E[(X - E[X])^2] . \quad (\text{A.9})$$

Using the property of the expectation value we can rewrite the variance as

$$\text{Var}[X] = E[X^2] - E[X]^2 . \quad (\text{A.10})$$

We also define the standard deviation  $\sigma$  as the positive square root of the variance.

### A.4.3 Moment of order $k$

We define the moment of order  $k$ , called  $m_k$ , as

$$m_k(X) = \int dx x^k \mu(x) \quad \text{or} \quad m_k(X) = \sum_{i \in \mathcal{S}} x_i^k p_i . \quad (\text{A.11})$$

We also define the centred moment, called  $c_k$ , as

$$c_k(X) = \int d(x - m_1)^k x^k \mu(x) \quad \text{or} \quad m_k(X) = \sum_{i \in \mathcal{S}} (x_i - m_1)^k p_i , \quad (\text{A.12})$$

where  $m_1$  is the first moment. With this notation we have  $c_2(X) = \text{Var}[X]$ .

### A.4.4 Co-variance

Consider a random vector  $Z$  defined as a pair of random variables  $Z = (X, Y)$ . In the discrete case we associate to  $Z$  a probability distribution  $\mu(x, y)$ . In the continuous case we associate to  $Z$  a probability density  $p(x, y)$ .

The marginal law of  $Y$  is defined as  $\mu_X(y) = \sum_x \mu(x, y)$  for the discrete case and  $p_X(y) = \int dx p(x, y)$  in the continuous case. In the same fashion we define the marginal law of  $X$  by  $\mu_Y(x) = \sum_y \mu(x, y)$  and  $p_Y(x) = \int dy p(x, y)$ .

We say that the two random variables  $X$  and  $Y$  are independent when the probability distribution can be written as the product of the marginal laws:

$$\mu(x, y) = \mu_X(y)\mu_Y(x) \quad \text{or} \quad p(x, y) = p_X(y)p_Y(x) . \quad (\text{A.13})$$

We define the co-variance  $\text{Cov}[X, Y]$  of a pair of random variable  $X, Y$

$$\text{Cov}[X, Y] := \text{E}[(X - \text{E}[X])(Y - \text{E}[Y])] . \quad (\text{A.14})$$

The co-variance is a measure of the linear dependency of two random variables. Especially, *if two random variables are independent then their co-variance vanishes*. The converse is in general not true.

### A.4.5 Function of random variables

#### Single random variable

It is possible to define functions of a random variable. Suppose that  $f : \mathbf{R} \rightarrow \mathbf{R}$  is a measurable function defined at least on  $X(\Omega)$ , where  $X$  is a random variable going from  $\Omega$  to  $\mathbf{R}$ . Then the map  $Y = f \circ X$  is also a random variable and is noted  $f(X)$ . We can also define moments for this random variable using its distribution.

For the first moment, we can give a close expression using only the original distribution  $p(x)$  and not the probability distribution of  $Y$ . Under the condition that the integral converges, we obtain the *transfer formula*:

$$\text{E}[f(X)] = \int_{-\infty}^{+\infty} dt f(t)p(t) . \quad (\text{A.15})$$

When the random variable is discrete we are left with

$$\text{E}[f(X)] = \sum_{x \in X(\Omega)} f(x)\mu(x) . \quad (\text{A.16})$$

For the second moment there is no such transfer formulas in the general case. When  $f$  corresponds to an affine transformation,  $f(X) = aX + b$  with  $a, b \in \mathbf{R}$ , we can calculate explicitly the variance,

$$\text{Var}[f(X)] = a^2 \text{Var}[X] . \quad (\text{A.17})$$

For a general function  $f$ , one has to try to calculate directly the transformed variance.

#### Sum of random variables

Consider two random variables  $X$  and  $Y$  and their sum  $Z = X + Y$ . The variance of  $Z$  is then given by

$$\text{Var}[Z] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y] . \quad (\text{A.18})$$

## A.5 From probabilities to statistics

In the preceding sections we have presented the axioms and basic tools of probability theory. We now turn back to the direction of statistics. For the sake of simplicity we will relax our exigences of formalism. Especially we do not distinguish between discrete and continuous random variables. We use the term *probability distribution* or just *distribution* to refer indifferently to the probability density of a continuous random variable or to the probability of an event for a discrete random variable.

### A.5.1 Bienaymé-Tchebychev inequality and law of large number

**Theorem A.1 — BIENAYMÉ-TCHEBYCHEV INEQUALITY.**

Let  $X$  be a random variable with measure  $\mu$  and define the variance  $\text{Var}[X] = \sigma^2$  and the expectation value  $E[X] = m$ . Then for all  $\varepsilon > 0$ , we have

$$\mu(|X - m| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}. \quad (\text{A.19})$$

This theorem provides an upper bound for the probability that the distance between  $X$  and its mean is greater than  $\varepsilon$ .

**Theorem A.2 — LAW OF LARGE NUMBER.**

Consider a sequence of random variables  $\{X_n\}_{n \in \mathbf{N}}$ , pairwise independent, with the same measure  $\mu$  and thus having the same expectation value  $E[X_n] = m$ . We define a new random variable  $Y_n$  as

$$Y_n = \frac{X_1 + \cdots + X_n}{n}. \quad (\text{A.20})$$

Then we have

$$\mu\left(\lim_{n \rightarrow \infty} Y_n = m\right) = 1. \quad (\text{A.21})$$

Stated otherwise, this theorem asserts that when repeating a large number of times the same random process, we get *in average* the expectation value.

### A.5.2 The central limit theorem

Before stating this famous theorem we should briefly review the normal distribution  $\mathcal{N}(m, \sigma)$ , also known as the Gaussian distribution. It is a continuous distribution and corresponds to the probability density

$$\mu(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/\sigma^2}. \quad (\text{A.22})$$

The calculation of the expectation value and variance of a random variable  $X$  distributed according to  $\mathcal{N}(m, \sigma)$  reads

$$E[X] = m \quad \text{and} \quad \text{Var}[X] = \sigma^2. \quad (\text{A.23})$$

We can now state the central limit theorem:

**Theorem A.3 — CENTRAL LIMIT THEOREM.**

Consider a sequence of random variables  $\{X_n\}_{n \in \mathbf{N}}$ , pairwise independent, with the same measure  $\mu$  and thus having the same expectation value  $E[X_n] = m$  and the same variance  $\text{Var}[X_n] = \sigma^2$ . We define the random variables  $Y_n$  as

$$Y_n = \frac{X_1 + \cdots + X_n}{n}. \quad (\text{A.24})$$

Then the sequence  $\{Y_n\}_{n \in \mathbf{N}}$  converges to the normal law  $\mathcal{N}(m, \sigma)$ , meaning that

$$\mu(Y_n \in [-\infty, x]) \xrightarrow{n \rightarrow \infty} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x dx' e^{-(x'-m)^2/\sigma^2}. \quad (\text{A.25})$$

# Appendix B

## Perturbation theory

### B.1 Interaction picture

Consider the time-independent Hamiltonian  $H$  composed of two parts, the free evolution Hamiltonian  $H_0$  and the perturbation  $V$ :

$$H = H_0 + V . \quad (\text{B.1})$$

Usually the free evolution part represents a known dynamics and one is interested in the effect of  $V$  on the dynamics of the system. In the Schrödinger picture and starting with the initial state  $|\psi_0\rangle$ , the state after evolution during a time  $t$  is found by solving the Schrödinger equation

$$\frac{\partial}{\partial t} |\psi_S(t)\rangle = -iH |\psi_S(t)\rangle . \quad (\text{B.2})$$

The Hamiltonian being time-independent the solution of the equation is found immediately and reads

$$|\psi_S(t)\rangle = U_S(t) |\psi_0\rangle , \quad (\text{B.3})$$

where  $U_S(t) := e^{-iHt}$  is called the evolution operator.

The same dynamics can be represented in the Heisenberg picture. There it is not the states that evolve with time but the operators. Starting at  $t = 0$  with the operator  $A_0$ , the operator after an evolution time  $t$  is given by:

$$A_S(t) = U_S(t) A_0 U_S(t)^\dagger . \quad (\text{B.4})$$

Eventually there is a third possibility to represent a Hamiltonian dynamics, which is especially suited to deal with Hamiltonians of the form (B.1). In this so-called *interaction picture* both states and operators evolve with time,

$$A_I(t) = U_0(t)^\dagger A_0 U_0(t) \quad (\text{B.5})$$

$$|\psi_I(t)\rangle = U_0(t)^\dagger |\psi_S(t)\rangle , \quad (\text{B.6})$$

where  $U_0(t) = e^{iH_0 t}$  is the evolution operator for the free Hamiltonian. The two parts of the Hamiltonian in the interaction picture read  $H_{0,I} = H_0$  and  $V_I(t) = U_0(t) V U_0(t)^\dagger$ . The Schrödinger equation in the interaction picture is known as the Schwinger-Tomonaga equation and reads

$$\frac{\partial}{\partial t} |\psi_I(t)\rangle = -iH_I(t) |\psi_I(t)\rangle , \quad (\text{B.7})$$

where  $H_I(t) := V_I(t)$  is the Hamiltonian in the interaction picture (notice that what we call "Hamiltonian in the interaction picture" is *not*  $U_0(t)^\dagger H U_0(t)$ ). Here lies the interest of the interaction picture: all the dependence on the free Hamiltonian is encoded in the operators. The cost of this simplification is that now we have to solve a Schrödinger equation with a time-dependent Hamiltonian.

## B.2 Dyson series and perturbation theory

The formal solution of the Schwinger-Tomonaga equation, with initial condition  $|\psi_0\rangle$  at  $t = 0$  reads

$$|\psi_I(t)\rangle = |\psi_0\rangle - i \int_0^t dt_1 H_I(t_1) |\psi_I(t_1)\rangle. \quad (\text{B.8})$$

By iterating the solution we obtain the *Dyson series*<sup>1</sup>

$$\begin{aligned} |\psi_I(t)\rangle = & \left( \mathcal{I} - i \int_0^t dt_1 H_I(t_1) - \int_0^t dt_1 \int_0^{t_1} dt_2 H_I(t_1) H_I(t_2) + \dots \right. \\ & \left. + (-i)^n \int_0^t dt_1 \dots \int_0^{t_{n-1}} dt_n H_I(t_1) \dots H_I(t_n) + \dots \right) |\psi_0\rangle. \end{aligned} \quad (\text{B.9})$$

We will now show how to express the state in the interaction picture in an elegant form using the time ordering operator. The time-ordering operator  $\mathcal{T}$  is defined by its effect on a product of two operators:

$$\mathcal{T}[A(t_1)B(t_2)] := \begin{cases} A(t_1)B(t_2) & \text{if } t_1 > t_2, \\ B(t_2)A(t_1) & \text{if } t_2 > t_1. \end{cases} \quad (\text{B.10})$$

For a product of  $n$  operators the time-ordering operators act such that the resulting operator is time-ordered, meaning of the form  $A_1(t_1)A_2(t_2) \dots A_n(t_n)$  with  $t_1 > t_2 > \dots > t_n$ .

The third term in the Dyson series may be rewritten as

$$\int_0^t dt_1 \int_0^{t_1} dt_2 H(t_1)H(t_2) = \int_0^t dt_2 \int_{t_2}^t dt_1 H(t_1)H(t_2). \quad (\text{B.11})$$

Both expressions correspond to the pink area in the square illustrated in Fig. B.1, where  $t_2 < t_1$ . In the first expression we first perform the integration over  $t_2$  and then over  $t_1$ , and in the second expression we first integrate over  $t_1$  and then over  $t_2$ . Relabelling the integration variables in the right hand side, *i.e.* exchanging  $t_1$  and  $t_2$ , we obtain

$$\int_0^t dt_1 \int_0^{t_1} dt_2 H(t_1)H(t_2) = \int_0^t dt_1 \int_{t_1}^t dt_2 H(t_2)H(t_1). \quad (\text{B.12})$$

Now we have  $t_1 < t_2$  and the region of integration corresponds to the green area in Fig. B.1. As before, the Hamiltonian with the smaller time parameter stands on the right hand side of the Hamiltonian with the larger time parameter. Using the left hand side (L) and the right hand side

---

<sup>1</sup>The study of the convergence of Dyson series is actually a non-trivial topic. As we are here just interested in Dyson series for the purpose of perturbation theory we do not need to worry about convergence.

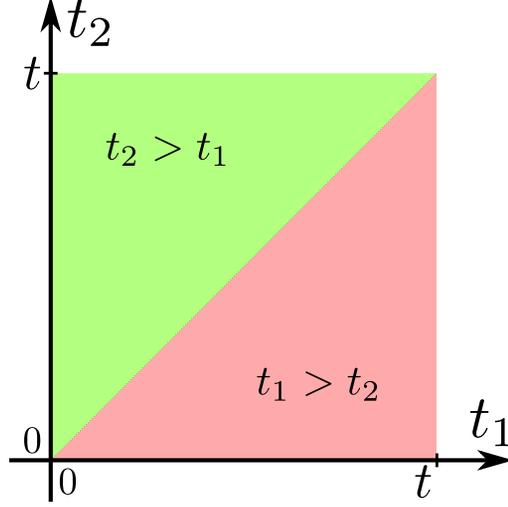


Figure B.1: Integration area for the third term of the Dyson series. The pink area corresponds to  $t_1 > t_2$  while the green area corresponds to  $t_2 > t_1$ .

(R) of the previous equation, and writing the trivial equation  $L = \frac{1}{2}(L + R)$ , we obtain

$$\begin{aligned} \int_0^t dt_1 \int_0^{t_1} dt_2 H(t_1)H(t_2) &= \frac{1}{2} \int_0^t dt_1 \left( \int_0^{t_1} dt_2 H(t_1)H(t_2) + \int_{t_1}^t dt_2 H(t_2)H(t_1) \right) \quad (\text{B.13}) \\ &= \frac{1}{2} \int_0^t dt_1 \int_0^t dt_2 \mathcal{T}[H(t_1)H(t_2)] \\ &= \frac{1}{2} \mathcal{T} \left[ \int_0^t dt_1 \int_0^t dt_2 H(t_1)H(t_2) \right]. \end{aligned}$$

Proceeding in the same way with the remaining terms in the Dyson series, we obtain

$$\begin{aligned} |\psi_I(t)\rangle &= \mathcal{T} \left[ \mathcal{I} - i \int_0^t H_I(t_1) dt_1 - \frac{1}{2} \int_0^t \int_0^{t_1} H_I(t_1)H_I(t_2) dt_1 dt_2 \right. \\ &\quad \left. + i \frac{1}{6} \int_0^t \int_0^{t_1} \int_0^{t_2} H_I(t_1)H_I(t_2)H_I(t'_2) dt_1 dt_2 dt'_2 + \dots \right] |\psi_0\rangle. \quad (\text{B.14}) \end{aligned}$$

Recognizing the Taylor expansion of the exponential function, we finally obtain

$$|\psi_I(t)\rangle = \mathcal{T} \left[ e^{-i \int_0^t H_I(t_1) dt_1} \right] |\psi_0\rangle. \quad (\text{B.15})$$

### B.3 Second order perturbation theory

Here we consider that the Hamiltonian in the interaction picture is "small" (we will see below how we can quantify this). Then we can neglect the terms of high order in the Dyson series. Up to second order we obtain

$$|\psi_I(t)\rangle = \left( \mathcal{I} - i \int_0^t H_I(t_1) dt_1 - \int_0^t dt_1 \int_0^{t_1} dt_2 H_I(t_1)H_I(t_2) + \mathcal{O}(H_I^3) \right) |\psi_0\rangle. \quad (\text{B.16})$$

We can try to set bounds to the condition of validity of the perturbation by looking at the norm of the first order term  $\mathcal{N} = \left\| \int_0^t dt_1 H_I(t_1) \right\|_2$ . The calculation of  $\mathcal{N}^2$  reads

$$\mathcal{N}^2 = \text{tr} \left[ \int_0^t dt_1 \int_0^t dt_2 H_I(t_1) H_I(t_2) \right] = \int_0^t dt_1 \int_0^t dt_2 \text{tr} [H_I(t_1) H_I(t_2)] . \quad (\text{B.17})$$

For the sake of simplicity, and also because we usually assume that the dynamics of the free Hamiltonian is known, we work in the eigenbasis of the free Hamiltonian  $H_0 = \sum_i \lambda_i |u_i\rangle\langle u_i|$ .

We obtain

$$\text{tr} [H_I(t_1) H_I(t_2)] = \sum_{i,j} e^{2i \Delta_{ij} t_1} e^{-2i \Delta_{ij} t_2} |V_{ij}|^2 , \quad (\text{B.18})$$

with  $\Delta_{ij} = (\lambda_i - \lambda_j)/2$ .

Finally we get for  $\mathcal{N}$ :

$$\mathcal{N}^2 = t^2 \sum_i |V_{ii}|^2 + t^2 \sum_{\substack{i,j;i \neq j; \\ \lambda_i = \lambda_j}} |V_{ij}|^2 + \sum_{\substack{i,j;i \neq j; \\ \lambda_i \neq \lambda_j}} \frac{\sin^2(\Delta_{ij} t)}{\Delta_{ij}^2} |V_{ij}|^2 , \quad (\text{B.19})$$

that can also be written as

$$\mathcal{N}^2 = t^2 \sum_{\substack{i,j \\ \lambda_i = \lambda_j}} |V_{ij}|^2 + \sum_{\substack{i,j;i \neq j; \\ \lambda_i \neq \lambda_j}} \frac{\sin^2(\Delta_{ij} t)}{\Delta_{ij}^2} |V_{ij}|^2 , \quad (\text{B.20})$$

As expected we see that the order of magnitude of  $V$  plays a great role. Moreover, the role of the time evolution and of  $H_0$  is made clearer. Although without knowing the Hamiltonians we cannot predict anything in detail, we can make some general statements:

- $H_0$  plays a role in the conditions of validity not through the norm of it but via the spacing  $\Delta_{ij}$  of its eigenvalues.
- The degeneracy in the spectrum of  $H_0$  plays also a role. If we have a highly symmetric problem leading to a high degree of degeneracy, the second term in the sum will not be negligible in comparison to the others.
- There is a linear scaling in  $t$  for all terms where  $\Delta_{ij}$  is null. Again depending on the degree of degeneracy this will prove more or less important in the analysis of the conditions of validity.

# Appendix C

## Perturbative result for coherent averaging

### C.1 QFI in perturbation theory II

In the Sec. 3.7.2 we considered the Hamiltonian  $H(\lambda) = H_0(\lambda) + V$ . In the interaction picture we have

$$\begin{aligned} H_I(t) &= V_I = e^{iH_0(\lambda)t} V e^{-iH_0(\lambda)t} , \\ |\psi_I(t)\rangle &= e^{iH_0(\lambda)t} |\psi_S(t, \lambda)\rangle , \end{aligned}$$

where for the sake of concision we did not make the dependence on  $\lambda$  in the state or in the Hamiltonian explicit. We showed that the QFI for  $\lambda$  is given by

$$I_\lambda = 4 (I_{\lambda,1} + tI_{\lambda,t} + t^2 I_{\lambda,t^2}) ,$$

where

$$\begin{aligned} I_{\lambda,1} &= \langle \dot{\psi}_I(t, \lambda) | \dot{\psi}_I(t, \lambda) \rangle - |\langle \psi_I(t, \lambda) | \dot{\psi}_I(t, \lambda) \rangle|^2 , \\ I_{\lambda,t} &= 2\Im \left\{ \langle \dot{\psi}_I(t, \lambda) | \dot{H}_0(\lambda) | \dot{\psi}_I(t, \lambda) \rangle \right\} - 2i \langle \psi_I(t, \lambda) | \dot{H}_0(\lambda) | \psi_I(t, \lambda) \rangle \\ &\quad \times \Im \left\{ \langle \psi_I(t, \lambda) | \dot{\psi}_I(t, \lambda) \rangle \right\} \\ I_{\lambda,t^2} &= \langle \psi_I(t, \lambda) | (\dot{H}_0(\lambda))^2 | \psi_I(t, \lambda) \rangle - \langle \psi_I(t, \lambda) | \dot{H}_0(\lambda) | \psi_I(t, \lambda) \rangle^2 . \end{aligned}$$

For the sake of concision we denote the derivative with respect to  $\lambda$  by a dot. Under the assumption that

$$[H_0(\lambda), \dot{H}_0(\lambda)] = [H_0(\lambda), \ddot{H}_0(\lambda)] = [\dot{H}_0(\lambda), \ddot{H}_0(\lambda)] = 0 ,$$

we find up to second order in perturbation:

$$I_{\lambda,1}^{V_I} = \int_0^t \int_0^t dt_1 dt_2 \left( \langle \dot{H}_I(t_1) \dot{H}_I(t_2) \rangle - \langle \dot{H}_I(t_1) \rangle \langle \dot{H}_I(t_2) \rangle \right) \quad (\text{C.1})$$

$$\begin{aligned}
 I_{\lambda,t}^{V_1} = & -2i \int_0^t dt_1 \langle \dot{H}_I(t_1) \rangle \int_0^t dt_2 \langle [H_I(t_2), \dot{H}_0] \rangle - 2 \langle \dot{H}_0 \rangle \int_0^t \langle \dot{H}_I(t_1) \rangle + 2\Re \left( \int_0^t dt_1 \langle \dot{H}_0 \dot{H}_I(t_1) \rangle \right) \\
 & - 2i \langle \dot{H}_0 \rangle \int_0^t \int_0^{t_1} dt_1 dt_2 \langle [H_I(t_2), \dot{H}_I(t_1)] \rangle + 2\Im \left( \int_0^t \int_0^t dt_1 dt_2 \langle H_I(t_1) \dot{H}_0 \dot{H}_I(t_2) \rangle \right) \\
 & - 2\Im \left( \int_0^t \int_0^{t_1} dt_1 dt_2 \langle \dot{H}_0 \dot{H}_I(t_1) H_I(t_2) + \dot{H}_0 H_I(t_1) \dot{H}_I(t_2) \rangle \right) \tag{C.2}
 \end{aligned}$$

$$\begin{aligned}
 I_{\lambda,t^2}^{V_1} = & \left( \langle \dot{H}_0^2 \rangle - \langle \dot{H}_0 \rangle^2 \right) - 2 \langle \dot{H}_0 \rangle \int_0^t \int_0^t dt_1 dt_2 \langle H_I(t_1) \dot{H}_0 H_I(t_2) \rangle + \left( \int_0^t dt_1 \langle [H_I(t_1), \dot{H}_0] \rangle \right)^2 \\
 & + 2 \langle \dot{H}_0 \rangle \int_0^t \int_0^{t_1} dt_1 dt_2 \langle H_I(t_2) H_I(t_1) \dot{H}_0 + \dot{H}_0 H_I(t_1) H_I(t_2) \rangle + i \int_0^t dt_1 \langle [H_I(t_1), \dot{H}_0^2] \rangle \\
 & + \int_0^t \int_0^t dt_1 dt_2 \langle H_I(t_1) \dot{H}_0^2 H_I(t_2) \rangle - \int_0^t \int_0^{t_1} dt_1 dt_2 \langle \dot{H}_0^2 H_I(t_1) H_I(t_2) + H_I(t_2) H_I(t_1) \dot{H}_0^2 \rangle \\
 & - 2i \langle \dot{H}_0 \rangle \int_0^t dt_1 \langle [H_I(t_1), \dot{H}_0] \rangle \tag{C.3}
 \end{aligned}$$

## C.2 QFI in PT II for the coherent averaging Hamiltonian

If we inject the structure of coherent averaging we can further develop the term of order one and two in PT II.

For the Eq. (C.1) we have:

$$\begin{aligned}
 \int_0^t \int_0^t dt_1 dt_2 \left( \langle \dot{H}_I(t_1) \dot{H}_I(t_2) \rangle - \langle \dot{H}_I(t_1) \rangle \langle \dot{H}_I(t_2) \rangle \right) = & - \int_0^t \int_0^t dt_1 dt_2 \sum_{\nu,\mu} \left\{ N \langle R_\nu(t_1) R_\mu(t_2) \rangle \right. \\
 & \times \left( \langle [\dot{H}_I, S_\nu(t_1)] [ \dot{H}_I, S_\mu(t_2) ] \rangle - \langle [\dot{H}_I, S_\nu(t_1)] \rangle \langle [\dot{H}_I, S_\mu(t_2)] \rangle \right) \\
 & \left. + N^2 \left( \langle R_\nu(t_1) R_\mu(t_2) \rangle - \langle R_\nu(t_1) \rangle \langle R_\mu(t_2) \rangle \right) \langle [\dot{H}_I, S_\nu(t_1)] \rangle \langle [\dot{H}_I, S_\mu(t_2)] \rangle \right\}
 \end{aligned}$$

For the Eq. (C.2) we have:

$$\begin{aligned}
 -2i t \int_0^t dt_1 \langle \dot{H}_I(t_1) \rangle \int_0^t dt_2 \langle [H_I(t_2), \dot{H}_0] \rangle = & \\
 & 2i t \int_0^t \int_0^t dt_1 dt_2 \sum_{\nu,\mu} N^2 \langle R_\nu(t_1) \rangle \langle R_\mu(t_2) \rangle \langle [\dot{H}_I, S_\nu(t_1)] \rangle \langle [S_\mu(t_2), \dot{H}_I] \rangle \\
 -2t \langle \dot{H}_0 \rangle \int_0^t \langle \dot{H}_I(t_1) \rangle = & -2i t \int_0^t dt_1 t_1 N^2 \sum_\nu \langle \dot{H}_I \rangle \langle R_\nu(t_1) \rangle \langle [\dot{H}_I, S_\nu(t_1)] \rangle
 \end{aligned}$$

$$\begin{aligned}
 & 2t\Re\left(\int_0^t dt_1\langle\dot{H}_0\dot{H}_I(t_1)\rangle\right) = 2t\int_0^t dt_1t_1\sum_\nu\langle R_\nu(t_1)\rangle\left\{iN^2\langle\dot{H}_I\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle\right. \\
 & \quad \left.-N\left(\Im\left(\langle\dot{H}_I[\dot{H}_I,S_\nu(t_1)]\rangle\right)-i\langle\dot{H}_I\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle\right)\right\} \\
 & -2it\langle\dot{H}_0\rangle\int_0^t\int_0^{t_1} dt_1dt_2\langle[H_I(t_2),\dot{H}_I(t_1)]\rangle = \\
 & 2t\langle\dot{H}_0\rangle\int_0^t\int_0^{t_1} dt_1dt_2t_1\sum_{\nu,\mu}\langle\dot{H}_I\rangle\left\{N^3\langle[R_\mu(t_2),R_\nu(t_1)]\rangle\langle S_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle\right. \\
 & \quad +N^2\left(\langle S_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle\langle[R_\nu(t_1),R_\mu(t_2)]\rangle+\langle R_\mu(t_2)R_\nu(t_1)\rangle\langle S_\mu(t_2)[\dot{H}_I,S_\nu(t_1)]\rangle\right. \\
 & \quad \left.-\langle R_\nu(t_1)R_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]S_\mu(t_2)\rangle\right\} \\
 & 2t\Im\left(\int_0^t\int_0^t dt_1dt_2\langle H_I(t_1)\dot{H}_0\dot{H}_I(t_2)\rangle\right) = 2t\Im\int_0^t\int_0^t dt_1dt_2\sum_{\nu,\mu}i t_2\langle R_\nu(t_1)R_\mu(t_2)\rangle \\
 & \quad \left\{N\left(\langle S_\nu(t_1)\dot{H}_I[\dot{H}_I,S_\mu(t_2)]\rangle-\langle S_\nu(t_1)\dot{H}_I\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle-\langle S_\nu(t_1)\rangle\langle\dot{H}_I[\dot{H}_I,S_\mu(t_2)]\rangle\right. \right. \\
 & \quad \left.-\langle\dot{H}_I\rangle\langle S_\nu(t_1)[\dot{H}_I,S_\mu(t_2)]\rangle+2\langle S_\nu(t_1)\rangle\langle\dot{H}_I\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle\right) \\
 & \quad +N^2\left(\langle S_\nu(t_1)\dot{H}_I\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle+\langle S_\nu(t_1)\rangle\langle\dot{H}_I[\dot{H}_I,S_\mu(t_2)]\rangle+\langle\dot{H}_I\rangle\langle S_\nu(t_1)[\dot{H}_I,S_\mu(t_2)]\rangle\right. \\
 & \quad \left.-3\langle S_\nu(t_1)\rangle\langle\dot{H}_I\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle+N^3\langle S_\nu(t_1)\rangle\langle\dot{H}_I\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle\right\} \\
 & -2t\Im\left(\int_0^t\int_0^{t_1} dt_1dt_2\langle\dot{H}_0\dot{H}_I(t_1)H_I(t_2)+\dot{H}_0H_I(t_1)\dot{H}_I(t_2)\rangle\right) = \\
 & \quad -2t\Im\int_0^t\int_0^{t_1} dt_1dt_2\sum_{\mu,\nu}\left(\langle R_\nu(t_1)R_\mu(t_2)\rangle\left\{N\left(t_1\left[\langle\dot{H}_I[\dot{H}_I,S_\nu(t_1)]S_\mu(t_2)\rangle\right.\right.\right. \right. \\
 & \quad \left.-\langle\dot{H}_I S_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle-\langle\dot{H}_I\rangle\langle[\dot{H}_I,S_\nu(t_1)]S_\mu(t_2)\rangle\right. \\
 & \quad \left.-\langle S_\mu(t_2)\rangle\langle\dot{H}_I[\dot{H}_I,S_\nu(t_1)]\rangle+2\langle\dot{H}_I\rangle\langle S_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle\right) \\
 & \quad \left.+t_2\left[\langle\dot{H}_I S_\nu(t_1)[\dot{H}_I,S_\mu(t_2)]\rangle-\langle\dot{H}_I S_\nu(t_1)\rangle-\left(\langle\dot{H}_I\rangle\langle S_\nu(t_1)[\dot{H}_I,S_\mu(t_2)]\rangle\right.\right. \\
 & \quad \left.-\langle S_\nu(t_1)\rangle\langle\dot{H}_I[\dot{H}_I,S_\mu(t_2)]\rangle+2\langle\dot{H}_I\rangle\langle S_\nu(t_1)\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle\right) \\
 & \quad \left.+N^2\left(t_1\left[\langle\dot{H}_I\rangle\langle[\dot{H}_I,S_\nu(t_1)]S_\mu(t_2)\rangle+\langle S_\mu(t_2)\rangle\langle\dot{H}_I[\dot{H}_I,S_\nu(t_1)]\rangle\right.\right. \\
 & \quad \left.+ \langle\dot{H}_I S_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle-3\langle\dot{H}_I\rangle\langle S_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle\right) \\
 & \quad \left.+t_2\left[\langle S_\nu(t_1)\rangle\langle\dot{H}_I[\dot{H}_I,S_\mu(t_2)]\rangle+\langle\dot{H}_I\rangle\langle S_\nu(t_1)[\dot{H}_I,S_\mu(t_2)]\rangle\right. \\
 & \quad \left.+ \langle\dot{H}_I S_\nu(t_1)\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle-3\langle\dot{H}_I\rangle\langle S_\nu(t_1)\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle\right) \\
 & \quad \left.+N^3\left(t_1\langle\dot{H}_I\rangle\langle S_\mu(t_2)\rangle\langle[\dot{H}_I,S_\nu(t_1)]\rangle+t_2\langle\dot{H}_I\rangle\langle S_\nu(t_1)\rangle\langle[\dot{H}_I,S_\mu(t_2)]\rangle\right)\right\}
 \end{aligned}$$

For the Eq. (C.3) we have:

$$\begin{aligned}
 & t^2 \left( \langle \dot{H}_0^2 \rangle - \langle \dot{H}_0 \rangle^2 \right) = t^2 N \left( \langle \dot{H}_I^2 \rangle - \langle \dot{H}_I \rangle^2 \right) \\
 & \left( t \int_0^t dt_1 \langle [H_I(t_1), \dot{H}_0] \rangle \right)^2 = t^2 \int_0^t \int_0^t dt_1 dt_2 \sum_{\mu, \nu} N^2 \langle R_\nu(t_1) \\
 & \qquad \qquad \qquad \times \langle R_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \langle [S_\mu(t_2), \dot{H}_I] \rangle \\
 & 2t^2 \langle \dot{H}_0 \rangle \left( \int_0^t \int_0^{t_1} dt_1 dt_2 \langle H_I(t_2) H_I(t_1) \dot{H}_0 + \dot{H}_0 H_I(t_1) H_I(t_2) \rangle \right. \\
 & \left. - \int_0^t \int_0^t dt_1 dt_2 \langle H_I(t_1) \dot{H}_0 H_I(t_2) \rangle \right) = -2t^2 \int_0^t \int_0^{t_1} dt_1 dt_2 \sum_{\nu, \mu} \left\{ N^2 \langle \dot{H}_I \rangle \right. \\
 & \left( \langle R_\nu(t_1) R_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] S_\mu(t_2) \rangle - \langle R_\mu(t_2) R_\nu(t_1) \rangle \langle S_\mu(t_2) [S_\nu(t_1), \dot{H}_I] \rangle \right. \\
 & \left. - \langle [R_\mu(t_2), R_\nu(t_1)] \rangle \langle S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right) \\
 & \left. + N^3 \langle \dot{H}_I \rangle \langle [R_\mu(t_2), R_\nu(t_1)] \rangle \langle S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right. \\
 & \left. + i t^2 \int_0^t dt_1 \langle [H_I(t_1), \dot{H}_0^2] \rangle = i t^2 \int_0^t dt_1 \sum_\nu \langle R_\nu(t_1) \rangle \left\{ N \left( \langle [S_\nu(t_1), \dot{H}_I^2] \rangle \right. \right. \right. \\
 & \left. \left. - 2 \langle \dot{H}_I \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right) + 2N^2 \langle \dot{H}_I \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right\} \\
 & t^2 \left( \int_0^t \int_0^t dt_1 dt_2 \langle H_I(t_1) \dot{H}_0^2 H_I(t_2) \rangle - \int_0^t \int_0^{t_1} dt_1 dt_2 \langle \dot{H}_0^2 H_I(t_1) H_I(t_2) \right. \\
 & \left. + H_I(t_2) H_I(t_1) \dot{H}_0^2 \rangle \right) = t^2 \int_0^t \int_0^t dt_1 dt_2 \sum_{\nu, \mu} \left\{ N \left[ \langle R_\nu(t_1) R_\mu(t_2) \rangle \left( \langle [S_\nu(t_1), \dot{H}_I^2] S_\mu(t_2) \rangle \right. \right. \right. \\
 & \left. - 2 \langle \dot{H}_I \rangle \langle [S_\nu(t_1), \dot{H}_I] S_\mu(t_2) \rangle - \langle S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I^2] \rangle - 2 \langle \dot{H}_I S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right) \\
 & \left. + 4 \langle \dot{H}_I \rangle \langle S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right] + \langle R_\mu(t_2) R_\nu(t_1) \rangle \left( \langle S_\mu(t_2) [S_\nu(t_1), \dot{H}_I^2] \rangle \right. \\
 & \left. - 2 \langle \dot{H}_I \rangle \langle S_\mu(t_2) [S_\nu(t_1), \dot{H}_I] \rangle - \langle S_\mu(t_2) \rangle \langle [\dot{H}_I^2, S_\nu(t_1)] \rangle - 2 \langle S_\mu(t_2) \dot{H}_I \rangle \langle [\dot{H}_I, S_\nu(t_1)] \rangle \right) \\
 & \left. + 4 \langle \dot{H}_I \rangle \langle S_\mu(t_2) \rangle \langle [\dot{H}_I, S_\nu(t_1)] \rangle \right] \\
 & \left. + N^2 \left[ \langle R_\nu(t_1) R_\mu(t_2) \rangle \left( 2 \langle \dot{H}_I \rangle \langle [S_\nu(t_1), \dot{H}_I] S_\mu(t_2) \rangle + \langle S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I^2] \rangle \right) \right. \right. \\
 & \left. + 2 \langle \dot{H}_I S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle - 6 \langle \dot{H}_I \rangle \langle S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right) \\
 & \left. + \langle R_\mu(t_2) R_\nu(t_1) \rangle \left( 2 \langle \dot{H}_I \rangle \langle S_\mu(t_2) [S_\nu(t_1), \dot{H}_I] \rangle + \langle S_\mu(t_2) \rangle \langle [\dot{H}_I^2, S_\nu(t_1)] \rangle \right) \right. \\
 & \left. + 2 \langle S_\mu(t_2) \dot{H}_I \rangle \langle [\dot{H}_I, S_\nu(t_1)] \rangle - 6 \langle \dot{H}_I \rangle \langle S_\mu(t_2) \rangle \langle [\dot{H}_I, S_\nu(t_1)] \rangle \right) \\
 & \left. + N^3 \langle [R_\nu(t_1), R_\mu(t_2)] \rangle 2 \langle \dot{H}_I \rangle \langle S_\mu(t_2) \rangle \langle [S_\nu(t_1), \dot{H}_I] \rangle \right\}
 \end{aligned}$$

$$-2i t^2 \langle \dot{H}_0 \rangle \int_0^t dt_1 \langle [H_I(t_1), \dot{H}_0] \rangle = -2i t^2 \int_0^t dt_1 \sum_{\nu} N^2 \langle R_{\nu}(t_1) \rangle \langle \dot{H}_I \rangle \langle [S_{\nu}(t_1), \dot{H}_I] \rangle$$

### C.3 SNR in PT I for the coherent averaging Hamiltonian

Let us take an observable acting only on the bus:  $A^{(0)} = I \otimes I \otimes \dots \otimes I \otimes A$ . The SNR for  $x$  is equal to

$$S_x^{A^{(0)}} = \frac{|\partial_x \langle \psi(t) | A^{(0)} | \psi(t) \rangle|^2}{\langle \psi(t) | A^{(0)2} | \psi(t) \rangle - \langle \psi(t) | A^{(0)} | \psi(t) \rangle^2}, \quad (\text{C.4})$$

In perturbation theory to the second order the denominator is equal to

$$\begin{aligned} \langle \psi(t) | A^{(0)2} | \psi(t) \rangle - \langle \psi(t) | A^{(0)} | \psi(t) \rangle^2 &= (\langle A^2 \rangle - \langle A \rangle^2) + i\varepsilon \int_0^t dt_1 N \langle S_{\nu}(t_1) \rangle \langle [R_{\nu}(t_1), B] \rangle \\ &+ \varepsilon^2 \sum_{\nu, \mu} \int_0^t \int_0^{t_1} dt_1 dt_2 \left\{ \left( N(N-1) \langle S_{\nu}(t_1) \rangle \langle S_{\mu}(t_2) \rangle + N \langle S_{\nu}(t_1) S_{\mu}(t_2) \rangle \right) \right. \\ &\times \langle [R_{\nu}(t_1), B] R_{\mu}(t_2) \rangle + \left( N(N-1) \langle S_{\nu}(t_1) \rangle \langle S_{\mu}(t_2) \rangle + N \langle S_{\mu}(t_2) S_{\nu}(t_1) \rangle \right) \\ &\left. \times \langle R_{\mu}(t_2) [B, R_{\nu}(t_1)] \rangle \right\} + \varepsilon^2 \sum_{\nu, \mu} \int_0^t \int_0^t dt_1 dt_2 N^2 \langle S_{\nu}(t_1) \rangle \langle S_{\mu}(t_2) \rangle \\ &\times \langle [R_{\nu}(t_1), A] \rangle \langle [R_{\mu}(t_2), A] \rangle + \mathcal{O}(\varepsilon^3) \quad (\text{C.5}) \end{aligned}$$

where  $B = A^2 - 2\langle A \rangle A$ . The expectation values for  $A$ ,  $B$ , and  $R_{\mu}(t)$  are taken with respect to  $|\xi\rangle$  and the expectation value for  $S_{\mu}(t) := S_{\mu}(x, t)$  are taken with respect to  $|\varphi\rangle$ .

The derivative of the expectation value of  $A^{(0)}$  is given by

$$\begin{aligned} \frac{\partial}{\partial x} \langle \psi(t) | A^{(0)} | \psi(t) \rangle &= \frac{\partial}{\partial \theta} \left[ i\varepsilon \int_0^t dt_1 N \langle S_{\nu}(t_1) \rangle \langle [R_{\nu}(t_1), A] \rangle \right. \\ &+ \varepsilon^2 \sum_{\nu, \mu} \int_0^t \int_0^{t_1} dt_1 dt_2 \left\{ \left( N(N-1) \langle S_{\nu}(t_1) \rangle \langle S_{\mu}(t_2) \rangle + N \langle S_{\nu}(t_1) S_{\mu}(t_2) \rangle \right) \right. \\ &\times \langle [R_{\nu}(t_1), A] R_{\mu}(t_2) \rangle + \left( N(N-1) \langle S_{\nu}(t_1) \rangle \langle S_{\mu}(t_2) \rangle + N \langle S_{\mu}(t_2) S_{\nu}(t_1) \rangle \right) \\ &\left. \left. \times \langle R_{\mu}(t_2) [A, R_{\nu}(t_1)] \rangle \right\} \right] + \mathcal{O}(\varepsilon^3). \quad (\text{C.6}) \end{aligned}$$

where the expectation values of the operators (or product of operators)  $S_{\eta}$  (resp.  $R_{\eta}$  and  $A_{(0)}$ ) are taken to respect of the states  $|\varphi\rangle$  (resp.  $|\xi\rangle$ ).

In looking at those formulas one can see that it is possible to reach an HL scaling with almost any observable, *at least in the range of validity of the perturbation theory*.



## Appendix D

# Diagonalization of the matrix $K$ and $G$ for channel estimation<sup>1</sup>

In order to calculate the QFI we need to diagonalize the density matrix. For the states in which we are interested, there are two matrices  $K^{(m)}(a, b, c)$  and  $G^{(m)}(a)$  that recurrently appear in the block decomposition of the states:

- The  $m \times m$  matrix  $K^{(m)}(a, b, c)$ :

$$K^{(m)}(a, b, c) = \begin{pmatrix} a & \cdots & a & b \\ \vdots & \ddots & \vdots & \vdots \\ a & \cdots & a & b \\ b & \cdots & b & c \end{pmatrix}. \quad (\text{D.1})$$

This matrix has rank two, the two eigenvalues

$$\lambda_{\pm}^{(K)} = \frac{1}{2} \left( c + a(m-1) \pm \sqrt{(c - a(m-1))^2 + 4b^2(m-1)} \right), \quad (\text{D.2})$$

and the two corresponding non-normalized eigenvectors

$$\mathbf{v}_{\pm}^{(K)} = (2b, \dots, 2b, Y_{\pm}^{(K)}), \quad (\text{D.3})$$

with  $Y_{\pm}^{(K)} = c - a(m-1) \pm \sqrt{(c - a(m-1))^2 + 4b^2(m-1)}$ .

- The  $m \times m$  matrix  $G^{(m)}(a)$ :

$$G^{(m)}(a) = \begin{pmatrix} a & \cdots & a \\ \vdots & \ddots & \vdots \\ a & \cdots & a \end{pmatrix}, \quad (\text{D.4})$$

---

<sup>1</sup>This appendix is based on: "Quantum channel-estimation with particle loss: GHZ versus W states", Fraïsse, J. M. E. and Braun, D. (2017), *Quantum Measurements and Quantum Metrology*, 3(1) available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.

which only non-zero eigenvalue is

$$\lambda^{(G,m)} = ma, \tag{D.5}$$

and the non-normalized corresponding eigenvector is

$$\mathbf{v}^{(G,m)}(a) = (1, \dots, 1). \tag{D.6}$$

# Bibliography

- Abbott, B. P. et al. (2009). Ligo: the laser interferometer gravitational-wave observatory. *Reports on Progress in Physics*, 72(7):076901.
- Abbott, B. P. et al. (2016a). Gw151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Physical Review Letters*, 116:241103.
- Abbott, B. P. et al. (2016b). Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116:061102.
- Abbott, B. P. et al. (2016c). Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Reviews in Relativity*, 19(1):1.
- Acernese, F. et al. (2015). The Advanced Virgo detector. *Journal of Physics: Conference Series*, 610(1):012014.
- Acín, A. (2001). Statistical distinguishability between unitary operations. *Physical Review Letters*, 87:177901.
- Acín, A., Jané, E., and Vidal, G. (2001). Optimal estimation of quantum dynamics. *Physical Review A*, 64(5):050302.
- Agarwal, G. S. (1981). Relation between atomic coherent-state representation, state multipoles, and generalized phase-space distributions. *Physical Review A*, 24(6):2889–2896.
- Aguiar, O. D. (2011). Past, present and future of the resonant-mass gravitational wave detectors. *Research in Astronomy and Astrophysics*, 11(1):1.
- Amari, S.-I. and Nagaoka, H. (2007). *Methods of Information Geometry*. American Mathematical Society, Oxford.
- Arrad, G., Vinkler, Y., Aharonov, D., and Retzker, A. (2014). Increasing Sensing Resolution with Error Correction. *Physical Review Letters*, 112(15):150801.
- Bachor, H.-A. (2004). *A Guide to Experiments in Quantum Optics*. Wiley VCH, Weinheim, 2nd, revised and enlarged edition edition.
- Barndorff-Nielsen, O. E. and Gill, R. D. (2000). Fisher information in quantum statistics. *Journal of Physics A: Mathematical and General*, 33(24):4481.
- Beenakker, C. and Schönenberger, C. (2003). Quantum Shot Noise. *Physics Today*, 56(5):37–42.

- Beltrán, J. and Luis, A. (2005). Breaking the Heisenberg limit with inefficient detectors. *Physical Review A*, 72(4):045801.
- Bengtsson, I. and Życzkowski, K. (2006). *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press.
- Bennett, C. H. and Wiesner, S. J. (1992). Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884.
- Bergmann, P. G. (1957). Summary of the Chapel Hill Conference. *Reviews of Modern Physics*, 29(3):352–354.
- Beth Ruskai, M., Szarek, S., and Werner, E. (2002). An analysis of completely-positive trace-preserving maps on  $M_2$ . *Linear Algebra and its Applications*, 347(1):159–187.
- Bocko, M. F. and Onofrio, R. (1996). On the measurement of a weak classical force coupled to a harmonic oscillator: experimental progress. *Reviews of Modern Physics*, 68(3):755–799.
- Boixo, S., Flammia, S. T., Caves, C. M., and Geremia, J. (2007). Generalized Limits for Single-Parameter Quantum Estimation. *Physical Review Letters*, 98(9):090401.
- Boixo, S. and Somma, R. D. (2008). Parameter estimation with mixed-state quantum computation. *Physical Review A*, 77(5):052320.
- Bollinger, J. J., Itano, W. M., Wineland, D. J., and Heinzen, D. J. (1996). Optimal frequency measurements with maximally correlated states. *Physical Review A*, 54(6):4649–4652.
- Bondi, H. (1957). Plane Gravitational Waves in General Relativity. *Nature*, 179(4569):1072–1073.
- Bondurant, R. S. and Shapiro, J. H. (1984). Squeezed states in phase-sensing interferometers. *Physical Review D*, 30(12):2548–2556.
- Braginsky, V. B. (1967). Classical and quantum restrictions on the detection of weak disturbances of a macroscopic oscillator. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 53:1434–1441.
- Braginsky, V. B. (1970). *Physical Experiments with Test Bodies*. Nauka, Moscow.
- Braginsky, V. B. and Khalili, F. Y. (1996). Quantum nondemolition measurements: the route from toys to tools. *Reviews of Modern Physics*, 68(1):1–11.
- Braginsky, V. B., Khalili, F. Y., and Thorne, K. S. (1992). *Quantum Measurement*. Cambridge University Press.
- Braginsky, V. B., Khalili, F. Y., and Vorontsov, Y. (1977). Quantum singularities of a ponderomotive meter of electromagnetic energy. *Journal of Experimental and Theoretical Physics*, 46:705.
- Braginsky, V. B., Mitrofanov, V. P., and Panov, V. I. (1985). *Systems with Small Dissipation*. University of Chicago Press.
- Braginsky, V. B. and Vorontsov, Y. (1974). Quantum-mechanical limitations in macroscopic experiments and modern experimental technique. *Uspekhi Fizicheskikh Nauk*, 114:41–53.

- Braginsky, V. B., Vorontsov, Y. I., and Thorne, K. S. (1980). Quantum Nondemolition Measurements. *Science*, 209(4456):547–557.
- Brask, J. B., Chaves, R., and Kołodyński, J. (2015). Improved Quantum Magnetometry beyond the Standard Quantum Limit. *Physical Review X*, 5(3):031010.
- Braun, D. (2011). Ultimate quantum bounds on mass measurements with a nano-mechanical resonator. *EPL (Europhysics Letters)*, 94(6):68007.
- Braun, D., Adesso, G., Benatti, F., Floreanini, R., Marzolino, U., Mitchell, M. W., and Pirandola, S. (2017). Quantum enhanced measurements without entanglement. *ArXiv e-prints*. arXiv:1701.05152.
- Braun, D., Giraud, O., Nechita, I., Pellegrini, C., and Žnidarič, M. (2014a). A universal set of qubit quantum channels. *Journal of Physics A: Mathematical and Theoretical*, 47(13):135302.
- Braun, D., Jian, P., Pinel, O., and Treps, N. (2014b). Precision measurements with photon-subtracted or photon-added Gaussian states. *Physical Review A*, 90(1):013821.
- Braun, D. and Martin, J. (2011). Heisenberg-limited sensitivity with decoherence-enhanced measurements. *Nature Communications*, 2:223.
- Braunstein, S. L. and Caves, C. M. (1994). Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439–3443.
- Busch, P., Lahti, P., and Werner, R. F. (2013). Proof of heisenberg’s error-disturbance relation. *Phys. Rev. Lett.*, 111:160405.
- Cable, H., Gu, M., and Modi, K. (2016). Power of one bit of quantum information in quantum metrology. *Physical Review A*, 93(4):040304.
- Caves, C. M. (1980). Quantum-Mechanical Radiation-Pressure Fluctuations in an Interferometer. *Physical Review Letters*, 45(2):75–79.
- Caves, C. M. (1981). Quantum-mechanical noise in an interferometer. *Physical Review D*, 23(8):1693–1708.
- Caves, C. M. (1985). Defense of the Standard Quantum Limit for Free-Mass Position. *Physical Review Letters*, 54(23):2465–2468.
- Caves, C. M., Thorne, K. S., Drever, R. W. P., Sandberg, V. D., and Zimmermann, M. (1980). On the measurement of a weak classical force coupled to a quantum-mechanical oscillator. I. Issues of principle. *Reviews of Modern Physics*, 52(2):341–392.
- Chapeau-Blondeau, F. (2015). Optimized probing states for qubit phase estimation with general quantum noise. *Physical Review A*, 91(5):052310.
- Chaves, R., Brask, J. B., Markiewicz, M., Kołodyński, J., and Acín, A. (2013). Noisy Metrology beyond the Standard Quantum Limit. *Physical Review Letters*, 111(12):120401.
- Chin, A. W., Huelga, S. F., and Plenio, M. B. (2012). Quantum Metrology in Non-Markovian Environments. *Physical Review Letters*, 109(23):233601.

- Choi, M.-D. (1975). Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290.
- Collins, D. (2013). Mixed-state Pauli-channel parameter estimation. *Physical Review A*, 87(3):032301.
- Collins, D. and Stephens, J. (2015). Depolarizing-channel parameter estimation using noisy initial states. *Physical Review A*, 92(3):032324.
- Cooperstock, F. I. (2015). Gravitational waves, energy and Feynman’s “sticky bead”. *Modern Physics Letters A*, 30(27):1550143.
- Cramér, H. (1946). *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press.
- D’Ariano, G. M., Lo Presti, P., and Paris, M. G. A. (2001). Using Entanglement Improves the Precision of Quantum Measurements. *Physical Review Letters*, 87(27):270404.
- Datta, A., Flammia, S. T., and Caves, C. M. (2005). Entanglement and the power of one qubit. *Physical Review A*, 72(4):042316.
- Datta, A. and Shaji, A. (2012). Quantum metrology without quantum entanglement. *Modern Physics Letters B*, 26(18):1230010.
- Datta, A., Shaji, A., and Caves, C. M. (2008). Quantum Discord and the Power of One Qubit. *Physical Review Letters*, 100(5):050502.
- De Pasquale, A., Rossini, D., Facchi, P., and Giovannetti, V. (2013). Quantum parameter estimation affected by unitary disturbance. *Physical Review A*, 88(5):052117.
- de Pillis, J. K. (1967). Linear transformations which preserve hermitian and positive semidefinite operators. *Pacific Journal of Mathematics*, 23(1):129–137.
- Demkowicz-Dobrzański, R., Banaszek, K., and Schnabel, R. (2013). Fundamental quantum interferometry bound for the squeezed-light-enhanced gravitational wave detector geo 600. *Physical Review A*, 88:041802.
- Demkowicz-Dobrzański, R., Jarzyna, M., and Kołodyński, J. (2015). Chapter Four - Quantum Limits in Optical Interferometry. In Wolf, E., editor, *Progress in Optics*, volume 60, pages 345–435. Elsevier.
- Demkowicz-Dobrzański, R., Kołodyński, J., and Guță, M. (2012). The elusive heisenberg limit in quantum-enhanced metrology. *Nature Communications*, 3:1063.
- Demkowicz-Dobrzański, R. and Maccone, L. (2014). Using entanglement against noise in quantum metrology. *Physical Review Letters*, 113(25):250801.
- Dicke, R. H. (1954). Coherence in Spontaneous Radiation Processes. *Physical Review*, 93(1):99–110.
- Dooley, K. L. and Collaboration, L. S. (2015). Status of GEO 600. *Journal of Physics: Conference Series*, 610(1):012015.

- Dorner, U., Demkowicz-Dobrzanski, R., Smith, B. J., Lundeen, J. S., Wasilewski, W., Banaszek, K., and Walmsley, I. A. (2009). Optimal quantum phase estimation. *Physical Review Letters*, 102:040403.
- Dür, W., Skotiniotis, M., Fröwis, F., and Kraus, B. (2014). Improved Quantum Metrology Using Quantum Error Correction. *Physical Review Letters*, 112(8):080801.
- Elsasser, W. M. (1937). On Quantum Measurements and the Role of the Uncertainty Relations in Statistical Mechanics. *Physical Review*, 52(9):987–999.
- Escher, B. M., Davidovich, L., Zagury, N., and de Matos Filho, R. L. (2012). Quantum Metrological Limits via a Variational Approach. *Physical Review Letters*, 109(19):190404.
- Escher, B. M., de Matos Filho, R. L., and Davidovich, L. (2011). General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology. *Nature Physics*, 7(5):406–411.
- Fraïsse, J. M. E. and Braun, D. (2015). Coherent averaging. *Annalen der Physik*, 527(9-10):701–712.
- Fraïsse, J. M. E. and Braun, D. (2017a). Enhancing sensitivity in quantum metrology by hamiltonian extensions. *Physical Review A*, 95:062342.
- Fraïsse, J. M. E. and Braun, D. (2017b). Hamiltonian extensions in quantum metrology. *Quantum Measurements and Quantum Metrology*, 4(1):8–16.
- Fraïsse, J. M. E. and Braun, D. (2017c). Quantum channel-estimation with particle loss: GHZ versus W states. *Quantum Measurements and Quantum Metrology*, 3(1).
- Fréchet, M. (1943). Sur l’extension de certaines évaluations statistiques au cas de petits échantillons. *Revue de l’Institut International de Statistique / Review of the International Statistical Institute*, 11(3/4):182–205.
- Frey, M., Collins, D., and Gerlach, K. (2011). Probing the qudit depolarizing channel. *Journal of Physics A: Mathematical and Theoretical*, 44(20):205306.
- Fröwis, F., Skotiniotis, M., Kraus, B., and Dür, W. (2014). Optimal quantum states for frequency estimation. *New Journal of Physics*, 16(8):083010.
- Fujiwara, A. (2001a). Estimation of SU(2) operation and dense coding: An information geometric approach. *Physical Review A*, 65(1):012316.
- Fujiwara, A. (2001b). Quantum channel identification problem. *Physical Review A*, 63(4):042304.
- Fujiwara, A. (2004). Estimation of a generalized amplitude-damping channel. *Physical Review A*, 70(1):012317.
- Fujiwara, A. (2006). Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. *Journal of Physics A: Mathematical and General*, 39(40):12489.
- Fujiwara, A. and Algoet, P. (1999). One-to-one parametrization of quantum channels. *Physical Review A*, 59(5):3290–3294.

- Fujiwara, A. and Imai, H. (2003). Quantum parameter estimation of a generalized Pauli channel. *Journal of Physics A: Mathematical and General*, 36(29):8093.
- Fujiwara, A. and Imai, H. (2008). A fibre bundle over manifolds of quantum channels and its application to quantum statistics. *Journal of Physics A: Mathematical and Theoretical*, 41(25):255304.
- George Casella, E. L. L. (1998). *Theory of Point Estimation*. Springer Texts in Statistics. Springer New York.
- Gessner, M., Pezzè, L., and Smerzi, A. (2016). Efficient entanglement criteria for discrete, continuous, and hybrid variables. *Physical Review A*, 94(2):020101.
- Giffard, R. P. (1976). Ultimate sensitivity limit of a resonant gravitational wave antenna using a linear motion detector. *Physical Review D*, 14(10):2478–2486.
- Giovannetti, V., Lloyd, S., and Maccone, L. (2006). Quantum Metrology. *Physical Review Letters*, 96(1):010401.
- Girolami, D. (2015). Interpreting quantum discord in quantum metrology. *Journal of Physics: Conference Series*, 626:012042.
- Girolami, D., Souza, A. M., Giovannetti, V., Tufarelli, T., Filgueiras, J. G., Sarthour, R. S., Soares-Pinto, D. O., Oliveira, I. S., and Adesso, G. (2014). Quantum Discord Determines the Interferometric Power of Quantum States. *Physical Review Letters*, 112(21):210401.
- Hayashi, M. (2006). *Quantum Information - An Introduction*. Springer-Verlag Berlin Heidelberg.
- Helstrom, C. W. (1960). *Statistical Theory of Signal Detection*. Macmillan.
- Helstrom, C. W. (1967). Minimum mean-squared error of estimates in quantum statistics. *Physics Letters A*, 25(2):101–102.
- Helstrom, C. W. (1968a). Detection theory and quantum mechanics (II). *Information and Control*, 13(2):156–171.
- Helstrom, C. W. (1968b). The minimum variance of estimates in quantum signal detection. *IEEE Transactions on Information Theory*, 14(2):234–242.
- Helstrom, C. W. (1969). Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252.
- Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory*, volume 123 of *Mathematics in Science and Engineering*. Elsevier.
- Henderson, L. and Vedral, V. (2001). Classical, quantum and total correlations. *Journal of Physics A: Mathematical and General*, 34(35):6899.
- Hillery, M. and Mlodinow, L. (1993). Interferometers and minimum-uncertainty states. *Physical Review A*, 48(2):1548–1558.
- Holevo, A. S. (1973a). Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4):337–394.

- Holevo, A. S. (1973b). Statistical problems in quantum physics. In *Proceedings of the second Japan-USSR Symposium on probability theory*, pages 104–119. Springer Berlin Heidelberg.
- Holevo, A. S. (1974). A Generalization of the Rao–Cramer Inequality. *Theory of Probability & Its Applications*, 18(2):359–362.
- Holevo, A. S. (2001). *Statistical Structure of Quantum Theory*, volume 67 of *Lecture Notes in Physics Monographs*. Springer-Verlag Berlin Heidelberg.
- Holevo, A. S. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*. Monographs (Scuola Normale Superiore). Edizioni della Normale.
- Holland, M. J. and Burnett, K. (1993). Interferometric detection of optical phase shifts at the Heisenberg limit. *Physical review letters*, 71(9):1355.
- Huang, Z., Macchiavello, C., and Maccone, L. (2016). Usefulness of entanglement-assisted quantum metrology. *Physical Review A*, 94:012101.
- Hübner, M. (1992). Explicit computation of the Bures distance for density matrices. *Physics Letters A*, 163(4):239–242.
- Hübner, M. (1993). Computation of Uhlmann’s parallel transport for density matrices and the Bures metric on three-dimensional Hilbert space. *Physics Letters A*, 179:226–230.
- Huelga, S. F., Macchiavello, C., Pellizzari, T., Ekert, A. K., Plenio, M. B., and Cirac, J. I. (1997). Improvement of Frequency Standards with Quantum Entanglement. *Physical Review Letters*, 79(20):3865–3868.
- Hyllus, P., Laskowski, W., Krischek, R., Schwemmer, C., Wieczorek, W., Weinfurter, H., Pezzé, L., and Smerzi, A. (2012). Fisher information and multiparticle entanglement. *Physical Review A*, 85(2):022321.
- Jamiolkowski, A. (1972). Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278.
- Ji, Z., Wang, G., Duan, R., Feng, Y., and Ying, M. (2008). Parameter Estimation of Quantum Channels. *IEEE Transactions on Information Theory*, 54(11):5172–5185.
- Jiang, M., Luo, S., and Fu, S. (2013). Channel-state duality. *Physical Review A*, 87(2):022310.
- Jing, L., Xiao-Xing, J., Wei, Z., and Xiao-Guang, W. (2014). Quantum Fisher Information for Density Matrices with Arbitrary Ranks. *Communications in Theoretical Physics*, 61(1):45.
- Jones, J. A. (2001). NMR quantum computation. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 38(4):325–360.
- Jones, J. A. (2011). Quantum computing with NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 59(2):91–120.
- Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Kennefick, D. (2014). Relativistic Lighthouses: The Role of the Binary Pulsar in proving the existence of Gravitational Waves. *ArXiv e-prints*. arXiv:1407.2164.

- Kessler, E. M., Lovchinsky, I., Sushkov, A. O., and Lukin, M. D. (2014). Quantum Error Correction for Metrology. *Physical Review Letters*, 112(15):150802.
- Kiilerich, A. H. and Mølmer, K. (2015). Quantum zeno effect in parameter estimation. *Physical Review A*, 92(3):032124.
- Knill, E. and Laflamme, R. (1998). Power of One Bit of Quantum Information. *Physical Review Letters*, 81(25):5672–5675.
- Knysh, S., Smelyanskiy, V. N., and Durkin, G. A. (2011). Scaling laws for precision in quantum interferometry and the bifurcation landscape of the optimal state. *Physical Review A*, 83(2):021804.
- Kolodynski, J. (2014). *Precision bounds in noisy quantum metrology*. PhD thesis, University of Warsaw. arXiv: 1409.0535.
- Kosugi, S. (2010). Proof of the standard quantum limit for monitoring free-mass position. *Physical Review A*, 82(2):022118.
- Landau, L. and Peierls, R. (1931). Erweiterung des Unbestimmtheitsprinzips für die relativistische Quantentheorie. *Zeitschrift für Physik*, 69(1-2):56–69.
- LSC, A. (2011). A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nature Physics*, 7(12):962–965.
- Luis, A. (2004). Nonlinear transformations and the Heisenberg limit. *Physics Letters A*, 329(1-2):8–13.
- Ma, Y., Miao, H., Pang, B. H., Evans, M., Zhao, C., Harms, J., Schnabel, R., and Chen, Y. (2017). Proposal for gravitational-wave detection beyond the standard quantum limit through EPR entanglement. *Nature Physics*, advance online publication.
- Maccone, L. (2013). Intuitive reason for the usefulness of entanglement in quantum metrology. *Physical Review A*, 88(4):042109.
- Macieszczak, K. (2015). Zeno limit in frequency estimation with non-Markovian environments. *Physical Review A*, 92(1):010102.
- Martynov, D. et al. (2016). Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Physical Review D*, 93(11):112004.
- Matsuzaki, Y., Benjamin, S. C., and Fitzsimons, J. (2011). Magnetic field sensing beyond the standard quantum limit under the effect of decoherence. *Physical Review A*, 84(1):012103.
- Modi, K., Brodutch, A., Cable, H., Paterek, T., and Vedral, V. (2012). The classical-quantum boundary for correlations: Discord and related measures. *Reviews of Modern Physics*, 84(4):1655–1707.
- Modi, K., Cable, H., Williamson, M., and Vedral, V. (2011). Quantum Correlations in Mixed-State Metrology. *Physical Review X*, 1(2):021022.
- Monroe, C. (2011). Demolishing quantum nondemolition. *Physics Today*, 64(1):8.

- Moss, G. E., Miller, L. R., and Forward, R. L. (1971). Photon-noise-limited laser transducer for gravitational antenna. *Appl. Opt.*, 10(11):2495–2498.
- Nagaoka, H. (2005). *On the Parameter Estimation Problem for Quantum Statistical Models*, pages 125–132. World Scientific Publishing Co.
- Nielsen, M. A. and Chuang, I. L. (2011). *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition.
- Ollivier, H. and Zurek, W. H. (2001). Quantum Discord: A Measure of the Quantumness of Correlations. *Physical Review Letters*, 88(1):017901.
- Ou, Z. Y. (1997). Fundamental quantum limit in precision phase measurement. *Physical Review A*, 55(4):2598–2609.
- Ozawa, M. (1988). Measurement breaking the standard quantum limit for free-mass position. *Physical Review Letters*, 60(5):385–388.
- Ozawa, M. (2004). Uncertainty relations for noise and disturbance in generalized quantum measurements. *Annals of Physics*, 311(2):350–416.
- Pang, S. and Brun, T. A. (2014). Quantum metrology for a general Hamiltonian parameter. *Physical Review A*, 90(2):022117.
- Pang, S. and Brun, T. A. (2016). Erratum: Quantum metrology for a general Hamiltonian parameter [Phys. Rev. A 90, 022117 (2014)]. *Physical Review A*, 93(5):059901.
- Paris, M. G. A. (2009). Quantum estimation for quantum technology. *International Journal of Quantum Information*, 07(supp01):125–137.
- Petz, D. (1996). Monotone metrics on matrix spaces. *Linear Algebra and its Applications*, 244:81–96.
- Pezzé, L. and Smerzi, A. (2009). Entanglement, Nonlinear Dynamics, and the Heisenberg Limit. *Physical Review Letters*, 102(10):100401.
- Pirani, F. A. E. (1956). On the Physical significance of the Riemann tensor. *Acta Physica Polonica*, 15:389–405.
- Pirani, F. A. E. (1957). Invariant Formulation of Gravitational Radiation Theory. *Physical Review*, 105(3):1089–1099.
- Radcliffe, J. M. (1971). Some properties of coherent spin states. *Journal of Physics A: General Physics*, 4(3):313.
- Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–89.
- Rondin, L., Tetienne, J.-P., Hingant, T., Roch, J.-F., Maletinsky, P., and Jacques, V. (2014). Magnetometry with nitrogen-vacancy defects in diamond. *Reports on Progress in Physics*, 77(5):056503.
- Roussas, G. G. (1997). *A Course in Mathematical Statistics, Second Edition*. Academic Press, San Diego, Calif, 2 edition edition.

- Roy, S. M. and Braunstein, S. L. (2008). Exponentially Enhanced Quantum Metrology. *Physical Review Letters*, 100(22):220501.
- Sarovar, M. and Milburn, G. J. (2006). Optimal estimation of one-parameter quantum channels. *Journal of Physics A: Mathematical and General*, 39(26):8487.
- Saulson, P. (2011). Josh goldberg and the physical reality of gravitational waves. *General Relativity and Gravitation*, 43(12):3289–3299.
- Saulson, P. R. (2013). Gravitational wave detection: Principles and practice. *Comptes Rendus Physique*, 14(4):288–305.
- Schirhagl, R., Chang, K., Loretz, M., and Degen, C. L. (2014). Nitrogen-Vacancy Centers in Diamond: Nanoscale Sensors for Physics and Biology. *Annual Review of Physical Chemistry*, 65(1):83–105.
- Schnabel, R., Mavalvala, N., McClelland, D. E., and Lam, P. K. (2010). Quantum metrology for gravitational wave astronomy. *Nature Communications*, 1:121.
- Schottky, W. (1918). Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. *Annalen der Physik*, 362(23):541–567.
- Seveso, L., Rossi, M. A. C., and Paris, M. G. A. (2017). Quantum metrology beyond the quantum Cramér-Rao theorem. *Physical Review A*, 95(1).
- Smarr, L. L. (1979). *Sources of Gravitational Radiation: Proceedings of the Battelle Seattle Workshop*. CUP Archive.
- Smirne, A., Kołodyński, J., Huelga, S. F., and Demkowicz-Dobrzański, R. (2016). Ultimate Precision Limits for Noisy Frequency Estimation. *Physical Review Letters*, 116(12):120801.
- Snider, R. F. (1964). Perturbation Variation Methods for a Quantum Boltzmann Equation. *J. Math. Phys.*, 5(11):1580–1587.
- Strobel, H., Muessel, W., Linnemann, D., Zibold, T., Hume, D. B., Pezzè, L., Smerzi, A., and Oberthaler, M. K. (2014). Fisher information and entanglement of non-Gaussian spin states. *Science*, 345(6195):424–427.
- Thorne, K. S. (1980). Gravitational-wave research: Current status and future prospects. *Reviews of Modern Physics*, 52(2):285–297.
- Thorne, K. S., Drever, R. W. P., Caves, C. M., Zimmermann, M., and Sandberg, V. D. (1978). Quantum Nondemolition Measurements of Harmonic Oscillators. *Physical Review Letters*, 40(11):667–671.
- Tóth, G. (2012). Multipartite entanglement and high-precision metrology. *Physical Review A*, 85(2):022322.
- Trees, H. L. V. (2013). *Detection, Estimation, and Modulation theory - Part I*. John Wiley & Sons Apr-22-2013.

- Uden, T., Balasubramanian, P., Louzon, D., Vinkler, Y., Plenio, M. B., Markham, M., Twitchen, D., Stacey, A., Lovchinsky, I., Sushkov, A. O., Lukin, M. D., Retzker, A., Naydenov, B., McGuinness, L. P., and Jelezko, F. (2016). Quantum Metrology Enhanced by Repetitive Quantum Error Correction. *Physical Review Letters*, 116(23):230502.
- Unruh, W. G. (1978). Analysis of quantum-nondemolition measurement. *Physical Review D*, 18(6):1764–1772.
- Weber, J. (1960). Detection and Generation of Gravitational Waves. *Physical Review*, 117(1):306–313.
- Wilcox, R. M. (1967). Exponential Operators and Parameter Differentiation in Quantum Physics. *J. Math. Phys.*, 8(4):962–982.
- Wineland, D. J., Bollinger, J. J., Itano, W. M., and Heinzen, D. J. (1994). Squeezed atomic states and projection noise in spectroscopy. *Physical Review A*, 50:67–88.
- Wineland, D. J., Bollinger, J. J., Itano, W. M., Moore, F. L., and Heinzen, D. J. (1992). Spin squeezing and reduced quantum noise in spectroscopy. *Physical Review A*, 46(11):R6797–R6800.
- Wootters, W. K. (1981). Statistical distance and Hilbert space. *Physical Review D*, 23(2):357–362.
- Yuan, H. (2016). Sequential Feedback Scheme Outperforms the Parallel Scheme for Hamiltonian Parameter Estimation. *Physical Review Letters*, 117(16):160801.
- Yuan, H. and Fung, C.-H. F. (2015). Optimal Feedback Scheme and Universal Time Scaling for Hamiltonian Parameter Estimation. *Physical Review Letters*, 115(11):110401.
- Yuen, H. and Lax, M. (1973). Multiple-parameter quantum estimation and measurement of nonselfadjoint observables. *IEEE Transactions on Information Theory*, 19(6):740–750.
- Yuen, H. P. (1983a). Contractive States and the Standard Quantum Limit for Monitoring Free-Mass Positions. *Physical Review Letters*, 51(9):719–722.
- Yuen, H. P. (1983b). Contractive States and the Standard Quantum Limit for Monitoring Freemass Positions. *Physical Review Letters*, 51(17):1603–1603.
- Yurke, B., McCall, S. L., and Klauder, J. R. (1986). SU(2) and SU(1,1) interferometers. *Physical Review A*, 33(6):4033–4054.
- Zawisky, M., Hasegawa, Y., Rauch, H., Hradil, Z., Myska, R., and Perina, J. (1998). Phase estimation in interferometry. *Journal of Physics A: Mathematical and General*, 31(2):551.
- Zwierz, M., Pérez-Delgado, C. A., and Kok, P. (2010). General Optimality of the Heisenberg Limit for Quantum Metrology. *Physical Review Letters*, 105(18):180402.
- Zwierz, M., Pérez-Delgado, C. A., and Kok, P. (2012). Ultimate limits to quantum metrology and the meaning of the Heisenberg limit. *Physical Review A*, 85(4):042112.



# List of Figures

|     |   |     |
|-----|---|-----|
| 1.1 | Statistical linearity for the first moment estimator . . . . .                          | 15  |
| 1.2 | Maximum likelihood estimation for the exponential distribution . . . . .                | 17  |
| 2.1 | Bloch sphere . . . . .  | 45  |
| 2.2 | Geometric representation of qubit's channel . . . . .                                   | 47  |
| 2.3 | Effect of qubit errors on the Bloch sphere . . . . .                                    | 48  |
| 3.1 | Diagram for quantum parameter estimation theory . . . . .                               | 70  |
| 4.1 | Michelson interferometer . . . . .  | 90  |
| 4.2 | Mach-Zehnder interferometer . . . . .   | 91  |
| 4.3 | Parallel protocol for quantum metrology . . . . .                                       | 95  |
| 4.4 | Complete quantum metrological protocol . . . . .  | 95  |
| 4.5 | Sequential protocol for quantum metrology . . . . .                                     | 96  |
| 4.6 | Protocols based on the "power of one qubit" . . . . .                                   | 107 |
| 4.7 | Ancilla assisted metrology . . . . .  | 108 |
| 5.1 | Coherent averaging protocol: Full-system estimation . . . . .                           | 122 |
| 5.2 | Coherent averaging protocol: Local estimation . . . . .                                 | 124 |
| 6.1 | Exact solution for the full-system estimation of $x$ with the ZZZZ model . . . . .      | 135 |
| 6.2 | Contour plot for the local QFI for $x$ with the ZZZZ model . . . . .                    | 137 |
| 6.3 | Local QFI for $x$ with the ZZZZ model as a function of $N$ . . . . .                    | 138 |
| 6.4 | Perturbative SNR for local estimation in the ZZZZ model . . . . .                       | 141 |
| 6.5 | Effect of the number of probes $N$ on the validity of the perturbation theory . . . . . | 142 |
| 7.1 | Full-system estimation of $x$ for the ZZXX Hamiltonian . . . . .                        | 151 |
| 7.2 | Full-system estimation of $\omega_p$ for the ZZXX Hamiltonian . . . . .                 | 152 |
| 7.3 | Full-estimation of $\omega_b$ for the ZZXX Hamiltonian . . . . .                        | 155 |
| 7.4 | Local estimation of $x$ for the ZZXX Hamiltonian . . . . .                              | 156 |
| 7.5 | Local estimation of $\omega_p$ for the ZZXX Hamiltonian . . . . .                       | 157 |
| 8.1 | Ideal depolarizing and phase-flip channel estimation . . . . .                          | 172 |
| 8.2 | Depolarizing channel estimation with one ancilla lost . . . . .                         | 178 |
| 8.3 | Phase-flip channel estimation with one ancilla lost . . . . .                           | 179 |
| 8.4 | Depolarizing channel estimation with a W states with $l$ ancillas lost . . . . .        | 180 |
| 8.5 | Depolarizing channel estimation with $l$ ancillas lost . . . . .                        | 181 |

|     |   |     |
|-----|---|-----|
| 8.6 | Phase-flip channel estimation with $l$ ancillas lost . . . . .  | 182 |
| 8.7 | Optimal number of ancillas in a $W$ state . . . . .   | 183 |
| 9.1 | Parameter shift in the Fisher information . . . . .   | 194 |
| 9.2 | Signal flooding for NV center magnetometry . . . . .  | 202 |
| 9.3 | Signal flooding and Hamiltonian subtraction for the estimation of a direction of a magnetic field . . . . . | 204 |
| 9.4 | Effect of a perturbation in "Hamiltonian subtraction" . . . . .   | 205 |
| B.1 | Integration area for Dyson series. . . . .  | 221 |

# List of Tables

|     |   |     |
|-----|---|-----|
| 5.1 | The four cases for perturbation theory with the coherent averaging protocol . . . . | 127 |
| 7.1 | Summary for the ZZZZ model . . . . .  | 159 |
| 7.2 | Summary for the the ZZXX model with full-system estimation. . . . .                 | 160 |
| 7.3 | Summary for the the ZZXX model with local estimation . . . . .                      | 160 |



# Contents

|   |           |
|---|-----------|
| <b>Abstract</b>   | <b>i</b>  |
| <b>Zusammenfassung</b>  | <b>ii</b> |
| <b>Summary</b>  | <b>iv</b> |
| <b>Abbreviations</b>  | <b>vi</b> |
| <b>Introduction</b>   | <b>1</b>  |
| <b>I Theoretical Framework</b>  | <b>7</b>  |
| <b>1 Parameter estimation theory</b>  | <b>9</b>  |
| 1.1 The problem of parameter estimation . . . . .                                   | 9         |
| 1.1.1 Parametric families, $n$ -samples and estimators . . . . .                    | 9         |
| 1.1.2 Good and bad estimators . . . . .   | 11        |
| 1.2 Some estimators . . . . .   | 14        |
| 1.2.1 Constant estimator . . . . .  | 14        |
| 1.2.2 Moment estimator . . . . .  | 14        |
| 1.2.3 Maximum likelihood estimator . . . . .  | 17        |
| 1.3 Cramér-Rao theorem . . . . .  | 19        |
| 1.3.1 A lower bound for the variance . . . . .                                      | 19        |
| 1.3.2 Change of parameters . . . . .  | 19        |
| 1.3.3 Proof of the Cramér-Rao theorem . . . . .                                     | 21        |
| 1.3.4 Cramér-Rao theorem and biased estimators . . . . .                            | 23        |
| 1.4 Fisher Information . . . . .  | 23        |
| 1.4.1 Additivity of the Fisher information for $n$ -samples . . . . .               | 23        |
| 1.4.2 Re-writing the Fisher information . . . . .                                   | 24        |
| 1.4.3 Geometric interpretation of the Fisher Information . . . . .                  | 25        |
| 1.5 Saturation of the Cramér-Rao bound in the finite and asymptotic cases . . . . . | 26        |
| 1.5.1 Saturating the bound . . . . .  | 26        |
| 1.5.2 Exponential families . . . . .  | 27        |
| 1.5.3 Maximum likelihood estimator and asymptotic study . . . . .                   | 27        |
| 1.6 Locality . . . . .  | 28        |
| 1.6.1 Locality of the Fisher information . . . . .                                  | 28        |

|          |   |           |
|----------|---|-----------|
| 1.6.2    | Locality in the Cramér-Rao theorem . . . . .                      | 29        |
| 1.6.3    | Operational justification of local estimation . . . . .           | 30        |
| <b>2</b> | <b>Foundation of quantum theory</b>                               | <b>33</b> |
| 2.1      | States . . . . .  | 33        |
| 2.1.1    | Pure and mixed states . . . . .                                   | 33        |
| 2.1.2    | Composite systems and entanglement . . . . .                      | 34        |
| 2.2      | Evolution . . . . .   | 35        |
| 2.2.1    | Kraus representation . . . . .                                    | 36        |
| 2.3      | Measurement . . . . .   | 36        |
| 2.4      | Going to a larger Hilbert space . . . . .                         | 37        |
| 2.4.1    | The usual postulates . . . . .                                    | 37        |
| 2.4.2    | The Church of the larger Hilbert space . . . . .                  | 38        |
| 2.4.3    | Interpretation and ambiguities . . . . .                          | 39        |
| 2.5      | Distance between states . . . . .                                 | 42        |
| 2.5.1    | Fidelity and Bures distance . . . . .                             | 42        |
| 2.5.2    | Properties of the Bures distance . . . . .                        | 42        |
| 2.6      | Two-level system: the model of the qubit . . . . .                | 43        |
| 2.6.1    | State of a qubit and Bloch vector picture . . . . .               | 43        |
| 2.6.2    | Assembly of qubits and spins . . . . .                            | 44        |
| 2.6.3    | Quantum channels for qubits . . . . .                             | 46        |
| <b>3</b> | <b>Quantum parameter estimation theory</b>                        | <b>51</b> |
| 3.1      | Introduction to q-pet . . . . .                                   | 51        |
| 3.1.1    | Estimation theory in physics . . . . .                            | 51        |
| 3.1.2    | Formalization of quantum parameter estimation theory . . . . .    | 52        |
| 3.1.3    | Classical <i>versus</i> Quantum . . . . .                         | 52        |
| 3.2      | Quantum Cramér-Rao Bound . . . . .                                | 53        |
| 3.2.1    | Quantum version of the theorem . . . . .                          | 53        |
| 3.2.2    | Proof of the QCRB I: Mathematical approach . . . . .              | 54        |
| 3.2.3    | Physical proof . . . . .  | 56        |
| 3.3      | Quantum Fisher Information . . . . .                              | 59        |
| 3.3.1    | Closed form for the QFI . . . . .                                 | 59        |
| 3.3.2    | Geometrical considerations - QFI and the Bures distance . . . . . | 61        |
| 3.3.3    | Properties of the QFI . . . . .                                   | 63        |
| 3.4      | Locality and saturation of the bound . . . . .                    | 64        |
| 3.4.1    | Saturation . . . . .  | 64        |
| 3.4.2    | Locality . . . . .  | 65        |
| 3.5      | Channel estimation . . . . .                                      | 67        |
| 3.5.1    | A first upper bound and QFI as a minimization procedure . . . . . | 69        |
| 3.5.2    | Channel QFI and channel extensions . . . . .                      | 69        |
| 3.5.3    | Upper bound and purification based QFI . . . . .                  | 72        |
| 3.6      | Hamiltonian parameter estimation . . . . .                        | 73        |
| 3.6.1    | Hamiltonian which commutes with its own derivative . . . . .      | 73        |
| 3.6.2    | Arbitrary Hamiltonian . . . . .                                   | 75        |
| 3.6.3    | Explicit form of the local generator . . . . .                    | 77        |
| 3.7      | Perturbative Hamiltonian parameter estimation . . . . .           | 77        |
| 3.7.1    | Parameter in the perturbation: PT I . . . . .                     | 78        |

|          |  |           |
|----------|--|-----------|
| 3.7.2    | Parameter in the unperturbed Hamiltonian: PT II . . . . .                          | 79        |
| 3.8      | SNR in perturbation theory . . . . .   | 80        |
| 3.8.1    | SNR in PT I . . . . .  | 80        |
| 3.8.2    | SNR in PT II . . . . .   | 81        |
| <b>4</b> | <b>Quantum metrology and quantum enhanced measurements</b>                         | <b>83</b> |
| 4.1      | Historical approach of the standard quantum limit and the Heisenberg limit . . . . | 83        |
| 4.1.1    | Gravitational wave detection . . . . .   | 83        |
| 4.1.2    | Birth of the SQL . . . . .   | 85        |
| 4.1.3    | SQL and HL in interferometry . . . . .   | 88        |
| 4.1.4    | Remarks on the shot noise . . . . .  | 93        |
| 4.1.5    | Legacy and limits of metrology for GW detection . . . . .                          | 93        |
| 4.2      | Heisenberg limit and standard quantum limit . . . . .                              | 94        |
| 4.2.1    | Protocols and resources in quantum metrology . . . . .                             | 94        |
| 4.2.2    | Standard definition of the SQL and the HL . . . . .                                | 97        |
| 4.2.3    | SQL and HL in quantum optics . . . . .   | 99        |
| 4.3      | Quantum enhanced measurements I: Enhancing the sensitivity with complex dynamics   | 102       |
| 4.3.1    | Non-linear schemes and $k$ -body metrology . . . . .                               | 102       |
| 4.3.2    | More on the SQL and the HL . . . . .   | 104       |
| 4.3.3    | Power of one qubit . . . . .   | 105       |
| 4.4      | Quantum enhanced measurements II: Ancilla assisted metrology . . . . .             | 106       |
| 4.4.1    | Channel extension for quantum metrology . . . . .                                  | 107       |
| 4.4.2    | Enhancing the sensitivity with ancillas . . . . .                                  | 108       |
| 4.4.3    | Ancilla assisted with more than one probe . . . . .                                | 109       |
| 4.5      | Correlations in quantum metrology . . . . .  | 111       |
| 4.5.1    | Entanglement between probes . . . . .  | 111       |
| 4.5.2    | Discord for metrology . . . . .  | 112       |
| 4.5.3    | Role of correlations in ancilla assisted schemes . . . . .                         | 114       |
| 4.6      | Noise in quantum metrology . . . . .   | 115       |
| 4.6.1    | Effect of the noise . . . . .  | 115       |
| 4.6.2    | Fighting the noise . . . . .   | 116       |

## II Coherent averaging 119

|          |  |            |
|----------|--|------------|
| <b>5</b> | <b>Coherent averaging protocol and metrology</b>                                     | <b>121</b> |
| 5.1      | Coherent averaging protocol . . . . .  | 121        |
| 5.1.1    | Protocol . . . . .   | 121        |
| 5.1.2    | Metrology for coherent averaging . . . . .   | 123        |
| 5.2      | Perturbative parameter estimation for coherent averaging . . . . .                   | 125        |
| 5.2.1    | Interaction pictures, PT I and PT II . . . . .                                       | 125        |
| 5.2.2    | Perturbative parameter estimation for the interaction parameter $x$ . . . . .        | 127        |
| 5.2.3    | Perturbative parameter estimation for free evolution parameters $\omega_P$ . . . . . | 128        |
| 5.2.4    | Perturbative parameter estimation for free evolution parameters $\omega_B$ . . . . . | 129        |
| 5.2.5    | Perturbative parameter local estimation in PT I . . . . .                            | 129        |

|          |  |            |
|----------|--|------------|
| <b>6</b> | <b>Qubit model: ZZZZ Hamiltonian</b>                               | <b>133</b> |
| 6.1      | Hamiltonian and state of the system . . . . .                      | 133        |
| 6.2      | Exact solution for full-system estimation . . . . .                | 134        |
| 6.2.1    | Full-system estimation of $x$ . . . . .                            | 134        |
| 6.2.2    | Full-system estimation of $\omega_p$ or $\omega_b$ . . . . .       | 135        |
| 6.3      | Exact solution for local estimation . . . . .                      | 135        |
| 6.3.1    | Local QFI . . . . .  | 136        |
| 6.3.2    | Local observable estimation . . . . .                              | 138        |
| 6.4      | Perturbation theory as an example . . . . .                        | 139        |
| 6.4.1    | Full-system estimation . . . . .                                   | 140        |
| 6.4.2    | Local estimation . . . . .   | 140        |
| 6.5      | Probes in a thermal state . . . . .                                | 142        |
| 6.5.1    | State and evolution . . . . .                                      | 143        |
| 6.5.2    | Full-system estimation for thermal states . . . . .                | 143        |
| 6.5.3    | Local estimation for thermal states . . . . .                      | 144        |
| <b>7</b> | <b>Qubit model: ZZXX Hamiltonian</b>                               | <b>147</b> |
| 7.1      | General qubit model of coherent averaging . . . . .                | 147        |
| 7.1.1    | General qubit model . . . . .                                      | 147        |
| 7.1.2    | QFI for $\omega_p$ in PT I . . . . .                               | 148        |
| 7.2      | ZZXX Hamiltonian and methods . . . . .                             | 148        |
| 7.2.1    | The ZZXX Hamiltonian . . . . .                                     | 148        |
| 7.2.2    | Numerics . . . . .   | 149        |
| 7.3      | Full-system estimation on the ZZXX model . . . . .                 | 150        |
| 7.3.1    | Full-system estimation of $\omega_p$ . . . . .                     | 153        |
| 7.3.2    | Full-system estimation of $\omega_b$ . . . . .                     | 154        |
| 7.4      | Local estimation on the ZZXX model . . . . .                       | 154        |
| 7.4.1    | Local estimation of $x$ . . . . .                                  | 154        |
| 7.4.2    | Local estimation of $\omega_p$ . . . . .                           | 158        |
| 7.5      | General summary of the results for the qubit-based model . . . . . | 159        |

### III Extensions for metrology 163

|          |   |            |
|----------|---|------------|
| <b>8</b> | <b>Channel extension for qubit errors: robustness of W and GHZ states</b> | <b>165</b> |
| 8.1      | Depolarizing and phase-flip channel estimation . . . . .                  | 165        |
| 8.1.1    | The depolarizing and phase-flip channels . . . . .                        | 165        |
| 8.1.2    | Optimal estimation of a Pauli channel . . . . .                           | 166        |
| 8.1.3    | Non-optimal estimation of the depolarizing channel . . . . .              | 167        |
| 8.2      | Estimation of the ideal quantum channels . . . . .                        | 169        |
| 8.2.1    | Extension of the channels . . . . .                                       | 169        |
| 8.2.2    | States considered . . . . .   | 169        |
| 8.2.3    | Ideal estimation of depolarizing channels . . . . .                       | 170        |
| 8.2.4    | Phase-flip channel . . . . .  | 173        |
| 8.3      | Losing particles . . . . .  | 174        |
| 8.3.1    | Modelling the loss . . . . .  | 174        |
| 8.3.2    | Losing one ancilla . . . . .  | 176        |
| 8.3.3    | Generalization to the loss of $l$ ancillas . . . . .                      | 178        |

|          |  |            |
|----------|--|------------|
| 8.3.4    | Depolarizing channel . . . . .                                     | 178        |
| 8.3.5    | Phase-flip channel . . . . .                                       | 181        |
| 8.3.6    | Gain <i>versus</i> robustness . . . . .                            | 182        |
| <b>9</b> | <b>Hamiltonian engineering</b>                                     | <b>185</b> |
| 9.1      | From channel extension to Hamiltonian extension . . . . .          | 185        |
| 9.2      | Upper bound to the channel QFI . . . . .                           | 186        |
| 9.2.1    | Channel QFI as a semi-norm . . . . .                               | 187        |
| 9.2.2    | Upper bound to the channel QFI . . . . .                           | 189        |
| 9.2.3    | Condition of saturation . . . . .                                  | 190        |
| 9.2.4    | History of the bound . . . . .                                     | 191        |
| 9.3      | Reaching maximal sensitivity with Hamiltonian extensions . . . . . | 191        |
| 9.3.1    | Case study: The broken phase-shift . . . . .                       | 192        |
| 9.3.2    | Hamiltonian subtraction . . . . .                                  | 194        |
| 9.3.3    | Signal flooding . . . . .  | 197        |
| 9.3.4    | Parameter dependence and ancillas . . . . .                        | 198        |
| 9.4      | Time engineering with Hamiltonian extension . . . . .              | 199        |
| 9.5      | Known results . . . . .  | 199        |
| 9.6      | Examples and applications . . . . .                                | 200        |
| 9.6.1    | NV center magnetometry . . . . .                                   | 200        |
| 9.6.2    | Vector magnetometry with spin-1 . . . . .                          | 202        |
| 9.6.3    | Application to the coherent averaging . . . . .                    | 204        |
|          | <b>Conclusion</b>  | <b>207</b> |
|          | <b>Appendices</b>  | <b>211</b> |
| <b>A</b> | <b>Probability theory</b>  | <b>211</b> |
| A.1      | Foundation of probabilities . . . . .                              | 211        |
| A.2      | Probability space . . . . .  | 212        |
| A.2.1    | Interpretation of probabilities . . . . .                          | 213        |
| A.3      | Real random variables . . . . .                                    | 213        |
| A.3.1    | Continuous random variables . . . . .                              | 214        |
| A.3.2    | Discrete random variables . . . . .                                | 215        |
| A.4      | Moments . . . . .  | 215        |
| A.4.1    | Expectation value . . . . .  | 215        |
| A.4.2    | Variance . . . . .   | 216        |
| A.4.3    | Moment of order $k$ . . . . .                                      | 216        |
| A.4.4    | Co-variance . . . . .  | 216        |
| A.4.5    | Function of random variables . . . . .                             | 217        |
| A.5      | From probabilities to statistics . . . . .                         | 217        |
| A.5.1    | Bienaymé-Tchebychev inequality and law of large number . . . . .   | 218        |
| A.5.2    | The central limit theorem . . . . .                                | 218        |

|          |   |            |
|----------|---|------------|
| <b>B</b> | <b>Perturbation theory</b>  | <b>219</b> |
| B.1      | Interaction picture . . . . .   | 219        |
| B.2      | Dyson series and perturbation theory . . . . .  | 220        |
| B.3      | Second order perturbation theory . . . . .  | 221        |
| <b>C</b> | <b>Perturbative result for coherent averaging</b>   | <b>223</b> |
| C.1      | QFI in perturbation theory II . . . . .   | 223        |
| C.2      | QFI in PT II for the coherent averaging Hamiltonian . . . . .                                 | 224        |
| C.3      | SNR in PT I for the coherent averaging Hamiltonian . . . . .                                  | 227        |
| <b>D</b> | <b>Diagonalization of the matrix <math>K</math> and <math>G</math> for channel estimation</b> | <b>229</b> |
|          | <b>Bibliography</b>   | <b>241</b> |
|          | <b>List of figures</b>  | <b>244</b> |
|          | <b>List of tables</b>   | <b>245</b> |
|          | <b>Contents</b>   | <b>252</b> |