

Universidade de São Paulo  
Instituto de Física

Estimação Bayesiana para termometria  
colisional e computação quântica  
não-holonômica em tempo ótimo

Gabriel Oliveira Alves

Orientador: Prof. Dr. Gabriel Teixeira Landi

Dissertação de mestrado apresentada ao Instituto de Física da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre em Ciências.

Banca Examinadora:

Prof. Dr. Gabriel Teixeira Landi - Instituto de Física da Universidade de São Paulo

Prof. Dr. -

Prof. Dr. -

São Paulo  
2023

**FICHA CATALOGRÁFICA**  
**Preparada pelo Serviço de Biblioteca e Informação**  
**do Instituto de Física da Universidade de São Paulo**

Alves, Gabriel Oliveira

Estimação bayesiana para termometria colisional e computação quântica não-holonômica em tempo ótimo / Bayesian estimation for collisional thermometry and time-optimal holonomic quantum computation. São Paulo, 2023.

Dissertação (Mestrado) - Universidade de São Paulo. Instituto de Física. Depto. de Física dos Materiais e Mecânica.

Orientador(a): Prof. Dr. Gabriel Teixeira Landi

Área de Concentração: Física

Unitermos: 1. Informação quântica; 2. Termodinâmica; 3. Inferência bayesiana; 4. Computação quântica.

USP/IF/SBI-004/2023

University of São Paulo  
Physics Institute

Bayesian estimation for collisional thermometry  
and time-optimal holonomic quantum  
computation

Gabriel Oliveira Alves

Supervisor: Prof. Dr. Gabriel Teixeira Landi

Dissertation submitted to the Physics Institute of the  
University of São Paulo in partial fulfillment of the  
requirements for the degree of Master of Science.

Examining Committee:

Prof. Dr. Gabriel Teixeira Landi - Physics Institute of the University of São Paulo

Prof. Dr. -

Prof. Dr. -

São Paulo  
2023



## *Acknowledgements*

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), processo nº 2020/16050-0, pelo financiamento.

# *Abstract*

In this thesis we deal with two different topics. In the first half we investigate how the Bayesian formalism can be introduced into the problem of quantum thermometry – a field which exploits the high level of control in coherent devices to offer enhanced precision for temperature estimation. In particular, we investigate concrete estimation strategies, with focus on collisional thermometry, a protocol where a series of ancillae are sent sequentially to probe the system’s temperature. We put forth a complete framework for analyzing collisional thermometry using Bayesian inference. The approach is easily implementable and experimentally friendly. Moreover, it is guaranteed to always saturate the Cramér-Rao bound in the long-time limit. Subtleties concerning the prior information about the system’s temperature are also discussed and analyzed in terms of a modified Cramér-Rao bound associated with Van Trees and Schützenberger.

Meanwhile, in the last part of the thesis we approach the problem of non-adiabatic holonomic computation. Namely, we investigate the implementation based on  $\Lambda$ -systems. It is known that a three-level system can be used in a  $\Lambda$ -type configuration in order to construct a universal set of quantum gates through the use of non-Abelian nonadiabatic geometrical phases. Such construction allows for high-speed operation times which diminish the effects of decoherence. This might be, however, accompanied by a breakdown of the validity of the rotating-wave approximation (RWA) due to the comparable timescale between counter-rotating terms and the pulse length, which greatly affects the dynamics. Here, we investigate the trade-off between dissipative effects and the RWA validity, obtaining the optimal regime for the operation of the holonomic quantum gates.

**Keywords:** Quantum thermodynamics; Quantum thermometry; Quantum metrology; Open quantum systems; Geometrical phases; Non-adiabatic quantum computing; Holonomic quantum computing.

# *Resumo*

Realizamos, nesta dissertação, uma investigação a respeito de dois tópicos distintos. Inicialmente lidamos com o formalismo Bayesiano e com o desafio de como introduzi-lo na termometria quântica – um campo que explora o alto nível de controle em dispositivos coerentes a fim de oferecer precisão aprimorada na inferência de temperaturas. Em particular, nós investigamos estratégias de estimação concretas, com foco na termometria colisional, um protocolo onde uma série de ancilas interage sequencialmente com um sistema de interesse, sondando sua temperatura. Esta abordagem é facilmente implementável e também propícia experimentalmente. Ademais, ela também assegura que a desigualdade de Cramér-Rao seja saturada no limite assintótico. Sutilezas a respeito da informação prévia da temperatura do sistema também são discutidas e analisadas em termos de uma desigualdade de Cramér-Rao modificada, associada a Van Trees e Schützenberger.

Já na parte final deste trabalho, abordamos o problema de computação quântica não-adiabática. Isto é, nós investigamos uma implementação baseada em sistemas  $\Lambda$ . Sabe-se que um sistema de três níveis pode ser utilizado na configuração de tipo  $\Lambda$  de modo que um conjunto universal de portas lógicas quânticas seja construído. Este tipo de implementação permite tempos de operação muito curtos, que amenizam os efeitos da decoerência. Esta escolha pode ser, contudo, acompanhada da invalidação da aproximação de onda girante (RWA) devido às escalas de tempo equivalentes entre os termos girantes e a duração do pulso; o que afeta a dinâmica do sistema de maneira significativa. Nosso objetivo consiste em investigar a competição entre os efeitos dissipativos e a validade da RWA, obtendo um regime ótimo de operação para as portas quânticas holonômicas.

**Palavras-chave:** Termodinâmica quântica; Termometria quântica; Metrologia quântica; Sistemas quânticos abertos; Fases geométricas; Computação quântica não-adiabática; Computação quântica holonômica.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Bayesian estimation for collisional quantum thermometry . . . . .	1
1.2	Time-optimal holonomic quantum computation . . . . .	3
<b>I</b>	<b>Bayesian estimation for collisional thermometry</b>	<b>5</b>
<b>2</b>	<b>Estimation theory</b>	<b>6</b>
2.1	Estimation theory . . . . .	7
2.2	The mean square error . . . . .	11
2.3	Minimum variance (unbiased) estimators . . . . .	13
2.4	The Fisher information . . . . .	17
2.5	The Cramér-Rao bound . . . . .	20
2.6	The maximum likelihood estimation . . . . .	24
2.7	Properties of estimators . . . . .	25
<b>3</b>	<b>Bayesian statistics</b>	<b>30</b>
3.1	Bayes' theorem . . . . .	31
3.2	Bayesian estimators . . . . .	38
3.3	The Van Trees-Schützenberger inequality . . . . .	46
<b>4</b>	<b>Quantum metrology and thermometry</b>	<b>54</b>
4.1	The quantum Fisher information . . . . .	55
4.2	A trick for calculating the QFI . . . . .	60
4.3	Quantum thermometry . . . . .	72
4.4	Collisional quantum thermometry . . . . .	77
<b>5</b>	<b>Bayesian quantum thermometry</b>	<b>82</b>
5.1	Ancilla-ancilla correlations . . . . .	83
5.2	Numerical analysis . . . . .	85
5.3	Bayesian estimation and collisional thermometry . . . . .	88
5.4	Effect of noisy probes . . . . .	94



<b>II</b>	<b>Time-optimal holonomic quantum computation</b>	<b>97</b>
<b>6</b>	<b>Geometrical phases in quantum mechanics</b>	<b>98</b>
6.1	The Berry phase . . . . .	99
6.2	Non-adiabatic geometrical phases . . . . .	103
6.3	Non-adiabatic non-abelian geometrical phases . . . . .	106
<b>7</b>	<b>Non-adiabatic holonomic quantum computing</b>	<b>109</b>
7.1	The $\Lambda$ -type system . . . . .	109
7.2	Implementing single-qubit gates . . . . .	112
7.3	Implementing two-qubit gates . . . . .	116
7.4	Gate robustness . . . . .	118
<b>8</b>	<b>Time-optimality</b>	<b>123</b>
8.1	Counter-rotating effects in the bright-dark basis . . . . .	124
8.2	Single-qubit gates . . . . .	126
8.3	Heterogeneous frequencies . . . . .	130
8.4	Two-qubit gates . . . . .	133
<b>9</b>	<b>Conclusion</b>	<b>136</b>
<b>A</b>	<b>Vectorization</b>	<b>140</b>
<b>B</b>	<b>Bayesian inference for correlated ancillae</b>	<b>142</b>
<b>C</b>	<b>Uniform sampling over the Bloch sphere</b>	<b>144</b>

# List of Figures

2.1	Histogram count for the sample mean estimator. . . . .	10
2.2	Illustration of the bias and the variance of an estimator. . . . .	12
2.3	Illustration of a frequentist estimation procedure. . . . .	12
2.4	Variance of an estimator as a function of the parameter. . . . .	13
2.5	Estimators for the uniform sampling problem. . . . .	15
2.6	Fisher information of the Bernoulli distribution. . . . .	20
2.7	Comparison of estimators with the CRB. . . . .	26
2.8	Common definitions for estimators. . . . .	28
3.1	Illustration of the sequential updating scheme. . . . .	33
3.2	Posterior distribution in the coin toss experiment. . . . .	34
3.3	Plot of the Beta distribution for different hyperparameters. . . . .	36
3.4	Bayesian update of the normal distribution. . . . .	37
3.5	Steps for obtaining Bayesian estimators. . . . .	39
3.6	The <i>maximum a posteriori</i> and Bayesian average estimators for the Beta distribution. . . . .	42
3.7	MSE for the Beta distribution with $n = 2$ . . . . .	43
3.8	A diagram with the summary of different inference approaches. . . . .	46
4.1	Typical protocol in quantum metrology. . . . .	56
4.2	(Left) State $\rho_{\nu\theta}$ and (Right) the optimal measurement basis on the Bloch sphere. . . . .	62
4.3	Standard optimization strategies in metrology. . . . .	65
4.4	Illustration of qubit states in the Bloch sphere after an infinitesimal phase shift. . . . .	69
4.5	FI for the qubit in different scenarios. . . . .	71
4.6	Diagram illustrating a thermometry experiment. . . . .	73
4.7	Thermal Fisher information of a qubit. . . . .	76
4.8	An illustration of the setup used in collisional quantum thermometry. . . . .	78
4.9	QFI in the collisional thermometry scheme. . . . .	80
5.1	A log-linear plot for the correlation between an ancilla and its $n$ -th neighbor for different $SA$ coupling strengths. . . . .	84
5.2	Bayesian estimation for collisional thermometry. . . . .	89
5.3	CRB and VTSB in the collisional model. . . . .	90
5.4	Influence of the temperature interval and cutoff in the asymptotic accuracy of the model, with an optimization over the $SE$ coupling strength. . . . .	92

5.5	Effect of non-ideal probes in the model. . . . .	94
5.6	Persistent error in the model due to error in the probe preparation. . . . .	95
6.1	Adiabatic and non-adiabatic dynamics of an infinite potential well. . . . .	100
6.2	A closed path $C$ in the parameter space. . . . .	101
6.3	Illustration of the geometrical consequences of the transport of a vector on curved surfaces. . . . .	102
6.4	Schematic diagram of the Hilbert space and the ray space $\mathcal{P}$ . . . . .	104
6.5	Loop $C$ on the Bloch sphere for the periodic evolution of a qubit under a constant magnetic field. . . . .	105
7.1	Basic setup for the $\Lambda$ -type system. . . . .	110
7.2	Effects of the counter-rotating terms in the dynamics of the $\Lambda$ -system. . . . .	111
7.3	Dynamics of the $\Lambda$ -system in the bright-dark basis. . . . .	112
7.4	Illustration of sech pulses. . . . .	114
7.5	Setup for the ions in the two-qubit gate. . . . .	116
7.6	Basic setup for the $\Lambda$ -type system when an open quantum system approach is considered. . . . .	119
7.7	Open quantum system effects in the dynamics of the $\Lambda$ -system. . . . .	121
8.1	A diagram illustrating the competition between decoherence and the counter-rotating contributions. . . . .	124
8.2	Dynamics of the amplitudes of the bright and dark states in the $\Lambda$ -system . . . . .	125
8.3	$\Lambda$ -system in the Bloch sphere. . . . .	126
8.4	Time-optimal configuration of the S gate. . . . .	127
8.5	Contour plot of the infidelity for heterogeneous frequencies. . . . .	130
8.6	Infidelity as a function of the ratio $f_{1e}/f_{0e}$ for single-qubit gates. . . . .	132
8.7	Relevance of the $ e\rangle \leftrightarrow  d\rangle$ coupling in the $\Lambda$ -system. . . . .	132
8.8	Infidelity of the CZ gate as a function of $\beta/f_i$ . . . . .	134
C.1	Uniform distribution of points using the Fibonacci nodes algorithm. . . . .	144

# List of Tables

C.1	Mathematica code for the Fibonacci nodes algorithm. . . . .	145
C.2	Generates a set of points in the Bloch sphere from the Fibonacci nodes algorithm. . . . .	145

# Chapter 1

## Introduction

In this dissertation we engage with two different problems. From chapters 2 through 5 we tackle different aspects of (Bayesian) estimation theory and quantum metrology, applying these concepts into collisional quantum thermometry. Meanwhile, chapters 6 - 8 constitute the latter half of the dissertation, where we deal with holonomic quantum computing. Further contextualization and a more complete introduction to each subtopic can be found at the beginning of each of these chapters.

The contents of the first and second parts of this work were conducted under the supervision of prof. Dr. Gabriel T. Landi and prof. Dr. Erik Sjöqvist, respectively. I worked on the latter half mainly during my exchange studies at the University of Uppsala. Any mistakes, inaccuracies or oversights in this work are entirely of my own responsibility. This research has been supported by CAPES, FAPESP (Grant No.2020/16050-0) and the framework of Erasmus + KA107 International Credit Mobility, financed by the European Commission.

### **1.1 Bayesian estimation for collisional quantum thermometry**

Temperature is an ubiquitous notion which is prevalent in the everyday life and, despite finding its roots in thermodynamics, is also of widespread importance in several fields [1–5]. It is thus not surprising that further extending our comprehension of thermodynamics

– especially the formalization of the concept of temperature – and how it integrates with relevant technologies is constant endeavor in physics. In this context, quantum thermodynamics evolved as a natural extension of thermodynamics into the domain of quantum mechanics [6, 7].

In a similar fashion, the precise manipulation of parameters in quantum systems is, at the same time, both a central requirement for cutting-edge quantum technologies and also a very hallmark of the ever-growing technological prowess of quantum science. In a sense, the engineering of quantum systems and our conceptual grasp of quantum theory evolve in a symbiotic relationship, where a deeper understanding of fundamental concepts allows us to develop more advanced technology and vice-versa [8–10]. With such motivations in mind, quantum metrology provides a well-established foundation accompanied by many versatile tools [11]. Many parameters of interest in real applications cannot be directly measured in a laboratory. For that, their value must be *inferred* from other quantities which are directly accessible. This is where *statistics* and *estimation theory* come into rescue; they provide the proper framework to deal with this type of challenge.

These lines of investigation converge into what is known as *quantum thermometry*, a field which is concerned with the problem of temperature estimation in quantum systems, taking into account the distinct traits of quantum thermodynamics [12]. Translating the well known notions and ideas from classical theory into quantum thermometry and quantum thermodynamics is a herculean effort due to the very small scales and novel phenomena from quantum physics [13–15]. The challenges range from the reformulation of the laws of the thermodynamics [16, 17] and the very definition of temperature [18], to how we model interactions and how we design experiments [19–22].

Our objective in this dissertation is to incorporate the Bayesian framework into quantum thermometry, a research venue which, apart from recent works [23, 24], has been hitherto seldom explored. In particular, we employ tools from Bayesian estimation to study a concrete thermometric platform which makes use of collisional models [25] in what is known as *collisional quantum thermometry* [26]. In contrast to the paradigm known as *local* quantum thermometry, where one usually has a reasonably accurate estimate for the parameter beforehand, our approach falls into what is known as *global* quantum thermometry [27], which demonstrates good applicability even when previous

knowledge is insufficient and the temperature lies on a broader interval. The Bayesian approach has been regarded recently as very efficient and less restrictive in such scenarios [28].

In chapter 2 we introduce the central concepts from statistics and classical estimation theory, mainly from the optics of the frequentist approach. Afterwards, in chapter 3, we extend these investigations to the Bayesian framework. These two chapters lay the groundwork for the contents of the chapter 4, where we discuss quantum metrology in generality, and later on, quantum thermometry as well. We conclude the first half of this dissertation with our original contributions in chapter 5.

## 1.2 Time-optimal holonomic quantum computation

The advent of quantum mechanics in the past century has brought a plethora of different technologies upon us. Quantum computing, for instance, has seen an auspicious development throughout the years. Seminal works such as the ones by Deutsch [29] and Shor [30] have established a new ground for quantum technologies, showcasing how quantum mechanics can be used to optimize tasks in computation and information processing.

However, quantum computing naturally requires a suitable platform to be performed on. Those, in turn, should exhibit several desirable properties, such as scalability [31], robustness against error [32–34], long decoherence time [35, 36] and universality [37]. The search for such type of implementation is, in itself, an extensive area of research – one which has seen considerable progress over the last decades, with proposals ranging from trapped ions [38] to topological systems [39].

In this context, holonomic quantum computing emerged at the end of the nineties as an encouraging alternative [40, 41]. This approach is based on the use of geometrical phases for quantum computing due to their inherent robustness against certain types of errors and noise. Even within this field, several different implementations and protocols have been proposed [42]. In particular, early proposals were very often restricted to adiabatic implementations. These, however are more susceptible to open quantum system phenomena due to their long operation time. For that reason, non-adiabatic implementations [43, 44] are regarded as an effective strategy in mitigating these effects [45, 46].

In this work we examine one of these approaches, focusing our attention on a non-adiabatic scheme for quantum computing based on three-level  $\Lambda$ -type systems, proposed in [45]. This framework has seen a couple of generalizations and alternative constructions in the last few years, together with concrete experimental implementations. Our main focus here is to study this implementation outside the regime of validity of the rotating wave approximation (RWA). Namely, we show that the breakdown in the RWA and decoherence arise as competing effects in the model, which suggests a time-optimal implementation.

In chapter 6 we review basic concepts about geometrical phases, with special focus on non-Abelian non-adiabatic phases. In chapter 7 we review the proposal for non-adiabatic quantum computing using  $\Lambda$ -type systems and we introduce some of our results for the non-RWA case. Finally, in chapter 8 we discuss our main results.



# **Part I**

## **Bayesian estimation for collisional thermometry**

# Chapter 2

## Estimation theory

Statistics is one of the most ubiquitous branches of mathematics due to its inherent applicability in real-life scenarios and experimental investigations [47–57]. It deals with questions ranging from fundamental and rigorous formulations of the theory [58–60], in a large intersection with probability theory, to the development of practical statistical methods and algorithmic approaches [61–65]. Statistics, from its centuries of development, provides a plethora of available tools and theoretical frameworks [66, 67]. Our objective in this section is to give a brief overview of one of its sub-branches: the estimation theory.

Just like its parent field, estimation theory is a quite extensive subject and is contained within two major frameworks: the frequentist and the Bayesian approaches. In this section we will deal with the former. As we shall see throughout the next few chapters, there is no single answer or clear-cut formulation to every circumstance we will encounter. For that reason, we discuss relevant definitions and several tools that might be of use in order to construct a versatile toolbox. Ultimately, which approach will be the most useful will depend on the problem at hand and practical constraints.

Nevertheless, in most of these situations we are interested in drawing useful information from empirical data. This lead us to one of the central problems in estimation theory, which is known as point estimation. In this scenario, one is interested in obtaining a single numerical value which serves as an estimation for a given parameter. Typically, practical formulations are guided either by the estimator which is the most accurate or, at the expense of some precision, an estimator which is reasonably simple to construct. Beyond the point estimation, other approaches for parameter estimation might be based

on confidence intervals or posterior distributions. In later chapters we shall study a few of these concepts.

In the current section, we will discuss some essential definitions regarding the point estimation problem with some concrete examples. We also introduce a central result from information and estimation theory, known as the Cramér-Rao bound (CRB) [68, 69]. In chapter 5, these concepts will be used to investigate a quantum thermometer based on the so-called collision models.

## 2.1 Estimation theory

Let us start by considering a string of discrete outcomes  $\mathbf{X} = (X_1, \dots, X_n)$  obtained from a given probability distribution. The sample space, i.e., the values that an element  $X_i$  in  $\mathbf{X}$  can assume, will depend on the problem. For instance, the sample space can be binary, where  $X_i$  is given by either 0 or 1, describing a coin-toss experiment. Or, similarly,  $X_i = 1, \dots, 6$ , describing the outcomes of a dice. We might also have discrete outcomes with a continuous sample space  $X_i \in \mathbb{R}$  describing, e.g., the readout of a sensor at specific time intervals [70].

These outcomes, and their underlying probability density function (pdf)  $p_\theta(X)$ , will usually depend on a parameter  $\theta$  (or on a vector of parameters  $\boldsymbol{\theta}$ ). We define an estimator as an arbitrary function of  $\mathbf{X}$  which is used to estimate  $\theta$ . In other words, we can denote the estimator  $\hat{\theta}(\mathbf{X})$  as

$$\hat{\theta}(\mathbf{X}) = g(X_1, \dots, X_n), \quad (2.1)$$

where  $g(X_1, \dots, X_n)$  is a function of the outcomes. While this definition might seem trivial, the most important insight is that since  $X_1, \dots, X_n$  are random variables, the estimator  $\hat{\theta}(\mathbf{X})$  is *also* a random variable itself. Therefore, it is clear that the estimator will depend on a particular *realization* (or trajectory) of the problem, that is, on the particular streak of results  $X_1, \dots, X_n$ . Therefore, a different sequence of results might yield a different estimate. In this case, how can one judge the quality of an estimator? A certain estimator might perform very well for a particular realization of the problem and very bad for the others, for example.

Another important aspect is that we obviously do not want the estimator in Eq. (2.1)

to depend on the unknown parameter  $\theta$ , because this is the very object we are trying to estimate. However, since the parametrization  $\theta$  will influence which  $\mathbf{X}$  we might observe and how likely certain realizations are to occur, by consequence, any error measure for the estimator  $\hat{\theta}(\mathbf{X})$ , as well as its distribution, will implicitly depend on the unknown parameter  $\theta$  itself. Hence,

- The estimator is a random variable which depends on a particular realization of  $X_1, \dots, X_n$ , but does not depend explicitly on the parametrization  $\theta$ .
- An error measure associated with this estimator will typically depend both on the outcomes  $X_1, \dots, X_n$  and also on the particular value of  $\theta$ .

In other words, one can define error measures  $\epsilon$  which depend only on  $\mathbf{X}$  and are conditioned solely on the parameter:

$$\epsilon_\theta = \epsilon(\hat{\theta}(\mathbf{X})|\theta), \quad (2.2)$$

or some which depend only on the parameter but are conditioned on the data, so in this case we can think of the data as being fixed:

$$\epsilon_{\mathbf{X}} = \epsilon(\theta|\mathbf{X}), \quad (2.3)$$

and finally, others which are conditioned on neither and depend only on the estimator:

$$\epsilon_B = \epsilon(\hat{\theta}(\mathbf{X}); \theta). \quad (2.4)$$

This is one of the main points of divergence between different frameworks in estimation theory. The first case, as illustrated by Eq. (2.2), is the approach taken by *frequentist* statistics and it is known as *classical* estimation theory. In this scenario only the outcomes are random variables, and the parameter  $\theta$  is regarded as *fixed* but unknown. This will be our main framework during this chapter. The latter formulations are Bayesian approaches, which we will discuss in the next sections.

▣ **Example:** In order to make these definitions less abstract, we will work on a concrete example in order to motivate the first definition. For that, we shall discuss a coin-toss experiment, which will be a recurrent example in this dissertation. For that, we consider a

coin which is not necessarily fair. That is, the outcomes are either head or tails, which we denote by  $X = 0$  or  $X = 1$ , and they are obtained with a probability  $1 - \theta$  or  $\theta$ , respectively. The probability  $p(X|\theta)$  of obtaining an outcome  $X$ , given  $\theta$ , is thus determined by

$$p_\theta(X) := p(X|\theta) = \theta^X(1 - \theta)^{(1-X)}, \quad (2.5)$$

which we call the *Bernoulli* distribution. If we perform sequential tosses and assume that they are all independent and identically distributed<sup>1</sup> (i.i.d), the probability of obtaining  $k$  heads and  $n - k$  tails on  $n$  trials is given by the *binomial* distribution:

$$p(k, n|\theta) = \binom{n}{k} \theta^k (1 - \theta)^{(n-k)}. \quad (2.6)$$

Note that Eq. (2.6) does not depend on the particular order of the outcomes, and only on the total number of heads (and tails). This is due to the particular form of the distribution in Eq. (2.5) and the fact that trials are i.i.d. This property is also related to the concept of sufficient statistics, which we are going to discuss in later sections.

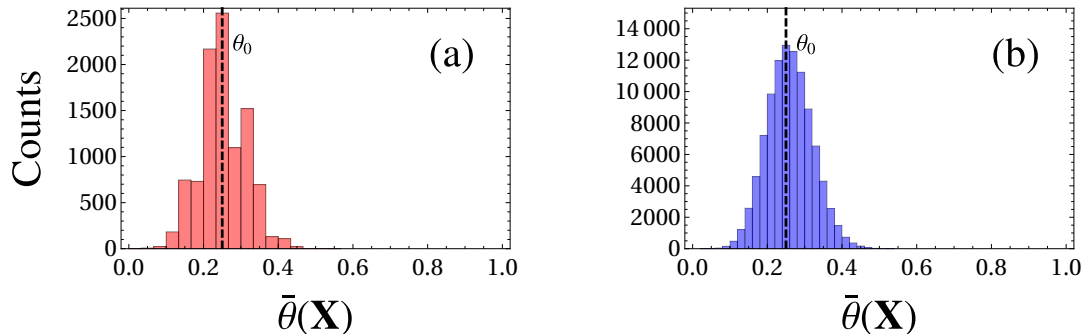
A very well-known estimator we can use to guess the true value of  $\theta$  in Eqs. (2.5) and (2.6) is the *sample mean*, defined as:

$$\bar{\theta}(\mathbf{X}) = \bar{\mathbf{X}} := \frac{X_1 + \dots + X_n}{n}. \quad (2.7)$$

As we will see, this estimator is useful due to the fact that, besides its simplicity, it is also optimal in a few senses. It is important to discuss what type of results Eq. (2.7) would yield in real experiments. For that, we perform a few numerical simulations with  $\theta_0 = 0.25$  being the true value of the parameter. We will consider that a single stochastic realization consists in a trial of 50 coin-tosses. Thus, in each realization we have a stochastic trajectory  $\mathbf{X}$  which is just a string of 50 zeroes or ones, representing heads or tails. We are interested in what values the estimator (2.7) might result in. We have claimed before that estimators are random variables themselves, so numerical simulations are a very convenient way of visualizing their distribution. One way to see this is by calculating  $\bar{\theta}(\mathbf{X})$  for different stochastic realizations. In Fig. 2.1 we plot a histogram for the sample mean

---

<sup>1</sup>The probability of each outcome is computed from the same probability distribution, in this case from Eq. (2.5), and each trial is independent.



**Figure 2.1:** Histogram count for the sample mean estimator. Results for  $\theta = 0.25$  and (a)  $10^4$  realizations and (b)  $10^5$  realizations.

$\bar{\theta}(\mathbf{X})$  for a few thousand trajectories.

We can see that the estimation will depend on a particular realization. We might get an estimate which yields  $\bar{\theta} = 0.3$  in one trial,  $\bar{\theta} = 0.17$  in another one, and so on. Still, in either case in Fig. 2.1 we can see that in this example the estimator is most likely to be found at its true value  $\theta_0 = 0.25$ . That of course makes sense, we can expect to get different proportions of heads and tails by pure chance, and different values of  $\theta$  could have resulted in the same sequence outcomes we have observed, so it is not possible to find the exact  $\theta_0$  from a finite sample. It would be unnatural to expect to learn *everything* about the model in a single-shot or just a few realizations. Nevertheless, it is true that  $\theta_0$  is more likely to have generated the particular values of  $\mathbf{X}$  that we have observed. That is why we actually get a histogram which peaks around  $\theta_0$  but which is still bound to assume other values.  $\square$

From Fig. 2.1 we can see that the estimation might vary a lot depending on the realization. A few outlier trajectories might result in values which are somewhat far away from the true value of the parameter. For that reason, we arrive at two natural questions. The first one is whether there is any sensible error figure to quantify how close, on average, the estimator is to  $\theta$ . The second one is whether we can quantify how spread-out the distribution for the estimator is going to be. While an estimator might result in the true value of the parameter *on average*, as shown in Fig. 2.1, for example, it still might fluctuate a lot on certain occasions. Meanwhile, other estimators might yield a much tighter distribution. We will now define a very standard error measure, which will give further insight into these two questions.

## 2.2 The mean square error

One of the standard error measures in statistics is the mean square error (MSE), defined as:<sup>2</sup>

$$\epsilon(\hat{\theta}|\theta) = \mathbb{E} \left[ (\hat{\theta} - \theta)^2 \right]. \quad (2.8)$$

Here,  $\mathbb{E}[\cdot]$  denotes the statistical average over the appropriate sample space. The pdf  $p(X)$  over which the integration is performed is specified in the index as  $\mathbb{E}_p[\cdot]$ , however we omit it whenever the distribution is clear from the context. We can expand the equation above as

$$\epsilon(\hat{\theta}|\theta) = \mathbb{E} \left[ \hat{\theta}^2 \right] - 2\theta \mathbb{E} \left[ \hat{\theta} \right] + \theta^2.$$

Here we have used the fact that  $\theta$  is deterministic. We can add extra terms on the RHS,

$$\epsilon(\hat{\theta}|\theta) = \mathbb{E} \left[ \hat{\theta}^2 \right] - \mathbb{E} \left[ \hat{\theta} \right]^2 + \mathbb{E} \left[ \hat{\theta} \right]^2 - 2\theta \mathbb{E} \left[ \hat{\theta} \right] + \theta^2.$$

to rewrite the MSE in terms of the famous bias-variance relationship:

$$\epsilon(\hat{\theta}|\theta) = \text{Var}(\hat{\theta}) + b^2(\hat{\theta}). \quad (2.9)$$

The term  $\text{Var}(\hat{\theta})$  denotes the *variance* and it is written as:

$$\text{Var}(\hat{\theta}) = \mathbb{E} \left[ \hat{\theta}^2 \right] - \mathbb{E} \left[ \hat{\theta} \right]^2. \quad (2.10)$$

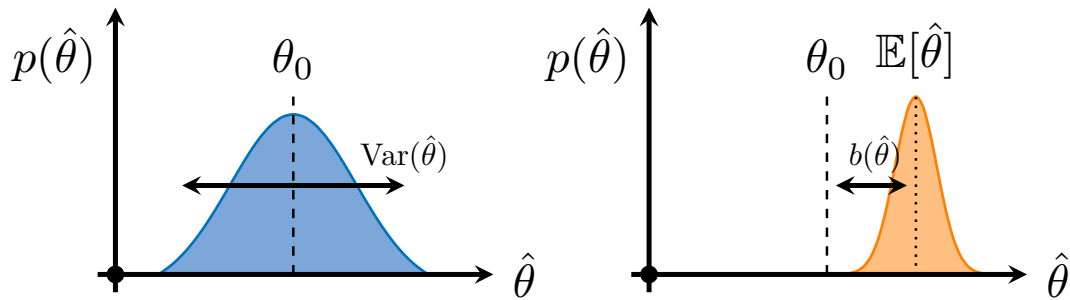
Meanwhile, we call the second term the *bias*, and it is defined as:

$$b(\hat{\theta}) = \mathbb{E} \left[ \hat{\theta} \right] - \theta. \quad (2.11)$$

The decomposition from Eq. (2.9) is also known as the bias-variance trade-off, which is very meaningful in fields such as, e.g., machine learning [71, 72]. The terms in Eqs. (2.10) and (2.11) also have very important interpretations in their own right. We illustrate their significance in Fig. 2.2. We can see that the bias is simply the difference between the expected value of the estimator and the true value of the parameter. Whenever  $b(\hat{\theta}) = 0$  for

---

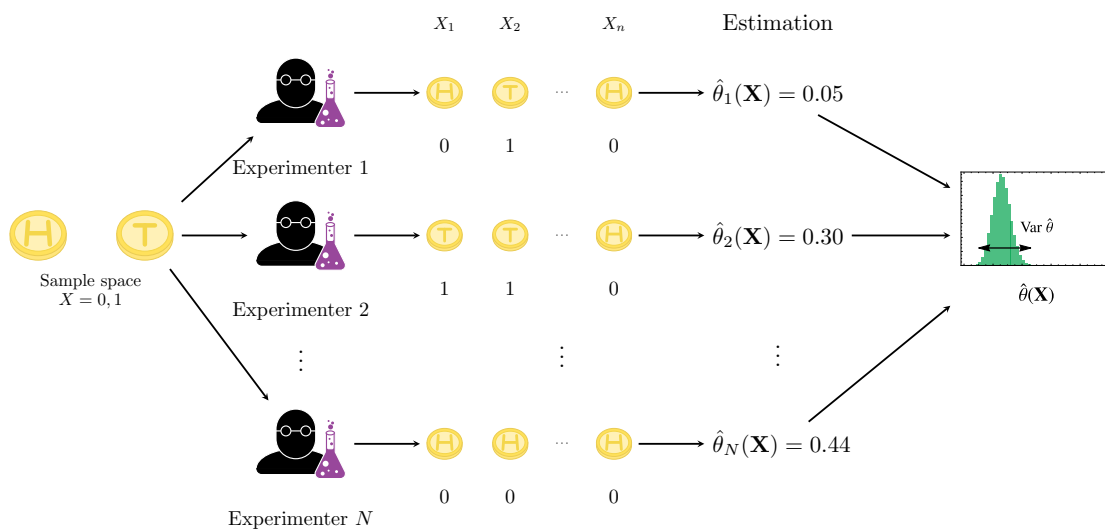
<sup>2</sup>In order to make the notation less cluttered, we express  $\hat{\theta}(\cdot)$  simply as  $\hat{\theta}$ .



**Figure 2.2:** Illustration of the bias and the variance of an estimator. On the left we depict the variance. On the right we show the bias. Note how we have a distribution with no bias but large variance, while on the right we see the converse; a distribution with a smaller variance but with the presence of a bias.

all  $\theta$ , we say that the estimator is *unbiased*. Similarly, the variance gives us information about how disperse the distribution for the estimator is. While only a proper probability distribution can give complete information about  $\hat{\theta}$ , the variance and the bias themselves already give a pretty important picture of the behavior of the estimation.

To make these definitions more intuitive, in Fig. 2.3 we describe a simple illustration of a usual estimation procedure in the frequentist sense. Going back to our example for the coin-toss experiment, we can suppose that  $N$  different agents will run a series of  $n$  coin tosses "in parallel". What we called a realization before is simply the  $n$  coin tosses that *each* of these experimenters perform. Of course, each of them will typically get



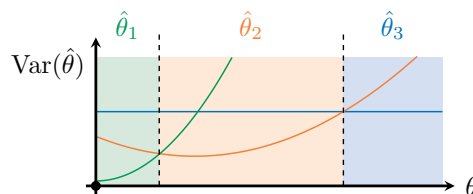
**Figure 2.3:** Illustration of an estimation procedure.



different results. This means that the estimator  $\theta_i(\mathbf{X})$  that the  $i$ -th agent will obtain will not necessarily agree with others. If we put all of their results in a histogram, we obtain a plot like the one seen in Fig. 2.1. In this scenario, the variance quantifies how well the different scientists will (dis)agree, quantifying how spread-out this distribution is going to be. The bias, on the other hand, is simply the average of their results, which is hopefully close to the true value of the parameter. Finally, the MSE will quantify how close to being correct their results are on average.

### 2.3 Minimum variance (unbiased) estimators

In Sec. 2.1 we saw a practical example of an estimation protocol, but note that we discussed the sample mean from Eq. (2.7) without any regard to its optimality or possible alternatives. A topic which naturally arises in this context is whether it is possible to minimize the MSE from Eq. (2.8), or even more generally, if there exists a procedure to construct other types of estimators. One could be tempted to define a minimum mean square error (MMSE) estimator, which minimizes Eq. (2.8). However, minimizing the MSE for *all* values of  $\theta$  with a single estimator is, in general, not possible [73] (p.19). For that reason, there is a couple of standard rules that one might follow in order to choose an appropriate estimator. One of them is to look for *minimum variance unbiased* (MVU) estimators. That is, we restrict ourselves to a class of unbiased estimators only, and choose the one with smallest variance. Nonetheless, even this strategy comes with a few caveats. For instance, the MVU estimator still might not exist. See Ref. [73] (p.20) or Ref. [74] (p.57) for examples. In Fig. 4.4 we depict a scenario where, depending on the value of  $\theta$ , different estimators are required for optimality. Even worse, there might be situations where *biased* estimators are a better choices; we arrive at a smaller MSE at the expense of an estimation which is,



**Figure 2.4:** Variance of an estimator as a function of the parameter. Its variance will depend on  $\theta$ . This means that there is possibly no single optimal estimator for the whole domain of the parameter. In this illustration there is no MVU estimator. A similar argument holds for the MMSE.

single estimator is, in general, not possible [73] (p.19). For that reason, there is a couple of standard rules that one might follow in order to choose an appropriate estimator. One of them is to look for *minimum variance unbiased* (MVU) estimators. That is, we restrict ourselves to a class of unbiased estimators only, and choose the one with smallest variance. Nonetheless, even this strategy comes with a few caveats. For instance, the MVU estimator still might not exist. See Ref. [73] (p.20) or Ref. [74] (p.57) for examples. In Fig. 4.4 we depict a scenario where, depending on the value of  $\theta$ , different estimators are required for optimality. Even worse, there might be situations where *biased* estimators are a better choices; we arrive at a smaller MSE at the expense of an estimation which is,

on average, slightly different from the true value. This is the case of ridge regression or regularization, for example [75, 76]. Regardless of these details, for now we focus on the MVU case and follow with a concrete example where we can discuss several estimators, their performance and some relevant definitions, commenting on further details about the optimality of estimators along the way.

▣ **Example:** We discuss a non-trivial example from Ref. [77] (pp.252-257, 274) where three very natural estimators arise. Consider that we sample a random variable  $X$  from the uniform distribution with height  $p_U(X) = 1/\theta$  in the interval  $[0, \theta]$ . Our objective is to estimate  $\theta$  from the array  $\mathbf{X} = (X_1, \dots, X_N)$ , that is, to estimate the largest value that  $X$  can assume.<sup>3</sup>

The first estimator we can try to define is the sample maximum  $\hat{\theta}_1 = \max \mathbf{X}$ . We need a few tricks in order to calculate the bias and the variance of this estimator. First, notice that for  $n$  samples, the probability that  $\hat{\theta}_1 = Y = \max(\mathbf{X})$  is smaller than a certain value  $y$  is given by

$$P(Y \leq y) = P(X_1 \leq y)P(X_2 \leq y)\dots P(X_n \leq y) = \frac{y^n}{\theta^n}, \quad (2.12)$$

since all outcomes are independent. We can differentiate the expression above to obtain:

$$p_Y(y) = \frac{ny^{n-1}}{\theta^n}, \quad 0 \leq y \leq \theta \quad (2.13)$$

with  $p_Y(y) = 0$  otherwise. The corresponding average reads:

$$\mathbb{E} [\hat{\theta}_1] = \int_0^\theta yp_Y(y)dy = \frac{n}{n+1}\theta.$$

We then find that the bias is

$$b(\hat{\theta}_1) = \mathbb{E} [\hat{\theta}_1] - \theta = -\frac{\theta}{n+1}, \quad (2.14)$$

showing that the estimator  $\hat{\theta}_1$  is biased. Fortunately, we can see that the bias goes to zero as  $n \rightarrow \infty$ , so in this case we say that the estimator is *asymptotically unbiased*. The

---

<sup>3</sup>The *German tank problem* is a historical problem which is very similar to this example [78].

minus sign in the equation above also makes sense: since we are taking the maximum of a finite sample, we are guaranteed to underestimate the actual maximum of the full distribution. From this point of view, the presence of a negative bias in the estimation is quite reasonable. In the same vein, we can then define a modified estimator

$$\hat{\theta}_2 = \frac{n+1}{n} \max(\mathbf{X}) \quad (2.15)$$

which is unbiased. Following the same steps we can actually show that  $\mathbb{E}[\hat{\theta}_2] = \theta$ . The extra factor of  $(n+1)/n$  correct for this underestimation that we mentioned before by overestimating some guesses, averaging out the bias. We can also calculate the variance of this estimator, finding [79]:

$$\text{Var}(\hat{\theta}_2) = \frac{\theta^2}{n(n+2)} \quad (2.16)$$

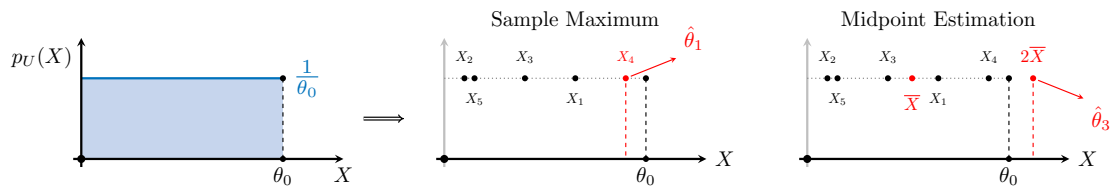
This, however, is not the only unbiased estimator we can define. Another sensible choice is the estimator

$$\hat{\theta}_3 = 2\bar{X}. \quad (2.17)$$

In this case the midpoint of the uniform distribution is approximated by the sample mean  $\bar{X}$ , and we double it in order to estimate  $\theta$ . However, we can show that the variance of this estimator is actually worse for  $n > 1$ , because it reads:

$$\text{Var}(\hat{\theta}_3) = \frac{\theta^2}{3n}. \quad (2.18)$$

Therefore, we can see that the modified sample maximum estimator given by Eq. (2.15)



**Figure 2.5:** Estimators for the uniform sampling problem. We show (left) the uniform distribution with unknown support, (center) the sample maximum estimator  $\hat{\theta}_1$  and (right) the midpoint estimator  $\hat{\theta}_3$ .

is better than the midpoint estimator  $\hat{\theta}_3$ . It is actually possible to show that this estimator is the MVU estimator of this distribution due to the fact that  $\mathbf{X}$  is a complete sufficient statistic [80] (p.342). In fact, we point Ref. [81] to the more interested reader for a more extensive discussion on this example and how different techniques can be used to build or improve upon estimators. We briefly go back to this discussion at the end of this chapter.

As a final comment, we check which of the estimators  $\hat{\theta}_1$  and  $\hat{\theta}_2$  has a smaller MSE. By using Eqs. (2.14) and (2.16) we can show, by direct calculation, that

$$\epsilon(\hat{\theta}_1|\theta) = \frac{2}{(n+1)(n+2)}\theta^2. \quad (2.19)$$

Meanwhile, since the modified sample mean estimator  $\hat{\theta}_2$  is unbiased, its MSE is simply given by its variance in Eq. (2.16), that is  $\epsilon(\hat{\theta}_2|\theta) = n\theta^2/(n+2)$ . Thus, we can see that  $\hat{\theta}_2$  is a better estimator even when we consider the MSE. One could be tempted to define a MMSE estimator for this example. However, as we claimed at the beginning of this section, this is not possible in general. As far as we are aware, this is a problem where there is no such strategy. To see this, consider the very trivial example of the constant estimator  $\hat{\theta}_c = c$ , with  $c > 0$ . In this case it is clear that the MSE is given by  $\epsilon(\hat{\theta}_c|\theta) = (c - \theta)^2$ . It is clear to see that this estimator is the best one for  $c = \theta$  but it is a very bad estimator otherwise, specially because its accuracy does not improve asymptotically. While this is a somewhat artificial example, it illustrates how we might come up with some estimators which are very good for some particular values of  $\theta$  but which very bad otherwise. That is why it is hard to obtain MMSE estimators in general (in the frequentist sense).

What we *can* do in this example, it to find an optimal estimator of the form  $\hat{\theta}_\alpha = \alpha \max(\mathbf{X}) = \alpha\hat{\theta}_1$ . To see this, note that we can simply write the MSE of this estimator as

$$\epsilon(\hat{\theta}_\alpha|\theta) = \mathbb{E} \left[ (\alpha\hat{\theta}_1 - \theta)^2 \right] = \alpha^2 \mathbb{E} \left[ \hat{\theta}_1^2 \right] - 2\alpha \mathbb{E} \left[ \hat{\theta}_1 \right] + \theta^2. \quad (2.20)$$

Using the results from Eqs. (2.14) and (2.16) once again, we get

$$\epsilon(\hat{\theta}_\alpha|\theta) = n\theta^2 \left( \frac{\alpha^2}{n+2} + 2\frac{\alpha}{n+1} - \frac{1}{n} \right). \quad (2.21)$$

Hence, by minimizing the equation above we can see that  $\alpha = (n+2)/(n+1)$  minimizes the MSE for this class of estimators. This result is quite interesting. Even though the modified sample mean  $\hat{\theta}_2$  is the best within the class of unbiased estimators, this new estimator  $\hat{\theta}_\alpha$ , which is a biased one, still wins when we use the MSE as a metric of error. The takeaway message is that looking for biased estimators might be useful in several scenarios, and that unbiasedness does not guarantee that the MSE is minimal.  $\square$

This example was very meaningful because i) it shows how we can construct an unbiased estimator from a biased one, going from  $\hat{\theta}_1$  to  $\hat{\theta}_2$ , and ii) it also illustrates a concrete scenario where we can directly compare a couple of different estimators in terms of their MSE and, more restrictively, unbiased estimators in terms of their variance. While we have introduced the sample maximum and the midpoint estimator in a somewhat arbitrary way in this example, in Sec. 2.6 we will show how these estimators can actually be obtained. Moreover, note how most of these estimators that we mentioned perform very well in the asymptotic limit. We will give some further thoughts on their asymptotic behavior in Sec. 2.7. However, in the next section we will take a slight detour, discussing a concept from statistics and information theory which will be central to the derivation of the Cramér-Rao bound in Sec. 2.5.

## 2.4 The Fisher information

Suppose we are given a continuous distribution  $p(X)$ . How can we introduce some notion of "closeness" with respect to another distribution  $q(X)$ ? A prevalent quantity in mathematical statistics which deals with this problem is the Kullback–Leibler (KL) divergence, also known as relative entropy, which is defined as:

$$D(p||q) := \mathbb{E}_p \left[ \ln \frac{p(X)}{q(X)} \right] = \int p(x) \ln \frac{p(x)}{q(x)} dx. \quad (2.22)$$

In the discrete case we simply substitute the integral by a summation. Strictly speaking, this is not really a measure of distance because it is not symmetric, nor it obeys the triangle inequality. Nevertheless, the KL divergence is still a intuitive way of comparing

distributions.<sup>4</sup>

Now, consider that the distribution  $p_\theta(x)$  is parametrized by  $\theta$ , as we had in the previous sections, and twice differentiable. We are interested in calculating the KL divergence for this distribution with itself but for *different* parametrizations. That is, given two parameters  $\theta$  and  $\theta_0$ , where we regard the latter as the "true" one, we can compute the KL divergence as:

$$D(p_{\theta_0}||p_\theta) = \int p_{\theta_0}(x) \ln \frac{p_{\theta_0}(x)}{p_\theta(x)} dx. \quad (2.23)$$

Since we expect to have  $D(p||q) \geq 0$ , with equality if, and only if,  $p = q$ , we can conclude that  $\theta = \theta_0$  is a minimum point for Eq. (2.23), as any other value of  $\theta$  should increase the KL divergence. So, if we regard  $p_\theta(x)$  as an estimated distribution, the KL divergence is a measure of how close we are to the true model  $p_{\theta_0}(x)$ .<sup>5</sup> Additionally, this also guarantees that the first derivative of Eq. (2.23) evaluates to zero at  $\theta = \theta_0$  and that it is locally convex around this minimum [85]. These observations allow us to Taylor-expand the KL divergence between the two parametrizations as:

$$D(p_{\theta_0}||p_\theta) = \frac{(\theta - \theta_0)^2}{2} F_{\theta_0}(X) + O((\theta - \theta_0)^3). \quad (2.24)$$

The term  $F_{\theta_0}(X)$  appearing in the equation above is defined as the *Fisher information*:

$$F(\theta) := \left. \frac{\partial^2}{\partial \theta^2} D(p_{\theta_0}||p_\theta) \right|_{\theta=\theta_0} \geq 0. \quad (2.25)$$

We will now follow some algebraic steps in order to make the definition above more operational and its interpretation more apparent. First, we re-express Eq. (2.25) in terms of the expected value at the point  $\theta = \theta_0$ :

$$F(\theta) = -\mathbb{E}_{p_\theta} \left[ \frac{\partial^2}{\partial \theta^2} \ln p_\theta(X) \right]. \quad (2.26)$$

---

<sup>4</sup>As its alternative name indicates, it is more correct to think of the KL divergence as a notion of entropy. In a way, information theory provides a better interpretation of this quantity. If we write the Shannon entropy of  $p$  as  $H(p) := -\sum_x p(x) \log p(x)$  and its cross-entropy with  $q$  as  $H(p, q) := -\sum_x p(x) \log q(x)$ , we can actually show that the identity  $D(p||q) = H(p, q) - H(p)$  holds. Therefore, we might think of  $D(p||q)$  as an entropy gain, or how much extra information (bits) should be provided in the context of e.g., optimal encoding, when we use a distribution  $q$  instead of a distribution  $p$  [82] (pp.19 - 21). This also makes clear why the relative entropy is not symmetric in general.

<sup>5</sup>A statistical-oriented interpretation of this distribution is to express the KL divergence as the expected value of the likelihood ratio [83] between  $p$  and  $q$  [84], as we see in the middle term of Eq. (2.22).

By using the product and chain rules and explicitly differentiating the expression inside the expected value we get:

$$F_{\theta}(X) = -\mathbb{E}_{p_{\theta}} \left[ \frac{1}{p_{\theta}(X)} \frac{\partial^2 p_{\theta}}{\partial \theta^2}(X) \right] + \mathbb{E}_{p_{\theta}} \left[ \left( \frac{1}{p_{\theta}(X)} \frac{\partial p_{\theta}}{\partial \theta}(X) \right)^2 \right]. \quad (2.27)$$

By using the normalization condition of the pdf, we can show that the first term vanishes:

$$-\mathbb{E}_{p_{\theta}} \left[ \frac{1}{p_{\theta}(X)} \frac{\partial^2 p_{\theta}}{\partial \theta^2}(X) \right] = -\frac{\partial^2}{\partial \theta^2} \int p_{\theta}(x) dx = 0.$$

Therefore, the alternative expression for the Fisher information reads:

$$F_{\theta}(X) = \int \frac{1}{p_{\theta}(x)} \left[ \frac{\partial p_{\theta}}{\partial \theta}(X) \right]^2 dx = \mathbb{E}_{p_{\theta}} \left[ \left( \frac{\partial}{\partial \theta} \ln p_{\theta}(X) \right)^2 \right]. \quad (2.28)$$

The Fisher information is a quantity which basically tell us how sensitive a distribution is to changes around  $\theta_0$ . If  $F_{\theta}$  is large, this means that the relative entropy  $D(p_{\theta_0} || p_{\theta})$  is very sharp in the neighborhood of  $\theta_0$ . We will probe a little bit further on this interpretation after we discuss the maximum likelihood method in Sec. 2.6. Moreover, the usefulness of the FI will become even clear after we discuss the Cramér-Rao bound in the next section.

In this derivation we have implicitly assumed some regularity conditions on the probability distribution  $p_{\theta}(x)$ . Besides the differentiability condition that we have briefly mentioned, it is also required that the support of the distribution does not depend on the parameter  $\theta$  itself. Unfortunately, this include some cases such as the example which we have discussed in the last section for the uniform distribution. In that problem, the support is given by  $[0, \theta]$ , with an explicit dependence on the parameter  $\theta$ . For now, we go back to the Bernoulli distribution from Sec. 2.1, discussing a concrete example:

▣ **Example:** Consider the Bernoulli distribution from Eq. (2.5). We have:

$$\frac{\partial}{\partial \theta} \ln p_{\theta}(X) = \frac{X}{\theta} - \frac{(1-X)}{(1-\theta)}. \quad (2.29)$$

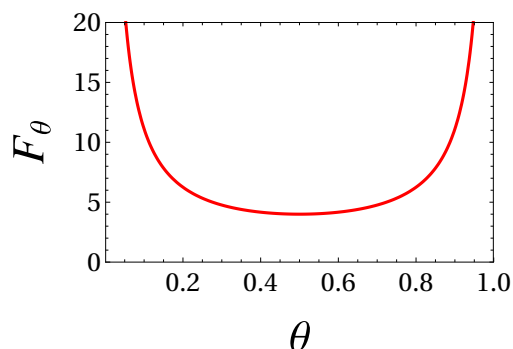
It holds that  $\mathbb{E}_{p_{\theta}} [X] = 0 \cdot p_{\theta}(X=0) + 1 \cdot p_{\theta}(X=1) = \theta$  and  $\mathbb{E}_{p_{\theta}} [X^2] = 0^2 p_{\theta}(X=0) + 1^2 p_{\theta}(X=1) = \theta$ . Thus, we can simply use Eqs. (2.28) and (2.29) to obtain the

Fisher information for the Bernoulli distribution

$$F_{\theta}(X) = \frac{1}{\theta(1-\theta)}. \quad (2.30)$$

Thus, we can see that the FI is minimum for  $\theta = 1/2$  and that it diverges for  $\theta \rightarrow 0$  or  $\theta \rightarrow 1$ . See the results in Fig. 2.6.  $\square$

Notice, both from Eq. (2.28) and from Fig. 2.6 in the example, that the Fisher information in general will depend on the parameter  $\theta$ . Thus, the information we have about a distribution will in general depend on the very parameter we are trying to estimate. It might therefore be difficult to be aware of the attainable information about a model beforehand. This is one of the difficulties we expect to address in subsequent sections.



**Figure 2.6:** Fisher information of the Bernoulli distribution.

## 2.5 The Cramér-Rao bound

We saw that the concept of Fisher information introduces a degree of knowledge, or of sensibility upon a parameter  $\theta$ , that we have about a model or a distribution. Is there a rigorous way of relating this with the framework of estimation theory? The very famous result by Cramér and Rao [68, 69], where one is able to relate the variance of an estimator with the Fisher information of the distribution, answers this question. This is one of the main results of the literature that we are going to review in this work and it will be fundamental for our review on metrology and thermometry in Chapter 4.

Here we introduce the Cramér-Rao bound for the more general case of *biased* estimator [74] (p.325). We follow a derivation similar to the one found in Ref. [73] (pp. 67-69). First, we require that the regularity condition

$$\mathbb{E}_{p_{\theta}} \left[ \frac{\partial \ln p_{\theta}(x)}{\partial \theta} \right] = 0. \quad (2.31)$$



holds for all  $\theta$ . The strategy to derive the CRB is to simply use the Cauchy-Schwartz inequality for integrals:

$$\int f(x)^2 w(x) dx \int g(x)^2 w(x) dx \geq \left( \int f(x) g(x) w(x) dx \right)^2. \quad (2.32)$$

It is intuitive to pick the weight here as the distribution, that is  $w(x) = p_\theta(x)$  so every integral becomes an expected value with respect to the  $p_\theta(x)$ . The other two choices we make are:

$$\begin{aligned} f(x) &= \frac{\partial \ln p_\theta(x)}{\partial \theta}, \\ g(x) &= \hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)]. \end{aligned} \quad (2.33)$$

The first function is also commonly called *score* in the literature, and  $g(x)$  are the fluctuations of the estimator around the mean. We can begin by solving the RHS of the inequality. Note that we can write this integral as:

$$\int f(x) g(x) w(x) dx = \int \frac{\partial \ln p_\theta(x)}{\partial \theta} \left( \hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)] \right) p_\theta(x) dx. \quad (2.34)$$

The first term is given by:

$$\int \hat{\theta}(x) \frac{\partial \ln p_\theta(x)}{\partial \theta} p_\theta(x) dx = \int \hat{\theta}(x) \frac{\partial p_\theta(x)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int \hat{\theta}(x) p_\theta(x) dx. \quad (2.35)$$

Note that here we used one of the regularity conditions which enable us to swap the integration and the derivative and also the assumption that the estimator is independent of  $\theta$ . The expression above is thus simply the derivative of the expected value of the estimator. We can then rearrange the term above in terms of the bias and the real parameter as:

$$\frac{\partial}{\partial \theta} \int \hat{\theta}(x) p_\theta(x) dx = 1 + b'(\theta), \quad (2.36)$$

since  $b(\theta) = \mathbb{E}[\hat{\theta}(x)] - \theta$ . Thus, we can write the inequality as:

$$\int f(x)^2 w(x) dx \int g(x)^2 w(x) dx \geq (1 + b'(\theta))^2. \quad (2.37)$$

Now we can show that the integral of the score is simply the Fisher information:

$$\int f(x)^2 w(x) dx = \int \left( \frac{\partial \ln p_\theta(x)}{\partial \theta} \right)^2 p_\theta(x) dx = F(\theta), \quad (2.38)$$

and finally, note that

$$\int g(x)^2 w(x) dx = \int \left( \hat{\theta}(x) - \mathbb{E}[\hat{\theta}(x)] \right)^2 p_\theta(x) dx = \text{Var} \hat{\theta}(x). \quad (2.39)$$

However, because of the bias-variance decomposition we have  $\epsilon(\hat{\theta}(x)|\theta) = \text{Var} \hat{\theta}(x) + b(\theta)^2$  [Eqs. (2.8) and (2.9)], and we can write

$$\int g(x)^2 w(x) dx = \epsilon(\hat{\theta}(x)|\theta) - b(\theta)^2. \quad (2.40)$$

Now we can simply rearrange the equations (2.37), (2.38) and (2.40) in order to arrive at the desired inequality [74]:

### The biased Cramér-Rao bound

If a pdf  $p_\theta(x)$  obeys the regularity condition (2.31) for all  $\theta$ , then any estimator  $\hat{\theta}(x)$  must satisfy:

$$\epsilon(\hat{\theta}(x)|\theta) \geq \frac{(1 + b'(\theta))^2}{F(\theta)} + b(\theta)^2. \quad (2.41)$$

If the estimator is unbiased, we obtain the standard and more famous result:

### The Cramér-Rao bound

Under the same conditions described above, any unbiased estimator  $\hat{\theta}(x)$  must satisfy:

$$\text{Var} \left( \hat{\theta}(X) \right) \geq \frac{1}{F(\theta)}. \quad (2.42)$$

This very important result establishes an ultimate limit of precision for *any* possible estimator. It is also clear that the existence of this bound also provides a very practical advantage: if we can find an estimator  $\hat{\theta}^*$  which saturates this bound, we know that  $\hat{\theta}^*$  will be optimal. This greatly simplifies the very troublesome quest of looking for MVU estimators. Any estimator which saturates Eq. (2.42) is said to be *efficient*.

A important property of the FI is its additivity: given two pdfs  $p_{1,\theta}(X_1)$  and  $p_{2,\theta}(X_2)$

with FI  $F_1(\theta)$  and  $F_2(\theta)$ , respectively, the FI of  $p_{1,\theta}(X_1)p_{2,\theta}(X_2)$  is given by their sum  $F_1(\theta) + F_2(\theta)$ . See Refs. [86, 87] for an extensive discussion on the properties of the FI. From the additivity of the FI, the CRB can be rewritten as

$$\text{Var} \left( \hat{\theta}(\mathbf{X}) \right) \geq \frac{1}{nF(\theta)}. \quad (2.43)$$

for a sequence  $\mathbf{X} = X_1, \dots, X_n$  of i.i.d. outcomes. If the estimator saturates the bound above asymptotically, it is said to be *asymptotically efficient*.

▣ **Example:** We found in the last example that the FI of the Bernoulli distribution is given by  $F_\theta(X) = 1/\theta(1 - \theta)$ . This means that the CRB for the binomial distribution (with  $n = 1$  for the Bernoulli distribution) reads  $\text{Var} \left( \hat{\theta}(\mathbf{X}) \right) \geq \theta(1 - \theta)/n$ . We would like to check how the variance of a real estimator compares with this bound. The natural candidate is simply the sample mean which we have discussed in Sec. 2.1. The variance of  $\bar{\theta}(\mathbf{X})$  is given by:

$$\text{Var} \left( \bar{\theta}(\mathbf{X}) \right) = \frac{\text{Var} X_1 + \dots + \text{Var} X_n}{n^2} = \frac{\theta(1 - \theta)}{n}. \quad (2.44)$$

Thus, we can see that the sample mean actually *saturates* the CRB in this case. It is not only a MVU estimator, but also an efficient estimator. However, the converse is not true in general; the existence of a MVU estimator does not guarantee its efficiency. See Ref.[88] (p.194) for example. □

The example above shows that the sample mean is the *best* unbiased estimator one can construct for the Bernoulli and the binomial distributions. This lead us to a very important question: under which conditions is it possible obtain efficient estimators which saturate the CRB? It can be shown that this is only possible for the exponential family [89, 90]. In more general scenarios, one will obtain estimators which, at best, satisfy the weaker condition of asymptotic efficiency. We will learn a method for obtaining such estimators in the next section.

## 2.6 The maximum likelihood estimation

R. A. Fisher introduced, during the early decades of the 20th century, a method known as *maximum likelihood estimation* (MLE). See Refs. [91] and [92] for a historical contextualization. This method is a quite general and practical approach, which is specially useful when obtaining a MMSE or a MVU estimator is either not trivial or not possible. Moreover, MLE results in estimators with several useful asymptotic properties.

The pdf  $p(X_1, \dots, X_n|\theta)$  for  $X_1, \dots, X_n$  conditioned of the parameter  $\theta$  is also known as the *likelihood*. The strategy of the MLE is very simple. Given the pdf  $p(\mathbf{X}|\theta)$  and the outcomes  $\mathbf{X} = X_1, \dots, X_n$ , what is the value of  $\theta$  which makes the observation of  $\mathbf{X}$  most likely? In other words, we define the maximum likelihood estimator as:

$$\hat{\theta}_{\text{MLE}}(\mathbf{X}) := \arg \max_{\theta} p(\mathbf{X}|\theta). \quad (2.45)$$

Since the logarithm is a monotonic function, it is sometimes more convenient, or numerically efficient, to minimize the *log-likelihood*  $\ln p(X_1, \dots, X_n|\theta)$  instead:

$$\hat{\theta}_{\text{MLE}}(\mathbf{X}) = \arg \max_{\theta} \ln p(\mathbf{X}|\theta). \quad (2.46)$$

▣ **Example:** Let us to back to the binomial distribution example. In this case we can take the derivative of the log-likelihood and equate it with zero to find:

$$\frac{\partial \ln p(n, k|\theta)}{\partial \theta} = \frac{k}{\theta} + \frac{n-k}{1-\theta} = 0,$$

which yields the sample mean estimator

$$\hat{\theta}_{\text{MLE}}(\mathbf{X}) = \frac{k}{n} = \frac{1}{n} \sum X_i. \quad (2.47)$$

▣

Similarly, it can be shown that the MLE also yields the sample maximum estimator  $\hat{\theta}_{\text{MLE}}(\mathbf{X}) = \max \mathbf{X}$  that we have discussed in the example of Sec. 2.3. This also shows that the MLE is not necessarily unbiased at finite sample size and that it is not guaranteed to coincide with the MVU.

We will briefly discuss some of the properties of the maximum-likelihood estimators. While we will not present the proofs, those can be obtained in any standard reference for point estimation theory (see e.g. [74, 80]).

- The MLE is asymptotically distributed as:

$$\hat{\theta}_{\text{MLE}}(\mathbf{X}) \underset{n \rightarrow \infty}{\sim} \mathcal{N}\left(\theta, \frac{1}{nF(\theta)}\right). \quad (2.48)$$

Here  $\mathcal{N}(\mu, \sigma^2)$  denotes a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ .

- The result above shows that the MLE is asymptotically unbiased and it is also asymptotically efficient, that is, its asymptotic variance is proportional to the reciprocal of the Fisher information. This implies that the maximum likelihood estimation is, at least asymptotically, an estimator of minimum variance.
- If  $g(\theta)$  is a one-to-one function of  $\theta$ , the MLE of  $g(\theta)$  is simply  $\hat{g}_{\text{MLE}} = g(\hat{\theta}_{\text{MLE}})$ .

These properties show why the MLE can be regarded as an asymptotically optimal estimator and why this method is a good rule-of-thumb when looking for estimators. This method is also numerically friendly. One can simply estimate the MLE numerically by using, e.g., Monte-Carlo methods. A disadvantage of this method which is worth mentioning is that it might be hard deciding beforehand how quickly it converges to a good asymptotic approximation [73]. Other procedures for obtaining estimators also exist, such as the method of moments.

## 2.7 Properties of estimators

In this section we give a brief summary the most relevant properties of estimators. We will review the properties that we have already encountered and also mention extra ones for completeness. Some of these properties will also motivate discussions in future chapters.

**Unbiasedness.** The estimator is said to be unbiased when its expected value coincides with the true value of the parameter for *every*  $\theta$ :

$$\mathbb{E}[\hat{\theta}] = \theta. \quad (2.49)$$

The sample mean estimator is an unbiased estimator for the binomial distribution, for example. We also mentioned the property of asymptotic unbiasedness before. This a weaker condition, which happens when the estimator is unbiased in the asymptotic limit  $n \rightarrow \infty$ . The sample maximum estimator that we have encountered in Sec. 2.3 was biased, but *asymptotically* unbiased.

**Efficiency.** An estimator is efficient when the equality in the CRB (2.42) holds. The sample mean for the binomial distribution is an example of efficient estimator. Analogously, an estimator is asymptotically efficient when the CRB is saturated in the asymptotic limit  $n \rightarrow \infty$ . See the chapter 10 of Ref. [80] for a discussion on the asymptotic properties of estimators. Note that even when the variance of an

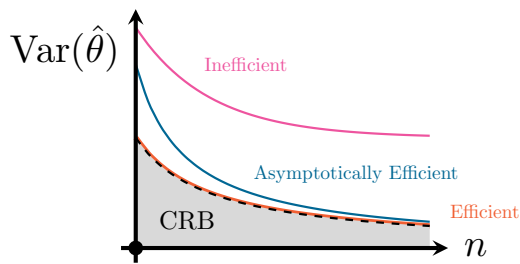
estimator goes to zero with  $n \rightarrow \infty$ , its asymptotic efficiency is not guaranteed. One such example is the median estimator for the average of a normal distribution [79] (p.290). That is, even in the asymptotic limit, some estimators do not attain the CRB.

Finally, we can also define the concept of *relative efficiency*. This is a property of unbiased estimators. If  $\text{Var } \hat{\theta}^* < \text{Var } \hat{\theta}$ , then  $\hat{\theta}^*$  is relatively more efficient than  $\hat{\theta}$ . Remember that if this inequality holds for *every*  $\hat{\theta}$ , then  $\hat{\theta}^*$  is the minimum variance unbiased estimator, as we saw in Sec. 2.3.

**Consistency.** This property means that by increasing  $n$  the estimator can approach the true value of the true parameter with arbitrary precision (in probability). This can be written as:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < \epsilon) = 1, \quad (2.50)$$

for all  $\epsilon > 0$ . While it is sometimes possible to explicitly check the consistency of an estimator by direct calculation, it is more usual to check for other properties: if the estimator is asymptotically unbiased and if its variance goes to zero with  $n \rightarrow \infty$ , then the estimator is consistent [80] (p. 469). For instance, note that this also means that the MLE is asymptotically consistent.



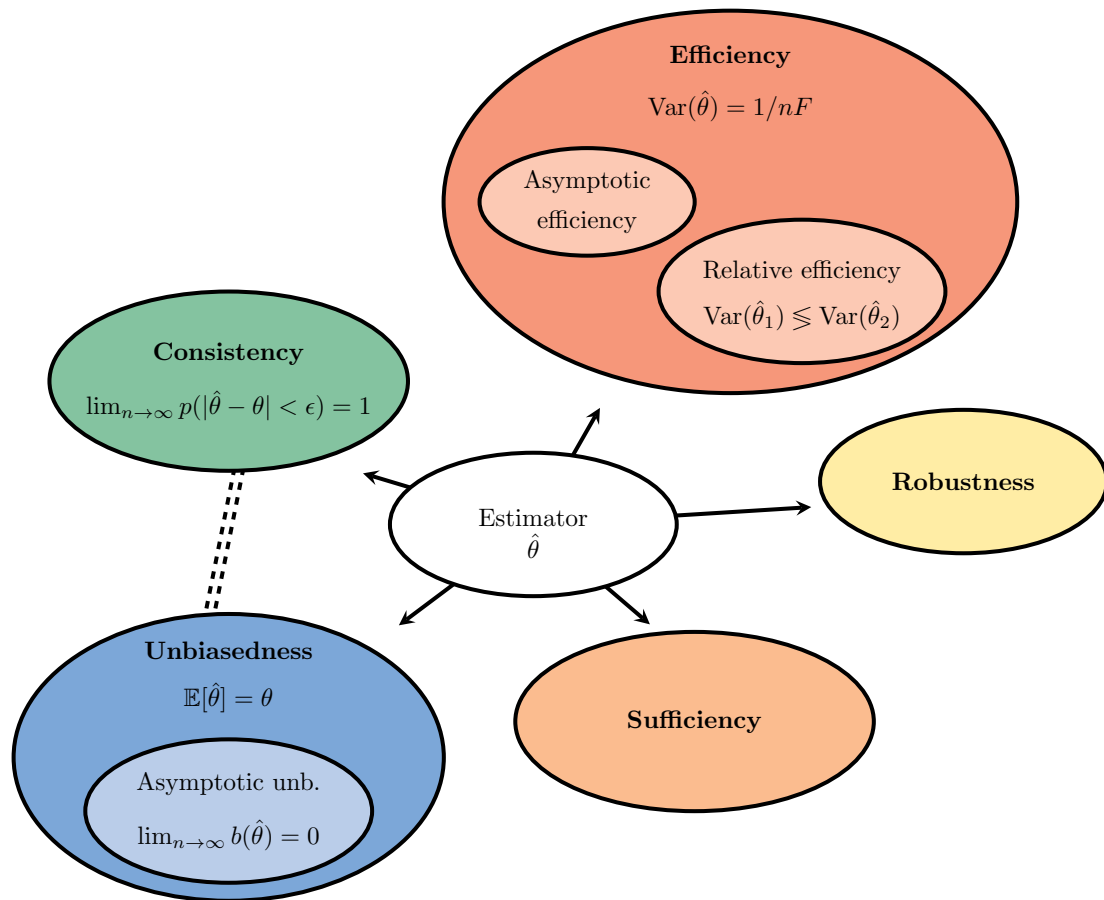
**Figure 2.7:** Comparison of estimators with the CRB.

**Sufficiency.** An estimator is said to be sufficient if, and only if, the pdf  $p(\mathbf{X}|\hat{\theta})$  of  $X_1, \dots, X_n$  conditioned on  $\hat{\theta}$  is independent of  $\theta$  [70] (pp. 305-307). This property means that no particular function or combination of the samples  $X_1, \dots, X_n$  in  $\theta$  is more likely than others. One such example is the binomial distribution from Eq. (2.6): the sample mean  $\bar{X} = (X_1 + \dots + X_N)/n = k/n$  happens to be a sufficient estimator in this case [79] (p. 295). Intuitively, this is due to the fact that only the number of heads and tails matter in the estimation, and not the order or any other complicated functions of the outcomes. In analogy, we can also express an estimator, such as the one above, in terms of a function  $T(\mathbf{X})$  of the outcomes  $\mathbf{X}$ , that is, of a statistic  $T$ . In the case of the sample mean, the sum  $X_1 + \dots + X_N$  of the outcomes is a sufficient statistic. A very helpful result, known as the *Fisher–Neyman factorization theorem*, can be used to verify whether a statistic is sufficient [73] (p. 104).

The idea of sufficiency is not only of conceptual interest: it can also aid us in finding a good estimator. Given an unbiased estimator  $\hat{\theta}$ , a sufficient statistic  $T$  and the conditional expectation  $\mathbb{E}[X|Y = x] := \sum_x x p(X = x|Y = y)$ , the Rao-Blackwell theorem [74] (p. 47) states that  $\hat{\theta}_T = \mathbb{E}[\hat{\theta}|T]$  is a better unbiased estimator of  $\theta$  for all  $\theta$ . ‘ In other words, this theorem gives a procedure for improving estimators simply by conditioning them on a sufficient statistics [73] (p. 109). This is of great practical aspect: one can start with a crude estimator and improve on it by using the Rao-Blackwell theorem. Equally importantly, as a corollary this theorem also implies that the MVU estimator is necessarily a function of a sufficient statistic.

This result from Rao and Blackwell leads to the Lehmann–Scheffé theorem, which states that if a statistics possesses a property called completeness, a unbiased estimator  $\hat{\theta}_T$  based on  $T$  is the *best* and *unique* unbiased estimator of  $\theta$ . We will not discuss this extra property here, but a statement of the theorem can be found in Ref. [80] (p. 369). With the aid of the factorization theorem, this construction can be used to show that the sample maximum is a sufficient statistics for the uniform distribution that we discussed in Sec. 2.3, for example. This proves that the unbiased sample maximum is precisely the MVU estimator [73] (p. 113).

Finally, it is worth stressing that regardless of these conclusions, biased estimators might still be preferable, so one should still be careful when looking for an optimal esti-



**Figure 2.8:** Common definitions for estimators.

mation strategy. See Ref. [93] for an interesting example.

**Robustness.** This is another property which we are not going to define mathematically, but we will nevertheless mention it for completeness. This is a more recent concept in relation to the others, and it may very often have different definitions in the literature. Nonetheless, robustness is related to the idea that an estimator should still perform well enough even under underlying assumptions about the models are not entirely correct or precise. One might, for instance, sacrifice optimality in exchange for an estimator which performs reasonably well even for a probability distribution (a model) slightly different from what is ideally expected [77] (p. 272). A concrete example is given in Ref. [80] (p.482).

A visual summary of these properties can also be found in Fig. 2.8. While we will not give further attention to the theoretical aspects of some of these properties in this work, it was important mentioning them in order to justify the use of a few estimation strate-



gies that we are going to be focusing from hereafter, such as the maximum likelihood-estimation, and a few Bayesian strategies that we are going to encounter in the next section. The interested reader can find several useful discussions and counter-examples for the theory of point estimation in Ref. [88].

# Chapter 3

## Bayesian statistics

Bayesian statistics constitutes a departure from the frequentist framework that we discussed in the last chapter. In the frequentist scenario we were mainly concerned with likelihood functions, which are conditioned on the deterministic, but unknown parameter, and also with quantities averaged over different realizations of the problem, such as the MSE. Meanwhile, here we focus on probability distributions over the parameters instead. Now, unknown parameters of a distribution are regarded as random variables themselves [94], while the relevant distributions are conditioned on the *data* instead. This is a useful approach in experimental scenarios, specially in problems where the notion of repeatability, very natural in frequentism, is less intuitive or difficult to implement. In those situations, the experimenter is probably interested in extracting information of his or her data as much as possible, and not from every possible realization. Similarly, maybe one is interested on how the model and how the estimators would behave for a large set (or interval) of different parameters.

This paradigm shift introduces a few conceptual and practical differences, as we shall see [95]. Many standard texts give a broad overview of the field those two schools. In fact, some authors even argue that Bayesian methods have a broader applicability nowadays [58]. We abstain from these discussions: the contents of this chapter are mainly concerned with some of the cornerstones of the Bayesian inference and their practical implications. In particular, in Secs. 3.1 and 3.2 we show how the Bayesian formulation can be used as an algorithmic procedure to obtain concrete estimators, in consonance with much of the development and the concepts which we have encountered in the last chapter. In the same

vein, the Bayesian approach results in a modified version of the Cramér-Rao bound [96] which we discuss in Sec. 3.3.

### 3.1 Bayes' theorem

The joint probability  $p(A, B)$  of two events  $A$  and  $B$  can be decomposed in a symmetric manner with the use of conditional probabilities:

$$p(A, B) = p(A|B)p(B) = p(B|A)p(A). \quad (3.1)$$

A rearrangement leads us to the *Bayes theorem*,

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}, \quad (3.2)$$

which is the basis of the Bayesian statistical inference. See Ref. [97] for a very complete introductory discussion and some brief historical remarks. A simple change of notation will bring an interpretation of the equation above which is closer to our current context. In the problem of estimation theory we can consider  $A$  to be the parameter  $\theta$  and  $B$  to be the sequence of random variables  $\mathbf{X} = X_1, \dots, X_n$ . By doing so, we rewrite Eq. (3.2) as:

$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int p(\mathbf{X}|\theta)p(\theta)} = \frac{1}{\mathcal{N}}p(\mathbf{X}|\theta)p(\theta). \quad (3.3)$$

The familiar term  $p(\mathbf{X}|\theta)$  is called the *likelihood*, as seen in the the problem of classical estimation. This terms tell us what is the probability of obtaining a particular trajectory (or detection record)  $\mathbf{X}$  if the distribution is parametrized by  $\theta$ . The term  $\mathcal{N} := p(\mathbf{X}) = \int p(\mathbf{X}|\theta)p(\theta)$  is just the marginalized distribution of  $\mathbf{X}$ , and it is essentially a normalization factor in the Bayes theorem.

The novelty of the Bayesian approach comes from the two other terms. The distribution  $p(\theta)$  is called the *prior distribution*, and it embeds any previous knowledge we might have about the parameter  $\theta$ . For instance, we might know that  $\theta$  is truncated at a certain value, or that it lies within a particular interval. All this information can be incorporated into the prior, ideally improving the estimates by including any previous

information about the model.

The final term,  $p(\theta|\mathbf{X})$  is called the *posterior distribution*, and is it the most important object in the Bayesian setting. One possible interpretation for this distribution is the following: given the realization  $X_1, \dots, X_n$ , how likely it is that these outcomes were generated by the parameter  $\theta$ ? Another very important view is that the posterior distribution represents an update of the prior: given  $p(\theta)$ , how does the observation of the outcomes  $X_1, \dots, X_n$  updates our state-of-knowledge about the model?

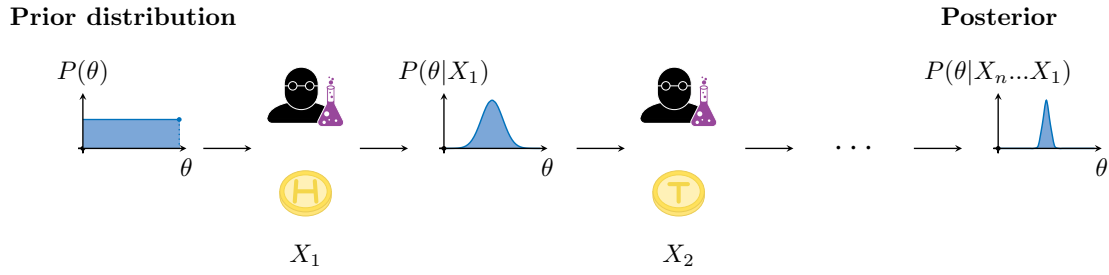
The latter interpretation thus allows for a sequential update scheme which is very natural in, e.g., real experiments or investigations. Suppose we start with a prior distribution  $p(\theta)$  and that we obtain an outcome  $X_1$ . The posterior distribution, given this outcome, is given by  $p(\theta|X_1) \propto p(X_1|\theta)p(\theta)$  up to the normalization factor. Now, suppose that the outcomes are i.i.d. Upon obtaining a second outcome  $X_2$ , we can use the first posterior  $p(\theta|X_1)$  as the prior in Eq. (3.3) and the new likelihood as the probability  $p(X_2|\theta)$ , so the updated distribution becomes  $p(\theta|X_2X_1) \propto p(X_2|\theta)p(\theta|X_1) \propto p(X_2|\theta)p(X_1|\theta)p(\theta)$ . With each new outcome  $X_{n+1}$ , one can update the distribution  $p(X_n \dots X_1|\theta)$  to  $p(X_{n+1}X_n \dots X_1|\theta)$ . In this case, after we observe a new outcome, the current posterior will become the prior in the subsequent step, in other words, we can depict this sequential updating scheme as: <sup>1</sup>

$$p(\theta|X_{n+1}X_n \dots X_1) \propto p(X_{n+1}|\theta)p(\theta|X_n \dots X_1) \propto p(X_{n+1}|\theta) \dots p(X_1|\theta)p(\theta). \quad (3.4)$$

A pictorial representation of this scheme is Shown in Fig. 3.1. Of course, if one is not interested in studying or visualizing the intermediate distributions, it is possible to update the prior with all the outcomes  $X_1, \dots, X_n$  at once in a single calculation; this avoids the trouble of normalizing the distribution at every step. Explicitly calculating the intermediate distribution is not strictly necessary, and both strategies are equivalent. We have included this description motivated by the fact that this is a strategy which is going to be used in later chapters. For now, we discuss in an example how the Bayes theorem can be used in practice.

---

<sup>1</sup>In the more general case where correlations might be present, the updated distribution would need to be conditioned both on  $\theta$  and on the previous outcomes  $X_1, \dots, X_n$ . We would then rewrite Eq. (3.4) as  $p(\theta|X_{n+1}X_n \dots X_1) \propto p(X_{n+1}|\theta, X_n \dots X_1)p(\theta|X_n \dots X_1)$ .



**Figure 3.1:** Illustration of the sequential updating scheme.

**Example:** Let us go back to the coin-toss example. In order to emulate ignorance about the parameter  $\theta$ , we will assume that the prior is given by a flat distribution  $p(\theta) = 1$ . The likelihood, as we are aware, is given by the Bernoulli distribution in Eq. (2.5). Now, suppose that one obtains an outcome  $X_1$ . The posterior distribution is given by:

$$p(\theta|X_1) = \frac{1}{\mathcal{N}_1} \theta^{X_1} (1 - \theta)^{1-X_1}. \quad (3.5)$$

Now, consider that we obtain a second outcome given by  $X_2$ . The posterior in this case becomes:

$$p(\theta|X_2X_1) = \frac{1}{\mathcal{N}_2} \theta^{X_1+X_2} (1 - \theta)^{2-X_1-X_2}. \quad (3.6)$$

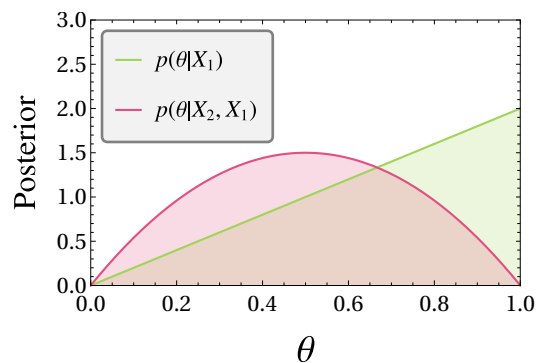
For concreteness, let us take a realization where  $X_1 = 1$  and that  $X_2 = 0$ . By doing so, the posterior in the equations above becomes:

$$\begin{aligned} p(\theta|X_1 = 1) &\propto \theta, \\ p(\theta|X_2 = 0, X_1 = 1) &\propto \theta(1 - \theta) \end{aligned} \quad (3.7)$$

By normalizing the distributions, we get  $\mathcal{N}_1 = \int_0^1 \theta d\theta = 1/2$  and  $\mathcal{N}_2 = 1/6$ , respectively. Thus, the posterior for the particular realization  $(X_1, X_2) = (1, 0)$  in the coin toss experiment is given by:

$$\begin{aligned} p(\theta|X_1 = 1) &= 2\theta, \\ p(\theta|X_2 = 0, X_1 = 1) &= 6\theta(1 - \theta). \end{aligned} \quad (3.8)$$

We show the results in Fig. 3.2. The plot for  $p(\theta|X_1)$  tell us that as we increase  $\theta$ , it is increasingly likely that one would obtain  $X_1 = 1$ . This probability is obviously zero (one) for  $\theta = 0$  ( $\theta = 1$ ), corresponding to the scenario where the coin is fully biased. Similarly,  $p(\theta|X_2, X_1)$  tells us that the greatest chance of obtaining exactly one head and one tail in two trials happens when the coin is perfectly fair, that is, when we have  $\theta = 1/2$ , as seen in the plot. Equivalently, we can say that parametrizing the Bernoulli distribution with  $\theta = 1/2$  maximizes our chances of obtaining the observed realization  $(X_1, X_2) = (1, 0)$ .



**Figure 3.2:** Posterior distribution in the coin toss experiment. Here we assume a flat prior and the realization  $(X_1, X_2) = (1, 0)$ .

Finally, note that for an arbitrary number of outcomes  $X_1, \dots, X_n$ , one would simply obtain the posterior distribution:

$$p(\theta|X_n \dots X_1) \propto \theta^{\sum_i X_i} (1 - \theta)^{n - \sum_i X_i}. \quad (3.9)$$

Normalizing the distribution above is not entirely trivial. We will go back to this example in the next section.  $\square$

The example above has shown how we can obtain the posterior for a desired parameter  $\theta$  that we expect to estimate. Additionally, we could explicitly see how the stochastic outcomes result in distributions which are updated with new measurements. However, differently from the point estimation scenario where we had a *single number* as an estimate, here we have a whole range of values that  $\theta$  might assume, with an associated probability  $p(\theta|\mathbf{X})$ . This is a departure from the classical estimation theory and one of the main differences which we encounter with the Bayesian procedures. Nevertheless, it is still possible to perform point estimation even in the Bayesian case, as we shall discuss in Sec. 3.2. This means, of course, that it would be erroneous to think that point estimation is restricted to the classical paradigm and that interval estimation is restricted to the Bayesian theory: the Bayesian and the frequentist approaches are, in different ways,

capable of both [74].

For now, let us take a slight detour in order to discuss a potential obstacle which might have been apparent from the last example. Obtaining the normalization factor in Eq. (3.3) is, in general, a very taxing effort. This is valid both in the theoretical and numerical scenarios [64]. The analytical normalization factor for the posterior from Eq. (3.9) in the example, for instance, requires some work to be found. Even so, a non-standard prior could result in an even more difficult integral, and, in the worst case scenario, one which might not have a solution in terms of elementary functions.

This complication motivates the definition of *conjugate priors*. We say that a prior  $p(\theta)$  is conjugate to a likelihood function  $p(\mathbf{X}|\theta)$  whenever the application of the Bayes theorem results in a posterior  $p(\theta|\mathbf{X})$  which belongs to the same family of probability distributions as the prior [98]. This provides the enormous advantage of simplifying the normalization of the posterior: instead of calculating a (possibly highly dimensional) integral for several outcomes at once, or re-normalizing it at every step, as we did in the last example, one can simply find the normalization factor *once* for pertinent family of distributions and update its parameters accordingly. We proceed with the Bernoulli distribution as an example.

**Example:** The conjugate for the Bernoulli distribution is the *Beta* distribution

$$\mathcal{B}_{(\alpha,\beta)}(\theta) := \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}, \quad (3.10)$$

defined in terms of *hyperparameters* (parameters which define the prior distribution)  $\alpha$  and  $\beta$ . The normalization constant is given by the Beta function, defined as:

$$B(\alpha,\beta) := \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \quad (3.11)$$

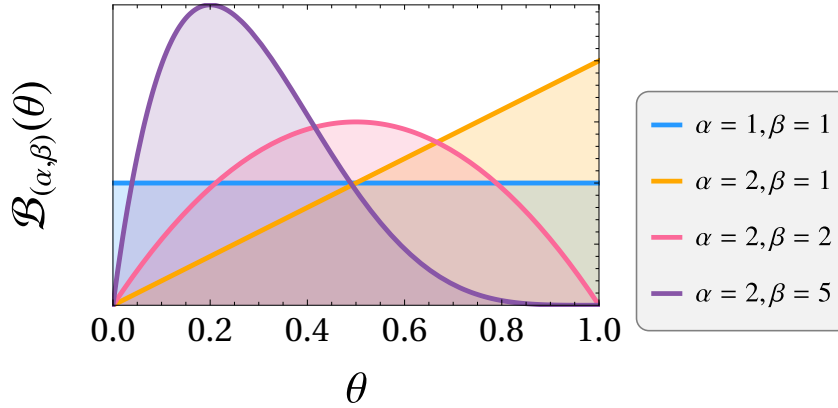
The second equality establishes a relationship between the Beta function and the well-known Gamma function  $\Gamma(x)$  (see Ref. [99]). While these definitions might come across as unnecessary complications, they will greatly simplify some calculations later on.

Suppose that our prior is given by the Beta distribution with hyperparameters  $\alpha_0$  and  $\beta_0$ , that is:  $p(\theta) = \mathcal{B}_{(\alpha_0,\beta_0)}(\theta)$ . If we perform  $n$  trials in the coin-toss experiment, obtain-

ing  $\sum_i X_i$  heads and  $n - \sum_i X_i$  tails, the posterior distribution is given by:

$$p(\mathbf{X}|\theta) \propto \underbrace{\theta^{\sum_i X_i} (1 - \theta)^{n - \sum_i X_i}}_{\prod_i p(x_i|\theta)} \underbrace{\theta^{(\alpha_0 - 1)} (1 - \theta)^{(\beta_0 - 1)}}_{\text{Prior}} = \theta^{\alpha_0 + \sum_i X_i - 1} (1 - \theta)^{n + \beta_0 - \sum_i X_i - 1}. \quad (3.12)$$

However, note that the RHS of the equation above is simply the Beta distribution  $\mathcal{B}_{(\alpha', \beta')}(\theta)$  with updated hyperparameters  $\alpha' = \alpha_0 + \sum_i X_i$  and  $\beta' = \beta_0 + n - \sum_i X_i$ , up to the normalization constant. Thus, if the normalization constant of the Beta distribution, or any other other property of interest such as the moments, is known beforehand, one can promptly find them for the posterior distribution simply by updating the hyperparameters accordingly.



**Figure 3.3:** Plot of the Beta distribution for different hyperparameters.

As a particular case, we mention that by taking  $\alpha_0 = \beta_0 = 1$  we obtain a flat distribution. Thus, the prior we chose in the previous example was actually a conjugate one. In that case, we could write the posterior as  $p(\theta|X_2 = 1, X_1 = 0) = \mathcal{B}(2, 2)$ . One can also easily find the normalization factor by calculating  $B(2, 2) = \Gamma(2)\Gamma(2)/\Gamma(4) = 1!1!/4! = 1/6$ . So we can see that this agrees with our previous brute-force results. For reference, we plot the beta distribution for a few different parameters in Fig. 3.3.  $\square$

We also discuss an important example for the normal distribution [73] (p. 332). The following results will be important for us in Sec. 3.3:

**Example:** Suppose we have a detection record given by i.i.d. data

$$X \sim \mathcal{N}(\mu_0, \sigma) \quad (3.13)$$



generated from a normal distribution with mean  $\mu_0 = 0$  and variance  $\sigma^2$ . How can we apply the Bayes rules to infer the mean of the Gaussian? We will suppose the likelihood is modeled by a Gaussian of known variance, given by  $P(X|\mu) = \mathcal{N}(\mu, \sigma)$ . For simplicity, we will also suppose that the prior is also a Gaussian, since the normal distribution is its own conjugate distribution [100]. We denote the mean and the variance of the prior by  $\mu_p$  and  $\sigma_p^2$ , respectively. By doing so, we can apply Bayes rule to arrive at the following posterior:

$$p(\mu|X) \propto p(X|\mu)p(\mu) = \exp\left(-\frac{(X - \mu)^2}{2\sigma^2} - \frac{(\mu - \mu_p)^2}{2\sigma_p^2}\right). \quad (3.14)$$

We can complete the squares by noticing that the term inside the exponential above may be written as

$$\frac{\sigma_p^2 + \sigma^2}{2\sigma_p^2\sigma^2} \left[ \mu^2 - 2\mu \frac{\sigma^2\mu_p + \sigma_p^2X}{\sigma_p^2 + \sigma^2} + \frac{\sigma_p^2\sigma^2}{\sigma_p^2 + \sigma^2} \left( \frac{X^2}{\sigma^2} \right) \right],$$

since the last term does not depend on  $\mu$  we can factor this part of the exponential out from the integral, since it will just be incorporated in the normalization factor. We can then complete the square. By doing so we arrive at the following posterior:

$$p(\mu|X) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left(-\frac{(\mu - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right). \quad (3.15)$$

Note how the normal distribution is indeed its own conjugate. Here, we defined the variance of the posterior as:

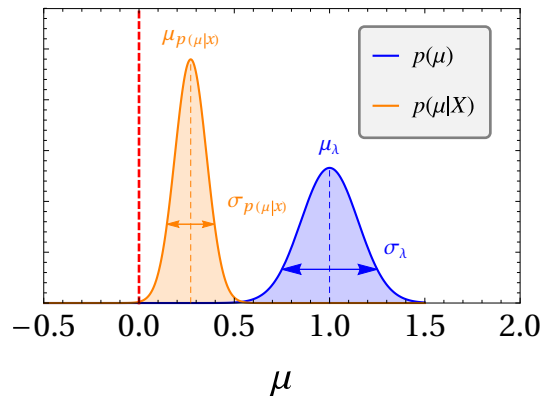
$$\tilde{\sigma}^2 := \frac{\sigma_p^2\sigma^2}{\sigma_p^2 + \sigma^2}, \quad (3.16)$$

and the mean of the posterior as:

$$\tilde{\mu} := \frac{\sigma^2\mu_p + \sigma_p^2X}{\sigma_p^2 + \sigma^2}. \quad (3.17)$$

This result has an intuitive interpretation: the

estimated mean is simply the average between the prior's mean  $\mu_p$  and the outcome  $X$ ,



**Figure 3.4:** Bayesian update of the normal distribution.

weighted by the variances. Essentially, the new mean is simply a "center of mass" between  $X$  and  $\mu_p$ , and the more they differ, the more the estimation is shifted towards  $X$ . It is also interesting to study the limiting case: if  $\sigma \rightarrow 0$  then  $\mu_p \rightarrow X$ . The reasoning behind this is that since the variance of the likelihood is small, it is very unlikely that  $X$  was sampled from a point far away from the mean, so we have a strong reason to believe that  $X$  itself is a good representative of the true average. See Fig. 3.4 for an illustration.  $\square$

To close this section, note the crucial difference from the frequentist paradigm in our example: earlier, we were mainly interested in quantities which are averaged over all possible stochastic realizations, such as the Fisher information in Eq. (2.28) or the MSE from Eq. (2.8). Here on the other hand, we can see that posterior is conditioned *on the data*, that is: different realizations obviously lead to different posteriors. In the classical case, we were very often conditioning over the *parameter* instead. This contrast in the approaches motivates many of the new definitions and results that we are going to encounter further into this chapter.

## 3.2 Bayesian estimators

We have seen in the last section how the Bayesian framework yield posterior distributions. These are simply probability densities containing information about the parameter of interest  $\theta$ . Here we will show that the Bayesian framework also admits point estimation strategies, in the same spirit of the classical estimation theory. In particular, the posterior distributions themselves can be used to construct concrete estimators. By doing so, one can condense the information contained in a distribution into a single random variable: the (Bayesian) estimator. This strategy is, of course, of great practical interest; specially due to the fact that we can take advantage of big part of the formalism constructed in Chapter 2 for frequentist estimation.

This framework requires two main ingredients. The first one is the use of Bayes theorem for constructing posteriors distributions. The second one, which we are going to introduce now, is the notion of Bayesian figures of error (or cost functions). For instance: now that the parameter  $\theta$  is *also* a random variable, how can one define the MSE in this situation? We are going to show that typical Bayesian estimators are defined as functions

which minimize some Bayesian figure of error with respect to the posterior distribution. In particular, the resulting classes of estimators depend on the type of cost function that is chosen.

More concretely, it is possible to define a cost, or loss, function  $C(\theta, \hat{\theta}(\mathbf{X}))$  which decides how to penalize the estimator  $\hat{\theta}$ . Note that  $C$  is a function of both the estimator  $\hat{\theta}$  and the parameter  $\theta$  (so it is also a random variable). We define the Bayesian error, also known as Bayesian loss, or risk, as the cost function integrated over  $p(\theta, \mathbf{x})$ :

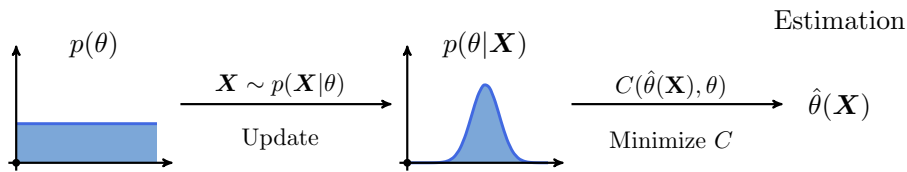
$$\mathcal{R} := \mathbb{E}_\theta \left[ \mathbb{E}_{\mathbf{X}} \left[ C(\theta, \hat{\theta}(\mathbf{X})) \right] \right] = \int d\theta \int C(\theta, \hat{\theta}(\mathbf{x})) p(\theta, \mathbf{x}) d\mathbf{x}. \quad (3.18)$$

By writing the joint pdf as  $p(\theta, \mathbf{X}) = p(\mathbf{X}|\theta)p(\theta)$ , we also have:

$$\mathcal{R} = \int p(\theta) d\theta \int C(\theta, \hat{\theta}(\mathbf{x})) p(\mathbf{x}|\theta) d\mathbf{x}. \quad (3.19)$$

We use  $\mathbb{E}_\theta [\cdot]$  and  $\mathbb{E}_{\mathbf{X}} [\cdot]$  to refer to the mean over the prior distribution and the stochastic average, respectively. By minimizing the error in Eq. (3.18) we obtain the *Bayesian estimator*. Note how this is equivalent to the integration of the frequentist error, such as the MSE from Eq. (2.8), over the prior. In other words, we are *averaging* the error over the parameter  $\theta$  and weighting it with respect to the prior: we give more importance to the values of  $\theta$  which are more likely to occur or to be true, and this is encoded precisely by the prior  $\theta$ . Deviations for unlikely values of  $\theta$  are penalized less, and vice-versa.

We illustrate this procedure in Fig. 3.5. As we will show in the following example, the procedure for obtaining a MMSE estimator in the Bayesian sense is much simpler than obtaining its analogue from the frequentist scenario.



**Figure 3.5:** Steps for obtaining Bayesian estimators.

**Example:** Let us investigate the MSE in the Bayesian framework. For that, we take the cost function  $C(\theta, \hat{\theta}(\mathbf{X})) = (\theta - \hat{\theta}(\mathbf{X}))^2$ . The Bayesian mean-squared error

(BMSE) reads:

$$\epsilon_B(\hat{\theta}) := \int p(\theta) d\theta \int (\theta - \hat{\theta}(\mathbf{X}))^2 p(\mathbf{x}|\theta) d\mathbf{x} = \mathbb{E}_\theta [\epsilon(\hat{\theta}|\theta)]. \quad (3.20)$$

Note how the BMSE is simply the frequentist MSE from Eq. (2.8) integrated over the prior. We find that the equation above is minimized if we take:

$$\left. \frac{d\epsilon_B(\hat{\theta})}{d\hat{\theta}} \right|_{\hat{\theta}=\hat{\theta}_{\text{BA}}} = 0 \implies \hat{\theta}_{\text{BA}}(\mathbf{X}) = \int \theta p(\theta|\mathbf{X}) d\theta. \quad (3.21)$$

The estimator above minimizes the MSE in the Bayesian sense. Note how this it is simply the mean value of the posterior distribution. For that reason, from hereafter we refer to  $\hat{\theta}_{\text{BA}}(\mathbf{X})$  either as the Bayesian average (BA) or as the posterior mean.

We also mention a few other common types of Bayesian estimators, showcasing the usefulness of this approach. In the case of the absolute error, we take the cost function  $C(\theta, \hat{\theta}(\mathbf{X})) = |\theta - \hat{\theta}(\mathbf{X})|$ . It can be shown that the corresponding Bayesian estimator  $\hat{\theta}_{\text{abs}}$  is the *median* of the posterior, which satisfies:

$$\int_{-\infty}^{\hat{\theta}_{\text{abs}}(\mathbf{X})} p(\theta|\mathbf{X}) d\theta = \int_{\hat{\theta}_{\text{abs}}(\mathbf{X})}^{\infty} p(\theta|\mathbf{X}) d\theta. \quad (3.22)$$

Finally, we mention a last important example. There exists a third estimator, which is also very standard, known as the *maximum a posteriori* (MAP) estimate. Rigorously defining the cost function which is minimized by the MAP is slightly difficult and we will omit the details here, but the proper construction can be found in Ref. [101]. In practice, this estimator is simply the the mode of the posterior:

$$\hat{\theta}_{\text{map}}(\mathbf{X}) = \arg \max_{\theta} p(\theta|\mathbf{X}). \quad (3.23)$$

One of the usefulness of this estimator, besides its simplicity, is that in certain conditions it agrees with the maximum likelihood estimate discussed in Sec. 2.6. In particular, we are going to argue in Sec. 3.3 that the MAP actually converges to the MLE in the asymptotic limit.  $\square$

The take away message from these examples is that each cost function has an associ-

ated estimator which minimizes it, per Fig. 3.5. For simplicity, we will focus mainly on the posterior mean; especially when discussing thermometry in later chapters. While the choice of a cost function may seem quite subjective, the MS estimator is actually also the optimal estimator for a wider class of cost functions. One such example happens when  $C(\theta, \hat{\theta}(\mathbf{X}))$  is simultaneously convex and symmetric [102] (p. 239). Thus, this estimator provides a good balance between generality and simplicity. This is an appropriate moment to employ these results in the practical example from the previous section. We will explicitly obtain the estimators, and their error, for the Bernoulli trials.

**Example:** We found that a for a prior distribution  $\mathcal{B}_{\alpha_0, \beta_0}(\theta)$ , the posterior becomes, after  $n$  Bernoulli trials:

$$p(\theta|\mathbf{X}) = \mathcal{B}_{\alpha_0+k, \beta_0+n-k}(\theta). \quad (3.24)$$

Here we wrote  $k := \sum_i X_i$ . How does the MAP and the BA estimates look like under the posterior distribution written above? By using the identity from Eq. (3.11) together with the definition in Eq. (3.10) we can show that the mean of the Beta distribution is given by  $\alpha/(\alpha + \beta)$ . Similarly, the mode is given by  $(\alpha - 1)/(\alpha + \beta - 2)$ . This means that the BA and the MAP evaluate to

$$\hat{\theta}_{\text{BA}}(\mathbf{X}) = \frac{\alpha_0 + k}{\alpha_0 + \beta_0 + n} \quad (3.25)$$

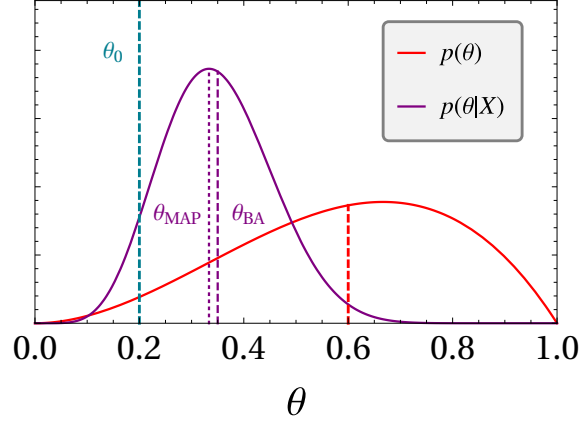
and

$$\hat{\theta}_{\text{MAP}}(\mathbf{X}) = \frac{\alpha_0 + k - 1}{\alpha_0 + \beta_0 + n - 2}, \quad (3.26)$$

respectively. We show a visual example in Fig. 3.6. The estimator in Eq. (3.25) has a particularly nice interpretation. It can be written as:

$$\frac{\alpha_0 + k}{\alpha_0 + \beta_0 + n} \equiv \frac{n}{\alpha_0 + \beta_0 + n} \left( \frac{k}{n} \right) + \frac{\alpha_0 + \beta_0}{\alpha_0 + \beta_0 + n} \left( \frac{\alpha_0}{\alpha_0 + \beta_0} \right), \quad (3.27)$$

which is the weighted average between the MLE and prior estimate [103] (pp. 42-43). To conclude this example, we will explicitly calculate the MSE for the posterior mean in Eq. (3.25). Using the properties for the mean and the variance of the Bernoulli distribu-



**Figure 3.6:** The *maximum a posteriori* and Bayesian average estimators for the Beta distribution. In this particular example we choose  $\mathcal{B}(3, 2)$  as the prior. The true value of the parameter is  $\theta_0 = 0.2$  and we update the prior with  $n = 15$  outcomes. In this particular realization we had  $k = 4$ , which yield the estimates  $\hat{\theta}_{\text{BA}}(\mathbf{X}) = 0.35$  and  $\hat{\theta}_{\text{BA}}(\mathbf{X}) = 1/3$ . The red dashed line shows the BA estimate of the prior.

tion, we find that the variance and the bias of the BA are given by:

$$\text{Var}(\hat{\theta}_{\text{BA}}) = \frac{1}{(\alpha_0 + \beta_0 + n)^2} \text{Var}\left(\sum_i X_i\right) = \frac{n\theta(1-\theta)}{(\alpha_0 + \beta_0 + n)^2} \quad (3.28)$$

and

$$b(\hat{\theta}_{\text{BA}}) = \frac{\alpha_0 + \mathbb{E}[\sum_i X_i]}{\alpha_0 + \beta_0 + n} - \theta = \frac{\alpha_0(1-\theta) - \beta_0\theta}{\alpha_0 + \beta_0 + n}, \quad (3.29)$$

respectively. Here we have used the definitions from Eqs. (2.10) and (2.11). Moreover, note how the BA is, in general, a biased estimator, as this example shows.

By using the bias-variance decomposition obtained in Eq. (2.9), we finally arrive at the (frequentist) MSE for the BA estimator:

$$\epsilon(\hat{\theta}_{\text{BA}}|\theta) = \frac{n\theta(1-\theta) + (\alpha_0(1-\theta) - \beta_0\theta)^2}{(\alpha_0 + \beta_0 + n)^2}. \quad (3.30)$$

We can now evaluate the *Bayesian MSE* the integrating the expression above over the prior. In the case of the flat prior, with  $\alpha_0 = \beta_0 = 1$ , we obtain:

$$\epsilon_B(\hat{\theta}_{\text{BA}}) = \int_0^1 p(\theta)\epsilon(\hat{\theta}_{\text{BA}}|\theta)d\theta = \frac{1}{6(n+2)}. \quad (3.31)$$

Now, let us compare this result with the MLE estimator from the last chapter, which is

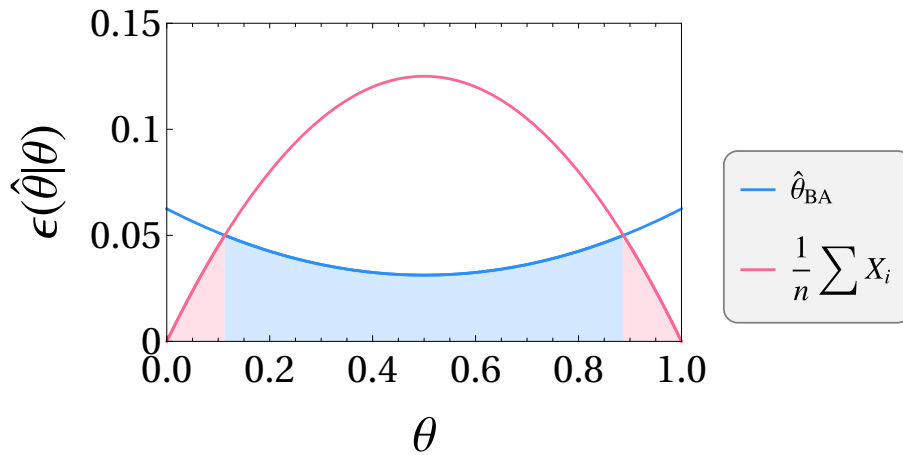
just the sample mean. Note that the MLE in this case is simply the MAP for a flat prior. The estimator in this case was unbiased and the MSE was given by Eq. (2.47):

$$\epsilon(\hat{\theta}_{\text{MLE}}|\theta) = \frac{\theta(1-\theta)}{n}. \quad (3.32)$$

This means that the BMSE in this case evaluates to:

$$\epsilon_B(\hat{\theta}_{\text{MLE}}) = \frac{1}{6n}. \quad (3.33)$$

Note how the BMSE of the BA is slightly lower than the one for the sample mean.



**Figure 3.7:** MSE for the Beta distribution with  $n = 2$ . We plot the results for the posterior mean (blue) and the sample mean (magenta) estimators. We shade the plot according to the optimal estimator. The blue (magenta)-shaded region corresponds to the interval where the BA (sample mean) estimator is better.

We also plot the results for the MSE of the BA and the MLE estimators in Fig. 3.7. This picture give us some intuition about the relevance of the different definitions of error that we have introduced throughout the last two chapters. First, note how the optimal estimator among these two choices depends on the particular value of  $\theta$ . The usefulness of the Bayesian error in this case is that it encapsulates the optimality of these estimates into a single figure: the BMSE of the BA is lower due to the fact that, *on average*, it performs better than the MLE. In particular, while the MLE is better for values around  $\theta = 0$  or  $\theta = 1$ , the BA works better for intermediate values of  $\theta$ .  $\square$

We just saw how to obtain concrete estimators in this example by means of Bayesian

tools. These results are, of course, a further merit of this approach. As we mentioned in the last chapter, MMSE estimators, or even MVU estimators, are very often unrealizable. If one consider these figures of error in the Bayesian sense instead, achieving MMSE estimators, or similar, is perfectly possible in general [102] (p. 239). Thus, we abandon the task of finding the best estimator for *every*  $\theta$  and we focus instead on finding the best estimator in the Bayesian sense, as we defined in Eqs. (3.18) and (3.19). By doing so, we obtain estimators which are good on *average*.

This formulation results in a sensible strategy in the scenario where, e.g., one is initially very ignorant about  $\theta$ , so it might be hard to choose a reasonable estimator which works well for the true value of the parameter. Of course, this example also serves to further stress that Bayesian and frequentist paradigms should not be seen as entirely antagonistic approaches, and neither that one of these formulations is necessarily better than the other. Their techniques might be actually very analogous or complementary at times: the last example just showed us how one can use Bayesian techniques as the starting point to construct concrete estimators. The BA and MAP estimators are perfectly valid estimators even beyond the Bayesian setting, as we have even calculated their MSE in the frequentist sense. Thus, even if one is not interested in the Bayesian interpretation of the data, it might still be useful to consider the Bayes theorem as a starting point and as a *tool* for obtaining concrete estimators.

The Bayesian framework and its notion of integrated error is also useful for, e.g., formalizing the notion of ordering of the risk, in consonance with much of our discussions throughout Sec. 2.3 and 2.7 regarding the optimality of estimators in a purely frequentist sense. A very famous problem, known as Stein's example results in what is known as the James-Stein estimator [104, 105]. This is a scenario where a biased estimator is *always* better than the MLE estimator and where we can make use of, e.g., the idea of *admissibility* to compare them [98].

We mention that some authors might not consider Eq. (3.18) a concept which is *entirely* Bayesian. A purely Bayesian loss is conditioned on the data instead, as we defined in Eq. (2.3). Thus, we could rather define a *posterior* loss

$$\epsilon(\hat{\theta}|\mathbf{X}) := \int C(\theta, \hat{\theta}(\mathbf{X}))p(\theta|\mathbf{X})d\theta \quad (3.34)$$



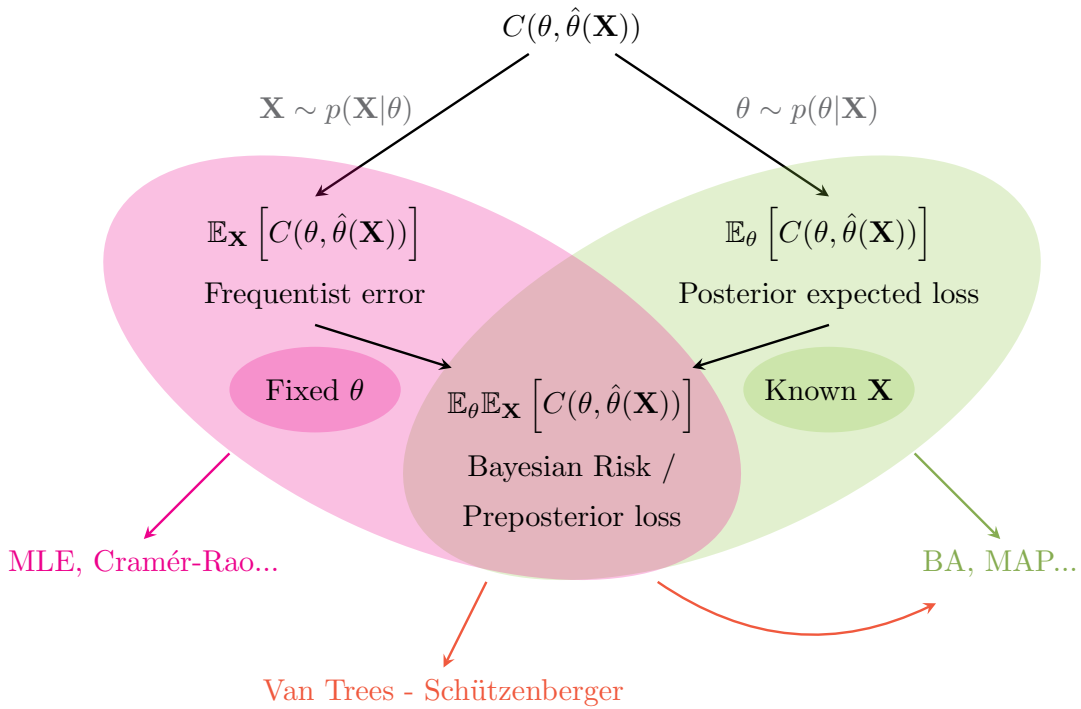
which disregards other realizations and is conditioned on the measured data  $\mathbf{X}$ . Thus, what we call the Bayesian error in this work, per Eq. (3.18), is sometimes called *preposterior* loss or integrated risk. It is an error which is averaged over *all* possible realizations, while Eq. (3.34) takes only the observed data into consideration. Therefore, the preposterior loss is the error that we expect to obtain *before* we perform any measurements. Hence, one could argue that Eq. (3.18) is useful in numerical or theoretical investigations where the experimenter does not have any data at hand. In this case he or she would like to have a general picture of what type of error should be expected for different realizations. With the data in hands, it is often more useful to use the posterior loss instead. Fortunately, this difference is not very significant when discussing estimators, since minimizing either the posterior or preposterior losses in Eqs. (3.18) and (3.34) is equivalent and yields the same estimators [97] (p. 62-63). Our current results are valid even in this scenario.

Dwelling into the philosophical and technical implications of the Bayesian and frequentist theories is well beyond the scope of this work [106–108]. For that, many appropriate references can be found. Refs. [58] and [97] discuss many of the implications and distinctions between the two schools of statistics. A special attention to this topic is given in Ref. [109].

Furthermore, obtaining appropriate priors and a sensible loss function, which we considered to be some of the main ingredients for inference, is another very important aspect of the Bayesian theory. This is specially true when one is interested in employing a fully Bayesian approach. In those scenarios, it is much more important to be careful with such details. We omit these topics since this is not the focus of this dissertation. However, we mention a few techniques which are standard. We have, throughout this chapter, used flat priors as a synonym of uninformative priors. This is not strictly true in general. Many models might require more sophisticated distributions to better convey the idea of ignorance. Among possible techniques to derive them, the idea of Jeffrey's priors and the maximum entropy principle [58] (pp. 181-183, 372-378) configure some of the main techniques. Sometimes, it is possible to obtain an appropriate prior by analyzing the symmetries and constraints of problem [110] (p. 366), such as scale estimation and phase estimation problems [58] (pp. 378-394). Thus, it is clear that assigning appropriate priors and error functions is a whole topic of study on itself. A self-contained introduction can

be found in Ref. [97] (Chapter 3.5).

While we are not going to directly employ most of these concepts in the development of our original results, we have nevertheless mentioned them in this thesis for completeness. This is motivated by the fact that part of the current literature on quantum thermometry borrows from many of these ideas: a certain familiarity with some of these topics might be useful in keeping up to date with the bibliography we are concerned with. We summarize this discussion, and the different definitions of error, such as Eqs. (2.2), (3.18) and (3.34) in Fig. 3.8. We introduce the remaining concept from the figure, the Van Trees-Schützenberger inequality, in the coming section.



**Figure 3.8:** A diagram of the summary of different inference approaches.

### 3.3 The Van Trees-Schützenberger inequality

As we saw, due to the gap between the Bayesianism and the frequentism, transposing some concepts from one paradigm to another requires some care. In particular, it is not immediately clear how the Cramér-Rao bound should look like in terms of an integrated loss function, in the sense of Eq. (3.20). Since the classical CRB is essentially a frequen-

tist concept, it naturally depends on the parameter of interest in general. Not only that, but the bound is also unable to capture any of the information contained in the prior distribution. In this chapter, we discuss an extension to the Bayesian case. This result was independently obtained by Van Trees [102] in 1968 and by Schützenberger [111] a few years earlier. We refer to their work as the Van Trees-Schützenberger bound (VTSB). This approach avoids one of the conundrums of the usual CRB; the bound is no longer dependent on the parameter being estimated. Instead, we obtain an inequality which depends only on the estimator and the prior distribution.

As it turns out, we can further adapt the Cramér-Rao inequality as a Bayesian bound through the use of the Cauchy-Schwarz theorem and a clever choice of joint densities. We can then turn it into an inequality for the MSE while also embedding information contained within the prior. This derivation is of very similar spirit to the one for the CRB in Sec. 2.5. The regularity conditions required for this bound [102] (p. 261), together with some generalizations and applications, can be found in Ref. [112].

First, we shall assume that integration and derivation are interchangeable, in the sense that:

$$\int \frac{\partial p(x|\theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int p(x|\theta) dx = 0. \quad (3.35)$$

We will also assume that the prior density  $p(\theta)$  vanishes at the limits of integration. By doing so, we can use integration by parts to show that the following identity holds:

$$\int \theta \frac{\partial (p(x|\theta)p(\theta))}{\partial \theta} d\theta = - \int p(x|\theta) d\theta, \quad (3.36)$$

and, by analogy, these conditions also imply that:

$$\int \frac{\partial (p(x|\theta)p(\theta))}{\partial \theta} d\theta = 0. \quad (3.37)$$

Since the estimator  $\hat{\theta}(X)$  is *not* a function of the random parameter itself, we can multiply the integral above by  $\hat{\theta}$ , subtract by Eq. (3.36) and then integrate with respect to  $x$  on both sides to obtain:

$$\int \int (\hat{\theta}(x) - \theta) \frac{\partial (p(x|\theta)p(\theta))}{\partial \theta} p(\theta) d\theta dx = \int \int p(x|\theta) dx d\theta = 1. \quad (3.38)$$

Note however that we can write the derivative above as:

$$\frac{\partial(p(x|\theta)p(\theta))}{\partial\theta} = \frac{\partial(p(x|\theta)p(\theta))}{\partial\theta} \frac{p(x|\theta)p(\theta)}{p(x|\theta)p(\theta)} = p(x|\theta)p(\theta) \frac{\partial \ln(p(x|\theta)p(\theta))}{\partial\theta},$$

so that the previous equation becomes:

$$\int \int (\hat{\theta}(x) - \theta) \frac{\partial \ln(p(x|\theta)p(\theta))}{\partial\theta} p(x|\theta)p(\theta) d\theta dx = 1 \quad (3.39)$$

The LHS can be written as an expectation value:

$$\mathbb{E}_\theta \mathbb{E}_X \left[ (\hat{\theta}(X) - \theta) \frac{\partial \ln(p(x|\theta)p(\theta))}{\partial\theta} \right] = 1. \quad (3.40)$$

Using the Cauchy-Schwarz inequality, in analogy to the derivation from Sec. 2.5, we can rewrite this as:

$$\mathbb{E}_\theta \mathbb{E}_X \left[ (\hat{\theta}(X) - \theta)^2 \right] \mathbb{E}_\theta \mathbb{E}_X \left[ \left( \frac{\partial \ln(p(x|\theta)p(\theta))}{\partial\theta} \right)^2 \right] \geq 1. \quad (3.41)$$

By using the product rule for logarithms, we can rewrite the expectation in Eq. (3.41) as:

$$\begin{aligned} \mathbb{E}_\theta \mathbb{E}_X \left[ \left( \frac{\partial \ln(p(x|\theta)p(\theta))}{\partial\theta} \right)^2 \right] &= \mathbb{E}_\theta \mathbb{E}_X \left[ \left( \frac{\partial \ln p(x|\theta)}{\partial\theta} \right)^2 \right] \\ &\quad + 2 \mathbb{E}_\theta \mathbb{E}_X \left[ \left( \frac{\partial \ln p(x|\theta)}{\partial\theta} \right) \left( \frac{\partial \ln p(\theta)}{\partial\theta} \right) \right] \\ &\quad + \mathbb{E}_\theta \mathbb{E}_X \left[ \left( \frac{\partial \ln p(\theta)}{\partial\theta} \right)^2 \right], \end{aligned} \quad (3.42)$$

where we can use the condition (3.35) to verify that the middle term vanishes:

$$\mathbb{E}_\theta \mathbb{E}_X \left[ \left( \frac{\partial \ln p(x|\theta)}{\partial\theta} \right) \left( \frac{\partial \ln p(\theta)}{\partial\theta} \right) \right] = 0.$$

Now, we can see that the inner expectation in the first term of Eq. (3.42) is simply the Fisher Information of the likelihood. Hence, the first expectation in Eq. (3.42) is simply

the Fisher information averaged over the prior:

$$\mathbb{E}_\theta \mathbb{E}_X \left[ \left( \frac{\partial \ln p(x|\theta)}{\partial \theta} \right)^2 \right] = \mathbb{E}_\theta [F(\theta)] \quad (3.43)$$

Meanwhile, the expectation of the last term in Eq. (3.42) with respect to  $X$  is trivial, since it only depends on  $\theta$ . Therefore, it results in an expression for the Fisher information of the prior distribution, which we denote by:

$$F_p := \mathbb{E}_\theta \left[ \left( \frac{\partial \ln p(\theta)}{\partial \theta} \right)^2 \right]. \quad (3.44)$$

Thus, having identified all the tree terms, we can finally use the inequality (3.41) to arrive at the desired result:

#### The Van Trees-Schützenberger inequality

The Bayesian mean square error (3.20) of any estimator  $\hat{\theta} = \hat{\theta}(X)$  of  $\theta$ , with  $X \sim p(X|\theta)$ , is bounded by the inequality

$$\epsilon_B = \mathbb{E}_\theta \mathbb{E}_X [(\hat{\theta} - \theta)^2] \geq \frac{1}{\mathbb{E}_\theta [F(\theta)] + F_p}, \quad (3.45)$$

given the prior distribution  $p(\theta)$  and the appropriate regularity conditions. In analogy to Eq. (2.43), one has

$$\epsilon_B \geq \frac{1}{n \mathbb{E}_\theta [F(\theta)] + F_p} \quad (3.46)$$

for a sequence  $\mathbf{X} = X_1, \dots, X_n$  of i.i.d. outcomes.

This bound is quite general in the sense that, differently from the standard CRB in Eq. (2.42) for *unbiased* estimators, no requirement about unbiasedness was made here. Moreover, we can immediately see that the VTSB does *not* depend on  $\theta$ . The FI of the likelihood is actually averaged over the prior. This is an upside of this result; we can obtain the limits of precision of an estimator independently  $\theta$ . Additionally, when comparing this inequality with the usual CRB we immediately see an extra term which accounts for the prior knowledge about the parameter.

Of course, this does not mean that the VTSB is more useful than the CRB. Ultimately, they both tell us very different things: the CRB is an inequality for the MSE given by

Eq. (2.8) in the frequentist sense. This bound is of course much more appropriate when one is interested in the precision of the estimator for a particular value of  $\theta$ . Thus, it might be reasonable to say that some information about the maximum achievable precision is "lost" when one integrates over the prior to use the VTSB. And the VTSB is valid, of course, only for the *Bayesian* analogue of the MSE in Eq. (3.20). For instance, maybe a certain estimator performs very well for a certain range parameters and very bad for others. By using the Bayesian risk and the VTSB we would have information of their Bayesian averages only, and not of specific parametrizations. Of course, it is also possible to argue that this is not so bad because the VTSB can incorporate a localized knowledge about  $\theta$  pretty well through  $F_p$  and the expectation over the FI anyway. Nevertheless, both approaches are just different perspectives on how to *present* the data and the associated bounds. It is on the hands of the theoretician and the experimentalist which of these two presentations would yield better interpretations for his or her particular problem; in some sense, the two constructions are just showing the same information in different ways.

More strikingly, and also going back to technical aspects, while the CRB is very often achievable in the asymptotic limit for, e.g., maximum likelihood estimators (as discussed in Sec. 2.6), the VTSB is not tight in general. This happens due to the asymptotic behavior of the posterior distribution and the structure of the VTSB. For independent outcomes, a result known as the Bernstein-von Mises theorem [113, 114] assures that the posterior converges, in the limit of large  $n$ , to a Gaussian whose mean is centered around the true value of the parameter  $\theta_0$  with variance  $1/nF(\theta_0)$ . In symbols,

$$P(\theta|\mathbf{X}) = \sqrt{\frac{nF(\theta_0)}{2\pi}} e^{-\frac{nF(\theta_0)(\theta-\theta_0)^2}{2}}, \quad (\text{large } n). \quad (3.47)$$

This is very similar to the asymptotic properties of the MLE which we mentioned in Sec. 2.6. A consequence of the result above is that the MSE scales with  $1/nF(\theta)$  in the asymptotic limit. This implies that the BMSE scales as

$$\epsilon_B(\hat{\theta}) \sim \mathbb{E}_\theta \left[ \frac{1}{nF(\theta)} \right], \quad (\text{large } n, \text{ MSE}). \quad (3.48)$$

In other words, the asymptotic BMSE scales simply with  $1/nF(\theta)$  averaged over the

prior. Meanwhile, the VTSB scales with

$$\epsilon_B(\hat{\theta}) \sim \frac{1}{\mathbb{E}_\theta [nF(\theta)]}, \quad (\text{large } n, \text{ VTSB}), \quad (3.49)$$

in the limit of many measurements. Thus, as a consequence of Jensen's inequality [60] we have that  $\mathbb{E}_\theta \left[ \frac{1}{F(\theta)} \right] \geq \frac{1}{\mathbb{E}_\theta [F(\theta)]}$ , meaning that the VTSB is, in general, not tight [102] (p. 265). In summary, this means that estimators which are asymptotically efficient, such as the MLE and the MAP, will converge to the *expected* Cramér-Rao bound, but not to the VTSB.

An elementary proof of the the result by Bernstein and von Mises [115, 116] can be sketched in terms of the Laplace approximation [117], which is often used to approximate posterior distributions [118]. A short and comprehensive summary of the method can be found in Ref. [119] (p. 341). A rigorous proof, with consideration for the appropriate regularity conditions, can be found in more technical texts [113, 120].

We also mention a generalization which eliminates the requirement for a prior with bounded supports where the distribution does not vanish at the end points [121]. This result might be useful when one is interested in, e.g., truncated Beta distributions. Needless to say, the VTSB is not the unique Bayesian lower bound of interest. Research concerning alternative, and possibly tighter, bounds is also active [122]. A compilation of several works concerned with Bayesian bounds can be found in Ref. [96]. We conclude this section with an example.

▣ **Example:** In the problem of estimating the mean of a normal distribution, we were able to obtain the mean of the posterior distribution in Eq. (3.17). Because the posterior is a Gaussian, the MAP and the BA naturally coincide, and the single-shot estimator is given by:

$$\hat{\mu}_{BA} = \hat{\mu}_{MAP} = \frac{\sigma_p^2 X + \sigma^2 \mu_p}{\sigma_p^2 + \sigma^2}. \quad (3.50)$$

The *squared* error associated with the estimator above will be:

$$(\hat{\mu}_{BA} - \mu)^2 = \frac{\sigma_p^4(x - \mu)^2 + 2\sigma_p^2\sigma^2(x - \mu)(\mu_p - \mu) + \sigma^4(\mu_p - \mu)^2}{(\sigma_p^2 + \sigma^2)^2}. \quad (3.51)$$

Now, per the definition (3.20), we are expected to take the stochastic average (with respect

to  $X$ ) in order to calculate the BMSE. We first should note that:

$$\mathbb{E}_X[X - \mu] = \mathbb{E}_X[X] - \mu = 0 \quad (3.52)$$

This means that the second term in the RHS of the equation above vanishes, and we also have that

$$\mathbb{E}_X[(X - \mu)^2] = \sigma^2, \quad (3.53)$$

so:

$$\mathbb{E}_X[(\hat{\mu}_{BA} - \mu)^2] = \frac{\sigma_p^4 \sigma^2 + \sigma^4 (\mu_p - \mu)^2}{(\sigma_p^2 + \sigma^2)^2}. \quad (3.54)$$

Now what remains is to take the average over the prior. The first term is independent of  $\mu$ , so we may as well leave it alone. The second term on the other hand depends on the parameter  $\mu$  we want to estimate, so we can verify that:

$$\mathbb{E}_\mu[(\mu_p - \mu)^2] = \int (\mu - \mu_p)^2 p(\mu) d\mu = \int (\mu_p - \mu)^2 \frac{e^{-\frac{(\mu - \mu_p)^2}{2\sigma_p^2}}}{\sqrt{2\pi}\sigma_p} d\mu = \sigma_p^2. \quad (3.55)$$

Here, we explicitly wrote the variance calculation as an integral for clarity. Now, we can finally write the BMSE error for this estimation procedure:

$$\mathbb{E}_\mu \mathbb{E}_X[(\hat{\mu}_{BA} - \mu)^2] = \frac{\sigma_p^4 \sigma^2 + \sigma_p^2 \sigma^4}{(\sigma_p^2 + \sigma^2)^2} = \frac{\sigma_p^2 \sigma^2}{\sigma_p^2 + \sigma^2}. \quad (3.56)$$

Finally, we can compare the result above with the Van Trees-Schützenberger inequality. The Fisher Information for the prior will be  $F_p = \frac{1}{\sigma_p^2}$ . Similarly, we have  $F(\mu) = \frac{1}{\sigma^2}$  for the likelihood. Since it does not depend on the parameter  $\mu$  we are estimating, it is trivial to check that

$$\mathbb{E}_\mu(F(\mu)) = \frac{1}{\sigma^2}. \quad (3.57)$$

Hence, it is easy to see that:

$$\frac{1}{\mathbb{E}_p(F(\mu)) + F_p} = \frac{\sigma_p^2 \sigma^2}{\sigma_p^2 + \sigma^2} \quad (3.58)$$



Therefore, we can finally verify that:

$$\mathbb{E}_\mu \mathbb{E}_X[(\hat{\mu}_{MAP} - \mu)^2] = \frac{1}{\mathbb{E}_\mu(F(\mu)) + F(p)} = \frac{\sigma_p^2 \sigma^2}{\sigma_p^2 + \sigma^2} \quad (3.59)$$

Thus the MAP and the BA also *saturate* the VTSB in this case. This happens precisely due to the fact that the Fisher information  $F(\mu)$  [Eq. (2.28)] depends only on the known  $\sigma$ , and not on the unknown parameter  $\mu$ , as we mentioned.  $\square$

## Chapter 4

# Quantum metrology and thermometry

Quantum metrology is the field which is concerned with measurement protocols of high precision [123–125] and with the problem of estimation theory [126] in quantum mechanics. In particular, much of the current advances and the literature are either preoccupied with the theoretical and experimental idiosyncrasies of the quantum theory in this context [127], or with the use of quantum resources and how they can bring quantum advantages in metrological settings [128–130].

Quantum advantages, in a very broad sense, have since long been discussed. Ubiquitous instances of this are the closely related field of quantum optics and [131], and also in the field of quantum algorithms since the 1990s [132], albeit in a very different context. An omnipresent topic which, even though we are not going to focus on, serves as an important example, is the idea of achieving the *Heisenberg limit*, where one can obtain a scaling of  $1/n^2$  for the variance of the estimation by means of quantum protocols, in contrast with, e.g., the usual bound of  $1/n$  for the CRB, arising from the classical estimation theory [133–136]. The latter is often called the *the standard quantum limit* (SQL).

Such type of improvement might be explored through entanglement [137, 138], quantum criticality [139–141] and squeezing [142, 143], to mention a few instances. A typical implementation consists in interferometric setup which makes use of highly entangled states, such as NOON states or GHZ states [144]. See Ref. [145] (p. 889-891) for the definition and further discussion about such entangled states. While this type of investigation is not exactly our focus in Chapter 5, they are naturally a great motivation for the underlying framework discussed in the upcoming sections.

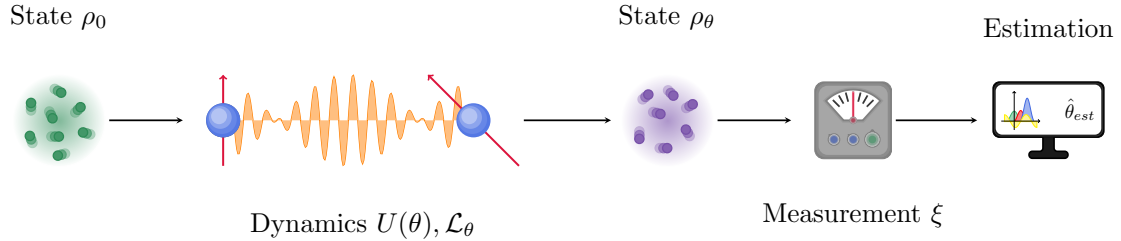
On a historical note, we cite the canonical works by Personick [146], Helstrom [147, 148] and Holevo [149] as early and very important contributions to the field, which were vital in kick-starting the foundations of quantum estimation theory. Furthermore, we also mention Giovannetti, Lloyd and Maccone's work as one of the earliest accounts of a more modern formulation of the field, which was central in shaping the current landscape for quantum metrology [150]. Further examples and applications of quantum metrology can be found in Ref. [151].

Needless to say, the evident usefulness of quantum metrology makes it a powerful instrument in several different fields, ranging from biology [152] to photonics [153]. Our focus in this dissertation is the application of this framework in the context of quantum thermodynamics, in a subfield known as *quantum thermometry* (see Ref. [154] for a review), which is concerned with the particular problem of temperature estimation. We introduce this topic in Sec. 4.3, following with a review of a concrete and recently proposed [155] implementation in Sec. 4.4.

## 4.1 The quantum Fisher information

We start this section by describing the standard steps in a typical metrology experiment, as illustrated by Fig. 4.1. We then introduce the relevant definitions, discussing the idea behind quantum metrology and quantum estimation theory. For that, let  $\rho$  be the density matrix describing the state of a (possibly mixed) system. One usually starts with an input (or probe) state denoted by  $\rho_0$ . A parameter of interest  $\theta$ , which can be, e.g., a phase or a frequency, is encoded into an output state  $\rho_\theta$  by means of a dynamical evolution. This embedding of the parameter is very often done through a unitary evolution  $U(\theta)$ , a superoperator  $\mathcal{L}_\theta$  describing a dissipative evolution or even through the measurement back-action described by Kraus operators [156]. Here, we will focus on protocols where the final state  $\rho_\theta$  is measured after unitary and dissipative dynamics. This measurement is described by a set of positive operator-valued measures (POVMs) [157], denoted by  $\{M_X\}$ . These measurements are then associated with outcomes  $X$  and, by the Born rule, with a probability distribution  $p(X|\theta) = \text{tr}\{M_X\rho_\theta\}$ . This pdf essentially plays the role of the likelihood that we have encountered in statistics and estimation theory. The empirical

data and the theoretical distribution  $p(X|\theta)$  can then be used to perform estimations about the parameter  $\theta$ , in the same spirit of point estimation theory and Bayesian inference.



**Figure 4.1:** Typical protocol in quantum metrology.

Of course, the discussion above shows that the quantum realm introduces an extra complication: we now have the freedom of selecting the measurement basis, or more generally, the POVMs. This means that the chosen POVM  $M$  defines the probability distribution associated with the measurements. These distributions might each have different degrees of sensibility upon the parameter, meaning that one should carefully consider how to perform a measurement: some choices might result in a better distinguishability within the neighbourhood of states  $\rho_\theta$ . Ideally, this suggests a maximization of the Fisher information over all possible measurements. The resulting quantity is known as the *quantum Fisher information*

$$\mathcal{F}(\theta) := \max_M F(\theta; M), \quad (4.1)$$

where  $F(\theta; M)$  is the FI associated to the distribution  $p(X|\theta) = \text{tr}\{M_X \rho_\theta\}$ .

The definition above is very intuitive on its meaning. However, maximizing over operators is, in general, not trivial at all. Fortunately, it is possible to arrive in a closed form expression for the optimal POVMs and for the QFI. For that, we closely follow Ref. [11], which provides a rich discussion about quantum metrology, furnishing us with some useful operational results. Moreover, we also include a result from Ref. [158], which consists in a very efficient trick trick for calculating the QFI matrix based on the vectorization process (see the Appendix A).

We start by defining an operator  $\Lambda_\theta$  known as the *Symmetric Logarithmic Derivative*

(SLD), which is a self-adjoint operator satisfying a Lyapunov equation of the form <sup>1</sup>

$$\Lambda_\theta \rho_\theta + \rho_\theta \Lambda_\theta = 2\partial_\theta \rho_\theta. \quad (4.2)$$

If we derive the expression for the probabilities for the Born rule with respect to  $\theta$

$$\partial_\theta p(X|\theta) = \text{Tr}\{(\partial_\theta \rho_\theta) M_x\}, \quad (4.3)$$

we can simply plug Eq. (4.2) in the equation above to find:

$$F(\theta; M) = \int dx \frac{\text{Re}(\text{Tr}\{\rho_\theta M_x \Lambda_\theta\})^2}{\text{Tr}\{\rho_\theta M_x\}}. \quad (4.4)$$

Here we have used the cyclic property of the trace and the identity  $\text{Tr}\{A_1 A_2 \dots A_n\} = (\text{Tr}\{A_n \dots A_2 A_1\})^*$  for Hermitian operators  $A_1, \dots, A_n$  [148] (p. 46). In the case of discrete outcomes, the integral above can be substituted by a summation sign. <sup>2</sup>

Proving the quantum analogue of the CRB is very similar to our derivations of the standard CRB in Chapter 2 and to the derivation of the VTSB in Chapter 3, where we used the Cauchy-Schwarz inequality in both cases, based on the standard inner product for integrals. Here, we consider instead the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product (See the Appendix A once again), which is given by  $|\text{Tr}\{A^\dagger B\}|^2 \leq \text{Tr}\{A^\dagger A\} \text{Tr}\{B^\dagger B\}$ . Thus, our objective is to bound the expression for the "measurement-based" FI in Eq. (4.9) by its maximum (over the space of POVMs). By doing so, we will be able to obtain an operational equation for the QFI, which we defined in Eq. (4.1).

---

<sup>1</sup>For a more detailed derivation of the upcoming results and a better motivation of the definitions of the SLD, see Refs. [146] and [148] (p. 266).

<sup>2</sup>To prove this, note that we can write  $\text{Tr}\{A_1 A_2 \dots A_n\} = \text{Tr}\{(A_n^\dagger \dots A_2^\dagger A_1^\dagger)^\dagger\} = \text{Tr}\{(A_n \dots A_2 A_1)^\dagger\} = (\text{Tr}\{A_n \dots A_2 A_1\})^*$ , since the operators inside the trace are all hermitian. Naturally,  $\rho_\theta$ ,  $\Lambda_\theta$  and  $M_x$  are all hermitian as well, so this property holds in the derivation of Eq. (4.9).

Algebraically, what we do is:

$$\begin{aligned}
 F(\theta; M) &\leq \int dx \left| \frac{\text{Tr}\{(\partial_\theta \rho_\theta) M_x \Lambda_\theta\}}{\sqrt{\text{Tr}\{\rho_\theta M_x\}}} \right|^2 \\
 &= \int dx \left| \text{Tr} \left\{ \frac{\sqrt{\rho_\theta} \sqrt{M_x}}{\sqrt{\text{Tr}\{\rho_\theta M_x\}}} \sqrt{M_x} \lambda_\theta \sqrt{\rho_\theta} \right\} \right|^2 \\
 &\leq \int dx \text{Tr}\{M_x \Lambda_\theta \rho_\theta \Lambda_\theta\} \\
 &= \text{Tr}\{\rho_\theta \Lambda_\theta^2\}.
 \end{aligned} \tag{4.5}$$

Note that we can bound the FI by the RHS in the first line simply because  $\text{Re}(\cdot)^2 \leq |\cdot|^2$ . Furthermore, in the second inequality we chose  $A^\dagger = \sqrt{\rho_\theta} \sqrt{M_x} / \sqrt{\text{Tr}\{\rho_\theta M_x\}}$  and  $B = \sqrt{M_x} \Lambda_\theta \sqrt{\rho_\theta}$ . Note how  $\text{Tr}\{A^\dagger A\} = \mathbf{I}$ , with these choices. Finally, in the last line we have used the "normalization" property of the POVM, that is,  $\int M_x dx = \mathbf{I}$ . The last expression in the equation above is precisely the QFI. In other words, we can write that the FI associated with the distribution of the measurements of  $\rho_\theta$ , with respect to the set of POVMs  $\{M_x\}$  is bounded by the quantum Fisher Information:

$$F(\theta; M) \leq \mathcal{F}(\theta) = \max_M F(\theta; M) = \text{Tr}\{\rho_\theta \Lambda_\theta^2\}. \tag{4.6}$$

If we use the QFI in conjunction with the CRB, we arrive at the quantum Cramér-Rao bound (QCRB):

$$\text{Var} \hat{\theta}(\mathbf{X}; M) \geq \frac{1}{\mathcal{F}(\theta)}. \tag{4.7}$$

In the case of  $n$  independent measurements where no correlations are present, we can simply include a factor of  $n$  the denominator of the RHS of the equation above, in analogy with the usual CRB. We included the POVM  $M$  in Eqs. 4.9 and 4.7 to make it explicit that the estimators (and the underlying pdf) *do* depend on the POVMs.

Finally, we comment on the attainability of the QFI, which translates into the choice of an optimal POVM (or measurement basis). The equality in Eq. (4.6) is satisfied if we choose the POVMs as the set  $\{|\Lambda_\theta\rangle_x \langle \Lambda_\theta|_x\}$ . That is, the optimal POVMs are projectors constructed from the eigenvectors of the SLD [11]. Refer also to [159] (p.25). To see this,

note that if calculate the trace appearing in the QFI we have:

$$\text{Tr}\{\rho\Lambda_\theta^2\} = \int \lambda_x^2 \langle \Lambda_\theta |_x \rho_\theta | \Lambda_\theta \rangle_x dx, \quad (4.8)$$

where we have used the spectral decomposition for the SDL, with  $\lambda_x$  representing its eigenvalues. Analogously, if we take  $M_x = |\Lambda_\theta\rangle_x \langle \Lambda_\theta|_x$  then Eq. (4.9) also becomes:

$$F(\theta; |\Lambda_\theta\rangle_x \langle \Lambda_\theta|_x) = \int dx \frac{\text{Re}(\text{Tr}\{\rho_\theta | \Lambda_\theta\rangle_x \langle \Lambda_\theta|_x \Lambda_\theta\})^2}{\langle \Lambda_\theta |_x \rho_\theta | \Lambda_\theta \rangle_x} = \int \lambda_x^2 \langle \Lambda_\theta |_x \rho_\theta | \Lambda_\theta \rangle_x dx. \quad (4.9)$$

where we use that  $\lambda_x$  is real (because the SLD is hermitian) and that  $|\Lambda_\theta\rangle_x \langle \Lambda_\theta|_x \Lambda_\theta = \lambda_x |\Lambda_\theta\rangle_x \langle \Lambda_\theta|_x$ . Showing that this choice of POVM indeed achieves the QFI.

We also provide a short summary here:

- In order to calculate the QFI in Eq. (4.6) the SLD must be known. It is obtained by solving Eq. (4.2) for  $\Lambda_\theta$ .
- Therefore, the QFI depends *only* on the state  $\rho_\theta$  and its parametrization.
- Since the SLD depends on  $\theta$ , this means that the optimal POVM might also depend on  $\theta$  as well, which is *unknown*. We refer back to this point at the end of Sec. (4.2).

We treat the operational aspects of the QFI and the SLD in Sec. 4.2, where we show a simple way of obtaining the QFI and the optimal POVMs based on the technique of Ref. [158]. Older and conventional approaches are based on, e.g., the diagonalization of the density matrix [11, 148].

The second point is also intuitive. The QFI essentially depends on the "statical manifold", or distinguishability of the states, around the parameter  $\theta$ . There is actually a deep connection between quantum estimation theory with geometrical aspects. See, for example, [126] for an earlier work. A very through review on this topic can be found Ref. [160]. Moreover, Ref. [11] also briefly discusses the connection of the QFI with the Bures metric. A few recent publications discuss some further aspects of these geometrical links [161, 162], including a very relationship between the Berry curvature and the QFI [163]. Note that this also explains why the notion of a Fisher information, or of a quantum version of a CRB, is slightly more complicated than the formulation from standard estimation theory. In the latter, we only consider probability distributions, which are

much simpler than quantum states. The mathematical structure of quantum mechanics introduces non-commuting operators and also the idea of measurements, which makes the structure of the problem much richer and more complex. The reader may check any of the references above for a rigorous discussion on this topic.

## 4.2 A trick for calculating the QFI

In this section we discuss the procedure described in Ref. [158], which will aid us in obtaining the SLD from Eq. (4.2) and the QFI from Eq. (4.6). This trick is based on the vectorization technique, which we discuss in Appendix A. To start, note that we can express the first term on the LHS in Eq. (4.2) as a product of three matrices:  $\mathbf{I}\Lambda_\theta\rho_\theta$ , allowing us to rewrite it as  $(\rho_\theta^T \otimes \mathbf{I})\text{vec}(\Lambda_\theta)$  through vectorization. We can proceed analogously for the second term. Therefore, this equation simply becomes a linear system in a bigger Hilbert space:

$$(\rho_\theta^T \otimes \mathbf{I} + \mathbf{I} \otimes \rho_\theta)\text{vec}(\Lambda_\theta) = 2\text{vec}(\partial_\theta\rho_\theta). \quad (4.10)$$

The solution is straightforward with orthodox methods:

$$\text{vec}(\Lambda_\theta) = 2(\rho_\theta^T \otimes \mathbf{I} + \mathbf{I} \otimes \rho_\theta)^{-1}\text{vec}(\partial_\theta\rho_\theta). \quad (4.11)$$

This avoids the hassle of diagonalizing the density matrix: we convert the problem of obtaining the QFI and the SLD to a simple linear algebra problem based on a linear system. Obtaining the inverse matrix becomes the numerical bottleneck in this case. Moreover, by using the identity  $\text{tr}\{A^\dagger B\} = \text{vec}(A)^\dagger \text{vec}(B)$ , we can rewrite the QFI in terms of the vectorized expressions as

$$\mathcal{F} = \text{vec}(\Lambda_\theta)^\dagger (\rho_\theta^T \otimes \mathbf{I} + \mathbf{I} \otimes \rho_\theta)^{-1} \text{vec}(\Lambda_\theta). \quad (4.12)$$

We now discuss an example, adapted from the contents of Ref. [158], showing how the result above can be used in practice.

■ **Example:** Consider the pure state given by  $|\psi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{e^{i\varphi}}{\sqrt{2}}|1\rangle$ , where  $\varphi$  is a phase. We will consider a scheme describing noise, which is quantified through a param-



eter  $\nu \in (0, 1)$ . The state of interest is a convex combination of the pure state  $|\psi\rangle$  and the maximally mixed state:

$$\rho_{\nu\theta} = (1 - \nu) |\psi\rangle \langle\psi| + \frac{\nu}{2} \mathbf{I} = \frac{1}{2} \begin{pmatrix} 1 & e^{i\theta}(1 - \nu) \\ e^{-i\theta}(1 - \nu) & 1 \end{pmatrix}. \quad (4.13)$$

The mixed state  $\rho_{\nu\theta}$  describes, e.g., some noise deterioration due to dephasing or an experiment where a machine prepares the state  $|\psi\rangle$  with a certain degree of imperfection. Our objective in this example is to obtain the QCRB for the parameters  $\nu$  and  $\theta$ , given the state  $\rho_{\nu\theta}$ . Note that in this example we are skipped the first two steps in the diagram show in Fig. (4.1). We will consider the parametrized state above as given, ignoring the underlying input states and the parametrization dynamics.

We begin by obtaining the QFI and the optimal POVMs for the noise parameter  $\nu$ . According to Eq. (4.10), we may start by vectorizing the expression

$$\text{vec}(\partial_\nu \rho_{\nu\theta}) = -\frac{1}{2} \begin{pmatrix} 0 \\ e^{-i\theta} \\ e^{i\theta} \\ 0 \end{pmatrix},$$

and the LHS of Eq. (4.2)

$$(\rho_{\nu\theta}^T \otimes I + I \otimes \rho_{\nu\theta}) = \begin{pmatrix} 1 & \frac{1}{2}e^{i\theta}(1 - \nu) & \frac{1}{2}e^{-i\theta}(1 - \nu) & 0 \\ \frac{1}{2}e^{-i\theta}(1 - \nu) & 1 & 0 & \frac{1}{2}e^{-i\theta}(1 - \nu) \\ \frac{1}{2}e^{i\theta}(1 - \nu) & 0 & 1 & \frac{1}{2}e^{i\theta}(1 - \nu) \\ 0 & \frac{1}{2}e^{i\theta}(1 - \nu) & \frac{1}{2}e^{-i\theta}(1 - \nu) & 1 \end{pmatrix},$$

respectively. By inverting the matrix above we obtain the following expression for the QFI:

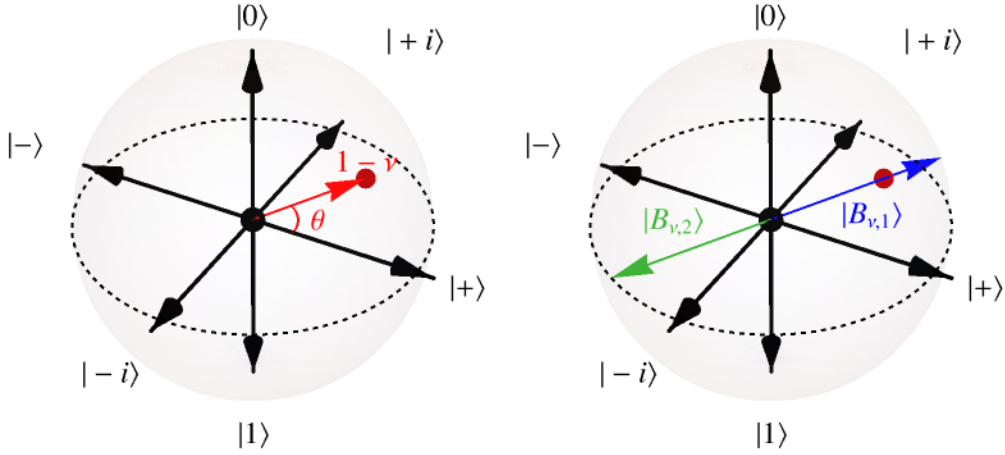
$$\mathcal{F}_\nu = 2\text{vec}(\partial_\nu \rho_{\nu\theta})^\dagger (\rho_{\nu\theta}^T \otimes I + I \otimes \rho_{\nu\theta})^{-1} \text{vec}(\partial_\nu \rho_{\nu\theta}) = \frac{1}{\nu(2 - \nu)}. \quad (4.14)$$

Furthermore, we can now use Eq. (4.11) to explicitly obtain the SLD

$$L_\nu = \frac{1}{\nu(2-\nu)} \begin{pmatrix} 1-\nu & e^{i\theta} \\ e^{-i\theta} & 1-\nu \end{pmatrix}. \quad (4.15)$$

By calculating its eigenvectors we can finally arrive at the optimal measurement basis:

$$|B_{\nu,1}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\theta} \\ 1 \end{pmatrix}, \quad |B_{\nu,2}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} -e^{i\theta} \\ 1 \end{pmatrix}. \quad (4.16)$$



**Figure 4.2:** (Left) State  $\rho_{\nu\theta}$  and (Right) the optimal measurement basis on the Bloch sphere.

The POVMs are then constructed simply by taking the projectors  $|B_{\nu,1}\rangle\langle B_{\nu,1}|$  and  $|B_{\nu,2}\rangle\langle B_{\nu,2}|$ . In Fig. 4.2 we illustrate the state and the basis on the Bloch sphere. Note how  $|B_{\nu,1}\rangle$  is in the same direction as the state. By using this measurement basis, we obtain the distributions

$$p(X = B_{\nu,1}|\nu) = 1 - \frac{\nu}{2} \quad \text{and} \quad p(X = B_2|\nu) = \frac{\nu}{2} \quad (4.17)$$

for the outcomes.<sup>3</sup> Note that measurements with this POVM are analogous to a coin toss experiment with  $\nu/2$  as the parameter in the Bernoulli distribution. In the ideal scenario where  $\nu = 0$ , we would *always* measure the system in the state  $|B_{\nu,1}\rangle$ , but *never* in  $|B_2\rangle$ ,

<sup>3</sup>Here we commit a slight abuse of notation. By writing  $p(X = B_i|\nu)$  we refer to the probability of measuring the system in the state  $|B_i\rangle$ , conditioned on the parameter  $\nu$ .

since the pure state  $|\psi\rangle$  defined at the beginning of this example is orthogonal to  $|B_2\rangle$ . On the other hand, in a configuration where  $\nu$  is non-zero, we would sometimes measure the state in  $|B_2\rangle$ . This happens not because of a superposition in the  $\{|B_{\nu,1}\rangle, |B_2\rangle\}$  basis, but rather, because of the classical probabilities associated with the mixture in  $\rho$ . Choosing a basis other than  $|B_{\nu,1}\rangle, |B_{\nu,2}\rangle$  would make it harder to decide whether the results arise from the superposition in the state  $|\psi\rangle$  or because of the impurity of  $\rho$ .

Finally, note that the QFI imparts the ultimate limit of precision associated to the estimation of the parameter of concern, while the eigenvectors of the SLD provide the optimal basis which is associated with the saturation of this bound. Nevertheless, they give no information about how the measurement data should be used. It is then clear that one should also formulate a concrete post-processing strategy. In this example, we can estimate  $\nu$  by counting the number of different outcomes in each state, in a large analogy with the scenario of Bernoulli trials. Thus, the obvious and best candidate for an estimator in this case is the sample mean, which we discussed in previous chapters. The FI associated with the distributions in Eq. (4.17) is simply the expression that we have obtained in Eq. (2.30), but with respect to the parameter  $\nu/2$  instead. You can see that the FI in this case will *coincide* with the QFI, as it should be, since we are purposely performing measurements in the optimal basis. Finally, note that due to the properties of the sample mean as an estimator for the Bernoulli distribution, we actually saturate the bound even in the non-asymptotic regime. This estimator is *efficient*, as we have thoroughly discussed before. These steps give a complete picture of how the noise parameter could be estimated in this toy model.

There are a few further things worth mentioning in this example. The first one is that the QFI, as it very often happens, depends on the unknown parameter  $\nu$ . Fortunately, that was not the case for the optimal POVM here, which depends only on  $\theta$ . That, however, is not guaranteed to happen every time. If we decided to probe the phase  $\theta$  instead of the parameter  $\theta$ , for instance, the optimal POVMs would also depend on the unknown  $\theta$  [158].<sup>4</sup> We can repeat the same steps with respect to the parameter  $\theta$  instead to see this

<sup>4</sup>See, e.g., Ref. [164] (p. 3) and the discussion therein.

in practice. By doing so, we find the SLD and the optimal measurement basis to be

$$L_\theta = i(\nu - 1) \begin{pmatrix} 0 & -e^{i\theta} \\ e^{-i\theta} & 0 \end{pmatrix} \quad (4.18)$$

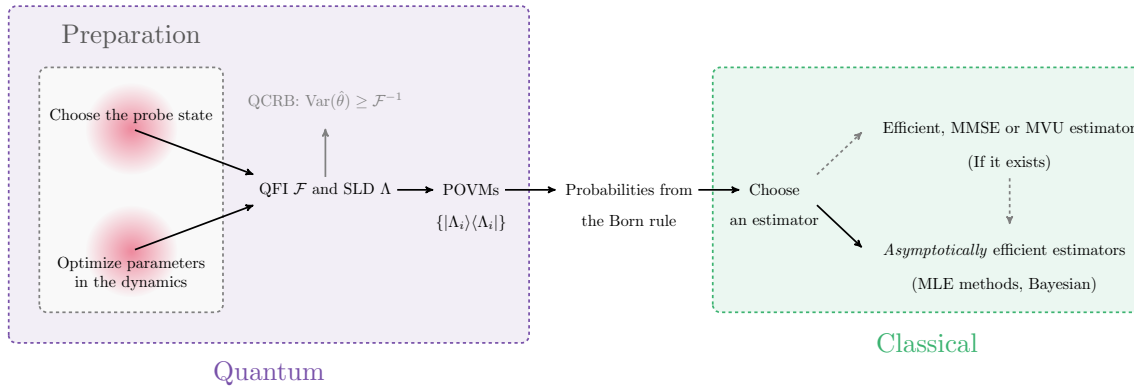
and

$$|B_{\theta,1}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} ie^{-i\theta} \\ 1 \end{pmatrix}, \quad |B_{\theta,2}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} -ie^{-i\theta} \\ 1 \end{pmatrix}, \quad (4.19)$$

respectively. Note how the latter explicitly depends on  $\theta$ . This example illustrates that it is not uncommon for the optimal basis to depend on the unknown parameter itself, introducing another layer of complexity to the problem. This often calls for adaptive strategies whenever possible, where one continuously change the measurement basis, or, e.g., parameters and the general configuration of the interaction [165–167].  $\square$

The previous example briefly showed the typical steps which constitute a metrology experiment in quantum mechanics. One of most striking aspects is the importance of the measurement choice in this scheme. We illustrate the main ideas with a diagram in Fig. 4.3. A very common point of focus in the literature is the notion of optimization over *the probe states* and quantities appearing in the parametrization process, such as parameters in a unitary evolution. This was not a point of concern in the previous example because we considered a given state. In more concrete scenarios these might be central points of investigation. The experimenter will typically try to choose a probe state which better encodes the parameters of interest after undergoing a predefined dynamics. Likewise, it is also important to perform this process in the best configuration possible. If some parameters of the interaction are known, they might be tuned in a way which results in a state with a larger QFI. For that reason, many works are devoted to establishing experimental setups and protocols with a sensible configuration which allow for very precise estimations. This is, in a first instance, quantified through the QFI, further justifying its importance.

Once the parametrized state has been obtained, it is then possible to calculate the QFI and the SLD. The latter is, in turn, used to obtain the optimal POVMs. More specifically, the projectors are constructed from the eigenvalues of the SLD, providing the best possible measurements, whose precision is bounded by the QCRB. These are quantities which are



**Figure 4.3:** Standard optimization strategies in metrology.

obtained entirely from the state, and no other elements are necessary in principle. An important point in this step, which has actually been omitted in the diagram, is that a few other strategies might result in very sharp improvements. One can, for example, exploit superposition and entanglement, performing joint measurements on several copies of the system or ancillas. By doing so it is often possible to capture correlations which result in more precise estimations [168]. Furthermore, protocols which employ other strategies or (quantum) resources, such as quantum discord, also configure a very active field in quantum metrology [169, 170]. See Ref. [130] for a review.

Regardless, once the POVMs have been decided upon, the estimation task essentially becomes a classical problem. At this point, the state and the POVM are the relevant objects, and we can use the Born rule to obtain the probability distribution for each of the outcomes. By doing so, the formalism that we have discussed in the previous chapters can be used as a powerful tool in the post-processing tasks, where we use the resulting distributions and the empirical data to estimate the parameters of interest. This is the point where it is necessary to pick a sensible estimator following the considerations that we have thoroughly discussed. Additionally, it is also possible to calculate the variance, or any figure of error which might be relevant, associated with the estimator.

These steps are illustrated on the green region highlighted on the diagram in Fig. 4.3. It is then possible to look for efficient estimators, or even MMSE and MVU estimators. This should provide an optimal post-processing treatment. If these do not exist, it is then common to resort to estimators which are *asymptotically* efficient, such as the ones obtained by MLE methods. Still, it is also important to remember that even MLE strategies

might contain some limitations. This is a hurdle encountered in, e.g., phase estimation [166]. These considerations also highlight the importance of the FI and the QFI: even if we are either unable to obtain an optimal estimator (or to employ the best measurement strategy), the (quantum) Fisher information still provides an important benchmark.

For a good application of this discussion in practice, the reader might refer to the Section III from Ref. [171] for example, where the problem of the qubit phase estimation is investigated. This publications provide a very good illustration of each of these steps that we have discussed. There, the author discusses the optimization of the QFI, the measurements and the estimators, in this order. An optimization over the probe states is given in a previous work by the same author, in Ref. [164]. Note how these contents overlap with the diagram shown in Fig. 4.3.

To complete our discussion, we dwell into question of the attainability of the QCRB. In the single-parameter case the QCRB can usually be achieved, at least asymptotically, by employing measurements in the appropriate basis per our discussions. Additionally, adaptative protocols might have to be employed.<sup>5</sup> The situation becomes much more involved when the multi-parameter case is considered: since different parameters will, in general, require a different set of POVMs for optimality, one can seldom saturate the multi-parameter QCRB. Surprisingly, the geometrical aspects of the QFI come at hand in this discussion: one can establish a link between the simultaneous indistinguishability of different parameters and the impossibility of attaining the multi-parameter QCRB with the non-zero Berry curvature and the geometrical properties of the quantum statistical model. These very elegant conclusions, which cleverly connect geometric and statistical aspects, can be found in Refs. [163] and [173] (see also [174] for an experimental follow-up). On the same spirit, other results provide useful insights about this problem: the QCRB is saturable if, and only if, the state  $\rho$  belongs to what is called the *quantum exponential family* [159] (p. 25). We are not going to define this class of states here, but the statement of the theorem can be found in Ref. [175] (pp. 118-120). This result depicts one of the differences between classical and quantum estimation theories quite well: in the former the CRB is saturated through the exponential family of *distributions*, while in the latter we consider a family of *states*. This illustrates how distributions are the fundamental objects

---

<sup>5</sup>See Ref. [172] for a discussion in the case of pure states.

in standard statistics, while it makes more sense to consider the manifold of states to play the analogous role in quantum metrology.

Finally, we shall conclude this section with a practical example. We begin by mentioning Ref. [176], where the authors provide some optimization over single-qubit probes for one-parameter qubit gates. Moreover, they also study the *stability* of these states as probes; in the sense that they investigate how small perturbations in the probe states affect the FI of the model. By doing so they were able to show that these small perturbations might introduce large deviations in the FI, making the protocol unstable. Surprisingly, one of their main conclusions was that using entangled two-qubit probes in the Bell states actually *protects* the protocol against fluctuations in the probe state. This is a concrete example in the literature which shows how the preparation process, and choosing good probe states, is important in metrology. An experimental approach of a very similar problem was done in [177]. Therefore, we introduce an example below which adapts the discussion from Refs. [172, 176, 177] in a simpler scenario. See also Refs. [178, 179] for very pedagogical approaches to the problem.

■ **Example:** Consider the initial state  $|\psi_0\rangle = \cos \frac{\alpha}{2} |0\rangle + e^{i\varphi} \sin \frac{\alpha}{2} |1\rangle$ , parametrized in terms of the polar angle  $\alpha$  in the Bloch sphere. Now, let us suppose that this state undergoes a unitary dynamics of the type  $U(\theta) = \exp\{-i\frac{\theta}{2}\sigma_z\}$ , resulting in the state  $|\psi_\varphi\rangle = \cos \frac{\alpha}{2} |0\rangle + e^{i(\varphi-\theta)} \sin \frac{\alpha}{2} |1\rangle$  (up to a phase factor) which encodes a phase-shift angle  $\varphi$  that we wish to estimate. Note that we are able to generate the pure state in Eq. (4.13) from the previous example through the same process. Our objectives with this example are two-fold:

- We want to see how the angle  $\theta$  in the probe states influences the QFI and the precision of the estimation.
- We extend the investigation to measurement bases beyond the optimal protocol. We would like to quantify how sub-optimal measurements fare.

We start by calculating the QFI. By repeating the steps of the previous example, we

find that the term  $(\rho_\theta^T \otimes I + I \otimes \rho_\theta)$  is given by

$$\begin{pmatrix} \cos^2\left(\frac{\alpha}{2}\right) & \frac{1}{4}\sin(\alpha)e^{i(\theta-\varphi)} & \frac{1}{4}\sin(\alpha)e^{-i(\theta-\varphi)} & 0 \\ \frac{1}{4}\sin(\alpha)e^{-i(\theta-\varphi)} & \frac{1}{2} & 0 & \frac{1}{4}\sin(\alpha)e^{-i(\theta-\varphi)} \\ \frac{1}{4}\sin(\alpha)e^{i(\theta-\varphi)} & 0 & \frac{1}{2} & \frac{1}{4}\sin(\alpha)e^{i(\theta-\varphi)} \\ 0 & \frac{1}{4}\sin(\alpha)e^{i(\theta-\varphi)} & \frac{1}{4}\sin(\alpha)e^{-i(\theta-\varphi)} & \sin^2\left(\frac{\alpha}{2}\right) \end{pmatrix},$$

for  $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ . This matrix is singular, so it cannot be used to calculate the QFI in this case. We avoid this conundrum by employing one of the strategies from Ref. [180] (See Theorem 4, p. 5). Consider the state  $\tilde{\rho}_\nu := (1 - \nu)\tilde{\rho} + \frac{\nu}{2}I$ , where  $0 < \nu < 1$  is a real parameter. Note that this is the same state from the previous exercise. This time, however, we are not interested on its physical features, but rather, on its mathematical convenience. We can invert the previous term by taking the limit:

$$(\rho_\theta^T \otimes I + I \otimes \rho_\theta) = \lim_{\nu \rightarrow 0} (\tilde{\rho}_\nu^T \otimes I + I \otimes \tilde{\rho}_\nu). \quad (4.20)$$

That is, we calculate the QFI for a mixed state and then obtain the limiting case of pure states by taking  $\nu \rightarrow 0$  [158]. This convenient technique avoids the problem of singular matrices.<sup>6</sup> Hence, by employing very similar steps to the previous exercise, we obtain that the QFI of the state  $\rho_\theta$  is given by:

$$\mathcal{F}(\rho_\theta) = \sin^2(\alpha). \quad (4.21)$$

This is a well-known result for qubits [172] (p. 4486)<sup>7</sup>, with some intuition behind it. Notice how Eq. (4.21) vanishes for  $\alpha = 0$  or  $\alpha = \pi$ , which corresponds to the states  $|0\rangle$  and  $|1\rangle$ , respectively, and how it achieves its maximum for  $\alpha = \pi/2$ , corresponding to states in the equator of the Bloch sphere, such as the  $|+\rangle$  and  $|-\rangle$  states, defined by  $(|0\rangle \pm |-\rangle)/\sqrt{2}$ , respectively.

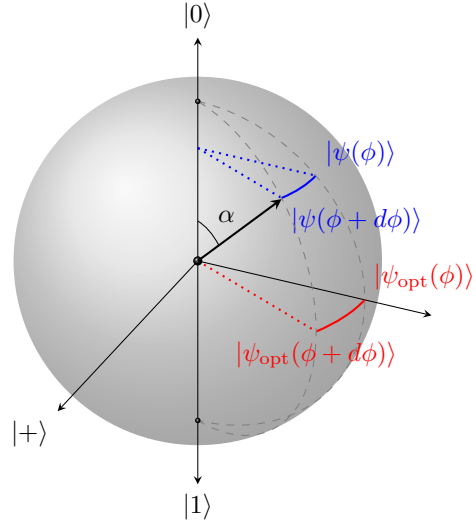
---

<sup>6</sup>In more generality, the result from Ref. [180] states that by defining the state  $\tilde{\rho}_\nu := (1 - \nu)\tilde{\rho} + \frac{\nu}{\dim \mathcal{H}}I$ , where  $\mathcal{H}$  is the finite-dimensional underlying Hilbert space, one can write the QFI of the state  $\tilde{\rho}$  as  $\mathcal{F}(\tilde{\rho}) = \lim_{\nu \rightarrow 0} \mathcal{F}(\tilde{\rho}_\nu)$ . This strategies is also used to deal with discontinuities.

<sup>7</sup>Since we are dealing with pure states, simple formulas which do *not* depend on vectorization can be found in any standard reference [11, 159]. We insist in these formulas for conciseness, because these are the only ones we are going to use in future sections.



The reason for that is very simple: it is very hard to distinguish the states around the pole by changing the phase-shift  $\theta$ . In the limiting case where the state is situated in one of the poles of the Bloch sphere, applying the unitary  $U(\theta)$  does not affect the state of the qubit at all, making it impossible to infer the parameter. Conversely, we find that distinguishing states on the equator of the Bloch sphere after applying the phase shift is the easiest: infinitesimal displacements of the azimuthal angle will be the most detectable in this region. One might notice how the QFI in Eq. (4.21) coincides with one of elements of the metric tensor of the unit sphere.



**Figure 4.4:** Illustration of qubit states in the Bloch sphere after an infinitesimal phase shift. Here  $|\psi_{\text{opt}}\rangle$  illustrates an optimal probe state lying in the equator.

Let us proceed by considering the measurements. This time, we will go beyond the optimal measurements. We shall, instead, consider a wider class of POVMs (which are not necessarily constructed from the SLD):

$$M_0 = |\Psi\rangle\langle\Psi|, \quad M_1 = I - M_0, \quad \text{with} \quad |\Psi\rangle = \cos\frac{\beta}{2}|0\rangle + e^{i\phi}\sin\frac{\beta}{2}|1\rangle. \quad (4.22)$$

Here  $\beta$  and  $\phi$  are angles which parametrize the measurement basis. Our objective is to see how the angles  $\alpha$  and  $\varphi$  in the probe states and the angles  $\beta$  and  $\phi$  parametrizing the POVMs affect the protocol. Namely, we will be able to quantify how well different probe states and POVMs perform. This is quantified through the FI. By using the projectors above, we obtain the probabilities

$$p_{0,1}(\theta) = \frac{1}{2} \pm \frac{f(\theta)}{2}, \quad \text{with} \quad f(\theta) := \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)\cos(\theta - \varphi + \phi), \quad (4.23)$$

associated with  $M_0$  and  $M_1$ , respectively. If we obtain  $k$  outcomes in 0 and  $n-k$  outcomes in 1, we can write the distribution for the outcomes of this experiments as  $p(X|\theta) =$

$p_0(\theta)^X(1 - p_0(\theta))^{1-X}$ , allowing us to explicitly calculate its FI. We should be careful with this distribution however. While it closely resembles the Bernoulli distribution, the "Bernoulli parameter" actually depends on  $\theta$  in a complicated manner, per Eq. (4.23). The simplest way to calculate the FI in this case, since we have already done that for the Bernoulli distribution, is to use the reparametrization rule for the Fisher Information [87].

<sup>8</sup> By doing so we can calculate the FI to be:

$$F(\theta|\alpha, \beta) = \left( \frac{dp_0(\theta)}{d\theta} \right)^2 \frac{1}{p_0(\theta)(1 - p_0(\theta))} = \frac{\sin^2(\alpha) \sin^2(\beta) \sin^2(\theta - \varphi + \phi)}{(1 - f(\theta))^2}. \quad (4.24)$$

We can optimize the expression above by taking  $\alpha = \beta = \pi/2$ . This means that one should take a probe state and a measurement basis which lie on the equator of the Bloch sphere. By doing so, we obtain  $F(\theta) = 1$  for all  $\theta$ , independent of  $\phi$  and  $\varphi$ . The FI in this case coincides with the maximum achievable QFI, with  $\alpha = \pi/2$ . <sup>9</sup> We plot these results in Fig. 4.5.

These calculations by themselves are interesting enough as an example. However, in Ref. [176] they bring a very interesting point into attention, which concerns the *stability* of the probe states: that is, how much do small fluctuations in the parameters of the probe state and of the POVMs influence the optimality of the protocol? We can follow Ref. [176] and Taylor expand Eq. (4.24) around  $\alpha = \pi/2$  to obtain (we consider that  $\varphi = \phi = 0$ , for simplicity):

$$F(\theta) \approx 1 - \delta\alpha^2 \frac{1}{\sin^2 \theta}, \quad (4.25)$$

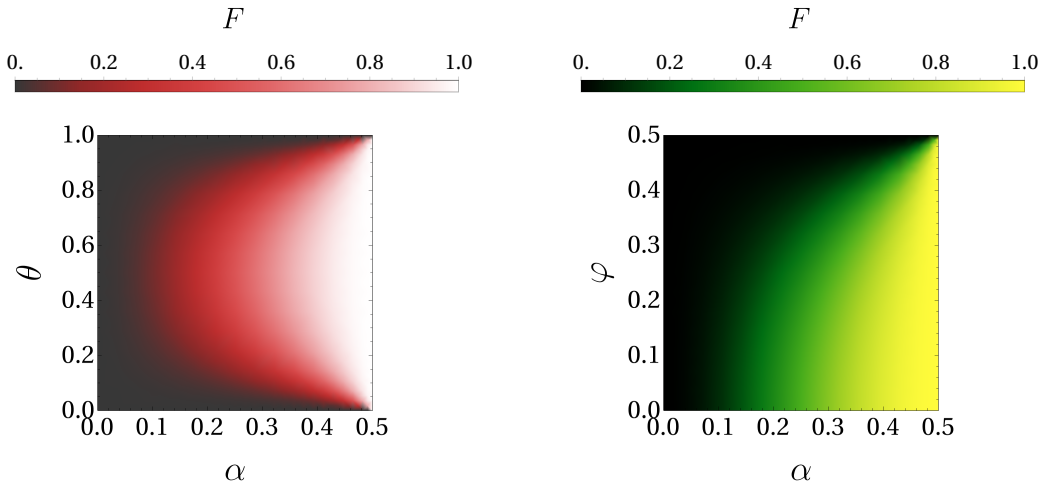
with  $\delta\alpha = \pi/2 - \alpha$ . Surprisingly, for small  $\theta$  we can see that this becomes  $F(\theta) \approx \delta\alpha^2/\theta^2$ . In other words, small deviations in the preparation of the probe *greatly* affect the achievable precision of the protocol, specially when small angles are considered.

This example showcases why the *preparation* step is very subtle when devising experiments in quantum metrology, and why one should also consider other aspects beyond only maximizing the QFI. The striking conclusion from the authors in Ref. [176] was to show that entangled probes are able to improve *the stability* of the protocol, making it

---

<sup>8</sup>Suppose that a parameter  $\eta$  depends on another parameter of interest  $\theta$ . The rule states that  $F(\eta) = \left( \frac{d\eta}{d\theta} \right)^2 F(\theta)$ .

<sup>9</sup>However, we briefly mention that mixed states are also considered, the optimal configuration is much different [178].



**Figure 4.5:** FI for the qubit in different scenarios. On the left we plot the FI as a function of the polar angle  $\alpha$  for the qubit state on the Bloch sphere and the interferometric parameter  $\theta$ . On the right we plot the QFI as a function of  $\alpha$  and  $\varphi$ , describing all the possible probe states. We can see that the optimality is achieved when  $\alpha = \pi/2$ , regardless of the azimuthal angle  $\varphi$  in the Bloch sphere.

optimal for a wider range of configurations and less sensitive to small perturbations in the preparation of the probe. This is an interesting example to show how aspects such as quantum improvements might be useful even beyond the scenario of surpassing the standard quantum limit.  $\square$

We finish this section with a brief disclaimer about some caveats which arise in quantum metrology. The previous discussion, for instance, very interestingly showed how entanglement could be used to improve the *stability* of a protocol. And while we omitted such an example here, an even larger body of the literature is concerned with the challenge of surpassing the SQL [181]; as we mentioned in the beginning of this chapter. However, employing quantum advantages is in practice, a difficult task. For instance, the choice of the estimator in phase estimation problems requires some care [182]. Additionally, it is very well known at this point that schemes based on the Heisenberg scaling are typically very sensitive to noise [144, 183, 184]. In Ref. [185] the authors provide some further bounds and concretely shows how estimations in the presence of noise might gradually transit from the Heisenberg scaling to the usual scaling in the SQL. Maybe even more importantly, the question of what counts as a "resource" for quantum metrology and the QCRB is certainly a fundamental question which we should dwell into. For instance,

Ref. [186] provides a unifying discussion of these aspects, establishing a connection between several different experiments and models. In summary, these observations, allied to the aforementioned references, allows us to better understand current the literature along with its challenges, and show us that many subtleties, such as technological feasibility, robustness and post-processing aspects should be taken into account when devising a sensible experiment in quantum metrology.

### 4.3 Quantum thermometry

Measuring and estimating temperature with high precision in nanoscale and quantum systems is an effort of increasingly importance in modern technologies [187–190]. We can find uses in, e.g., ultra-cold gases [191–194], biology [195] and even relativity [196], with proposals for the detection of the Unruh temperature [197] and the thermometry of gravitational waves [198]. Lying in a very rich intersection between quantum metrology and quantum thermodynamics, the field concerned with such problem is known as quantum thermometry, and it has seen a fast development in the past decade [199–204]. For a recent collection of important developments and works in the field, consult Refs. [205] (Part III, Sec. 3) and [206].<sup>10</sup> Rather than *solely* transposing the tools from quantum estimation theory, quantum thermometry emerges as a very careful framework which incorporates the physical and mathematical traits from thermodynamics into the underlying framework of quantum metrology. Due to its own peculiarities, quantum thermometry comes across theoretical and experimental challenges which are particular of it as its own field - some challenges which are not generally present in quantum metrology and other subfields. With that in mind, we use the discussions of the preceding chapters as the foundation of the upcoming results.

One of the main building blocks in the field is the notion of a *thermal* quantum Fisher information, which can be used as the main benchmark for thermometric protocols - specially for near-equilibrium configurations [208]. The standard setup is illustrated in Fig. 4.7. It consists on a probe which is left to interact with a thermal bath at a tempera-

---

<sup>10</sup>For a review in quantum thermodynamics instead, see Ref. [207].

ture  $T$ . The probe eventually thermalizes to the Gibbs state

$$\rho_{th} := \frac{e^{-\beta H}}{Z}, \quad (4.26)$$

associated with the Hamiltonian  $H$ , where we define  $\beta = 1/T$  as the reciprocal of the temperature (from hereafter we take the Boltzmann constant to be  $k_b = 1$ ). Hence, by performing measurements and post-processing strategies (per Fig. 4.3, for example) on the thermalized probe, we can infer the temperature  $T$  of the bath. All things considered, this is very similar to previous setups we have analyzed, with the complication of an extra step where the probe acts as a proxy for measuring the bath temperature. Modeling this interaction might be a big challenge on itself. We omit such a discussion here, but an example can be found in Ref. [209].

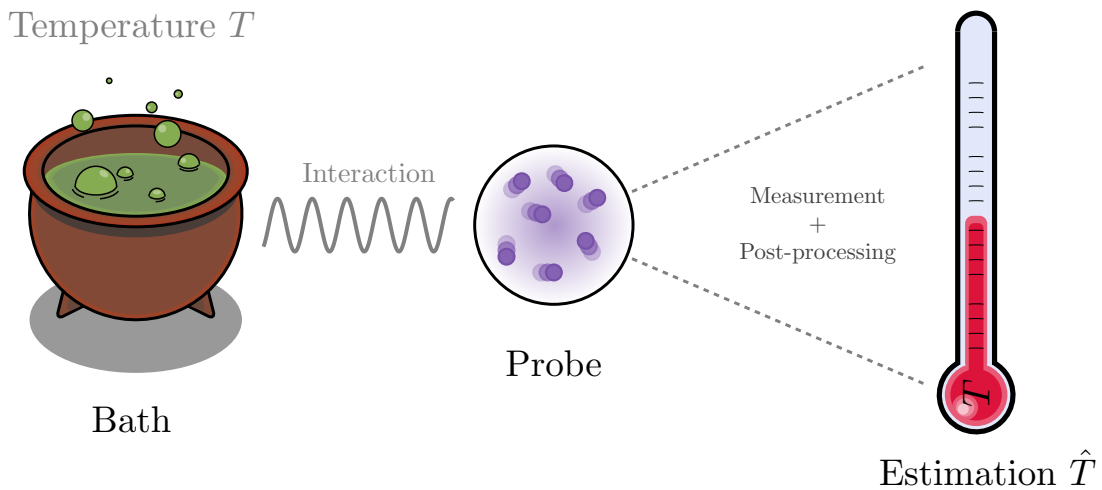
More concretely, the thermal Fisher information (TFI) is defined as the quantum fisher information of the thermal state, and it is given by

$$\mathcal{F}_{th} = \frac{C}{T^2}, \quad (4.27)$$

where

$$C = \frac{\langle H^2 \rangle - \langle H \rangle^2}{T^2} \quad (4.28)$$

is the heat capacity of the system. To arrive at this result we can write the Lyapunov



**Figure 4.6:** Diagram illustrating a thermometry experiment.

equation for the thermal state

$$\Lambda_T \rho_{th} + \rho_{th} \Lambda_T = 2 \partial_T \rho_{th}, \quad (4.29)$$

and solve it for the "thermal" SLD  $\Lambda_T$ . We briefly review some results from statistical physics in order to proceed. In particular, we will need to identities some results in order to obtain the derivative of the density matrix in Eq. (4.26). It can be written in a more convenient way through the use of the chain rule:

$$\partial_T \rho_{th} = -\beta^2 \frac{\partial \rho_{th}}{\partial \beta} = -\beta^2 \left[ -H \frac{e^{-\beta H}}{Z} - \frac{e^{-\beta H}}{Z} \left( \frac{1}{Z} \frac{\partial Z}{\partial \beta} \right) \right],$$

or equivalently,

$$\partial_T \rho_{th} = \beta^2 \left[ H - \frac{1}{Z} \frac{\partial Z}{\partial \beta} \right] \rho_{th}, \quad (4.30)$$

since  $\partial_T \beta = -1/T^2$ . Moreover, knowing that

$$\frac{\partial Z}{\partial \beta} = - \sum_n E_n e^{-\beta E_n} = -Z \langle H \rangle,$$

we have:

$$\frac{1}{Z} \frac{\partial Z}{\partial \beta} = \langle H \rangle. \quad (4.31)$$

Thus we can see that (4.30) becomes:

$$\partial_T \rho_{th} = \beta^2 (H - \langle H \rangle) \rho_{th}, \quad (4.32)$$

and, in turn, Eq. (4.29) yields:

$$\Lambda_T \rho_{th} + \rho_{th} \Lambda_T = 2\beta^2 (H - \langle H \rangle) \rho_{th}. \quad (4.33)$$

Given that the thermal state is simply the exponential of the Hamiltonian, we know that it commutes with the Hamiltonian itself, i.e.  $[H, \rho_{th}] = 0$ . This fact which makes the equation above much easier to solve, and the solution is straightforward if we assume that

$\Lambda_T \rho_{\text{th}} = \rho_{\text{th}} \Lambda_T$ , leading us to:

$$\Lambda_T = \beta^2(H - \langle H \rangle). \quad (4.34)$$

We can immediately verify that this solution for SLD is valid. Now, since  $\text{Tr}(H\langle H \rangle) = \langle H \rangle^2$  it is also easy to see that the QFI  $\text{Tr}(\rho_{\text{th}} \Lambda_T^2)$  will be given by the expression:

$$\mathcal{F}_{\text{th}} = \frac{\langle H^2 \rangle - \langle H \rangle^2}{T^4}, \quad (4.35)$$

thus proving the results in Eqs. (4.27) and (4.28). Finally, we mention that the measurement optimal basis in this scenario is simply a projection onto the energy eigenbasis, since the thermal state (4.26) is diagonal in this basis.

▣ **Example:** As an example, we shall apply the results above to a qubit. We start by considering the very simple case where the qubit is described by the Hamiltonian  $H = \frac{\Omega}{2} \sigma_z$ . Here,  $\sigma_z$  refers to the usual Pauli matrix  $\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$ . We find that the partition function in this case is given by  $Z = 1 + e^{\Omega/T}$ , and that

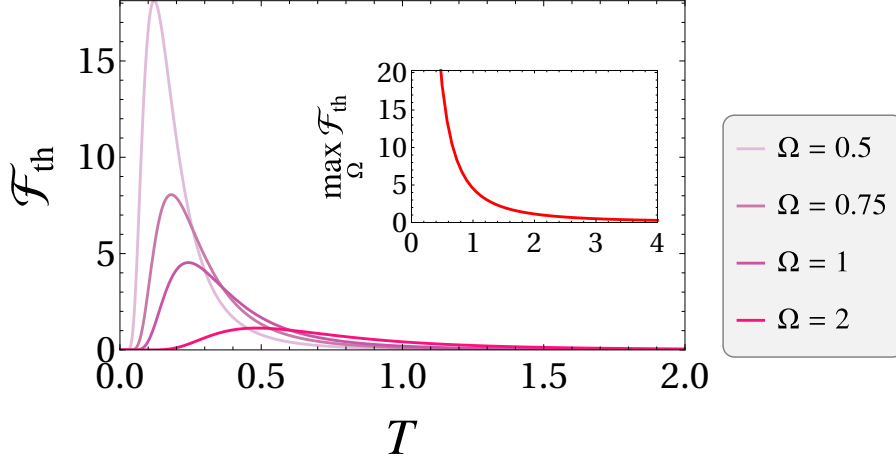
$$\rho_{\text{th}} = \frac{1}{1 + e^{\Omega/T}} \begin{pmatrix} 1 & 0 \\ 0 & e^{\Omega/T} \end{pmatrix}. \quad (4.36)$$

Since  $\sigma_z$  is idempotent, the second moment very easily evaluates to  $\langle H^2 \rangle = \Omega^2/4$ . Meanwhile, we find that  $\langle H \rangle = \text{Tr}\{H\rho_{\text{th}}\} = -\frac{\Omega}{2} \tanh\left(\frac{\Omega}{2T}\right)$ . Therefore, by substituting these results into (4.27) and by using the identity  $\text{sech}^2 x = 1 - \tanh^2 x$  we obtain

$$\mathcal{F}_{\text{th}} = \left(\frac{\Omega}{2T^2}\right)^2 \text{sech}^2\left(\frac{\Omega}{2T}\right) \quad (4.37)$$

as the thermal fisher information of a (fully thermalized) qubit.

We plot this result in Fig. 4.7. The first observation we can make is that the TFI depends on the temperature itself, reaching a maximum for a certain value of  $T$ , depending on the size  $\Omega$  of the energy gap. We perform a maximization over the temperature, plotting the maximum achievable TFI as a function of the gap in the inset of Fig. 4.7. We can see smaller gaps can result in much higher values for the TFI. We should be careful however in how we interpret these results. This plot tells us that smaller gaps yield



**Figure 4.7:** Thermal Fisher information of a qubit. We plot the TFI as a function of the temperature  $T$  for different values of the energy gap  $\Omega$ . In the inset we plot the maximum TFI (maximized over  $T$ ) as a function of  $\Omega$ . We can see that smaller gaps result in a higher achievable precision for the optimal temperature.

higher precision for *certain* values of the temperature, but not for all of them. We can see from the plots that while smaller values of  $\Omega$  are more precise for lower temperatures, they quickly decay for larger values of  $T$ . Hence, it might be preferable to opt out for a larger gap as  $T$  increases. This scenario is a further illustration on the challenges of optimization in metrology that we have encountered before, but this time in the context of thermometry. We can see how, e.g., optimization might also emerge as a very natural strategy in thermometry, for example.  $\square$

Another very important observation about the plots in Fig. 4.7 and Eq. (4.37) which is worth mentioning is that the TFI (and the heat capacity) both converge to zero as  $T \rightarrow 0$ . Thus, standard thermometry protocols become very imprecise in this regime. Many works in the existing literature are concerned with this challenge [201, 210, 211]. See for example Sec. 4.3 in Ref. [154]. Even worse, the heat capacity usually decays *exponentially* as  $T \rightarrow 0$ . Many schemes mitigate this problem by employing configurations and techniques which results in a sub-exponential, and sometimes, polynomial decay [200, 212].

A natural generalization would be to investigate how different probes, such as a quantum harmonic oscillator substituting the qubit in the previously described setup, would affect the protocol. See Sec. 2.1.1 from Ref. [207], for instance. There are already many thorough investigations in this sense [213]. Moreover, there also exist some surprisingly general results concerning other aspects of thermometry. To name a few, we mention the



results concerning the dimensionality and the structure of the eigenspectrum of the system from the discussion in Ref. [214]. There, the authors state that the optimal configuration in the (equilibrium) probe-based thermometry is to employ highly degenerate Hamiltonians. Similarly, a discussion on coarse-grained measurements in quantum thermometry can be found in Ref. [215].

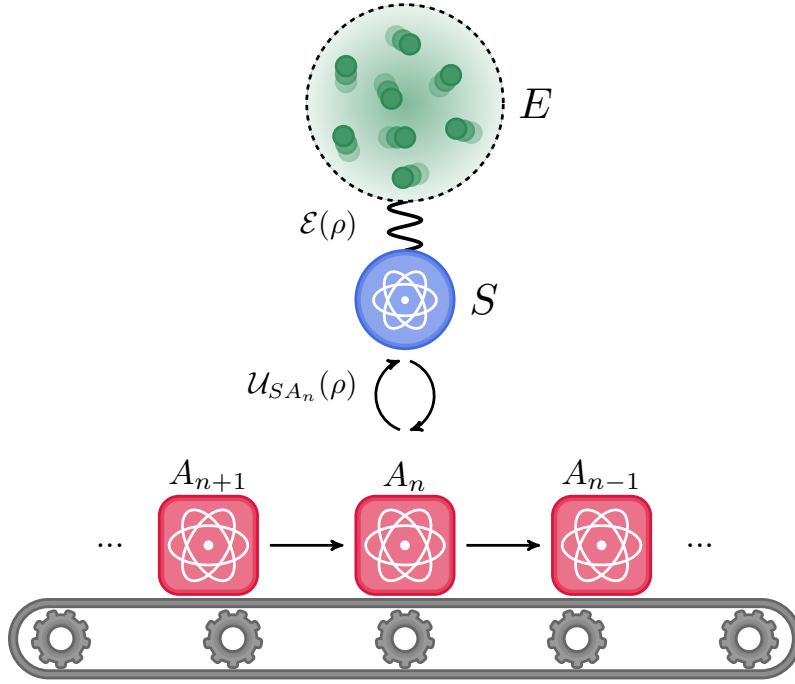
## 4.4 Collisional quantum thermometry

Throughout this chapter we have discussed the general aspects of quantum metrology and quantum thermometry, with a few concrete applications serving as examples. Our objective in this section is to introduce a concrete platform for thermometry, which will be the foundation for our investigations in Chapter 5. During this section, in particular, we shall discuss mostly model-specific results. Hence, we close this chapter by introducing the idea of *collisional quantum thermometry*, which employs collisional models in a thermometric setup. Collisional models have encountered a growing interest in the recent years in the field of open quantum systems [216–219]. Among important uses, we mention applications in quantum optics [25], quantum thermodynamics [220, 221] and, more recently, quantum batteries [222–224]. The reader might refer to Ref. [225] for a short and pedagogical introduction, and to Ref. [226] for a comprehensive review.<sup>11</sup>

We show now how collisional models can be tailored as quantum thermometry, closely following the original proposal for collisional quantum thermometry [155]. The basic idea behind collisional models is that an object of interest repeatedly interact (in other words, "collides") with other constituents of the model. In our case, an intermediate system, which we denote by  $S$ , is placed between two other constituents: the environment  $E$ , which usually refers to a thermal bath (at a temperature  $T$ ), and a trail of ancillae  $A_n$  which are eventually measured, as depicted in Fig. 4.8. One might notice that this setup is in contrast with the standard configuration for probe-based thermometry, where the ancillae (or probes), directly interact with the system of interest. This is precisely the objective of such a construction: the model from Ref. [155] is intrinsically out-of-equilibrium,

---

<sup>11</sup>The terms *collision models* or *repeated interaction models* often appear in the literature as well. Moreover, note how some adjacent formulations - which might not be strictly regarded as collisional models - also exist in the literature, such as the recent work from Ref. [227].



**Figure 4.8:** An illustration of the setup used in collisional quantum thermometry.

since the resulting ancillae are, in general, not fully thermalized. This departure from the standard strategy gives rise to two advantages. The first one is that the resulting states of the ancillae might better encode information about the temperature due to their dynamical relaxation rate, surpassing the TFI which we introduced in Eq. (4.37). This will be our main interest in the upcoming discussions. The second one is that we are also able to introduce correlations between the ancillae, which allow us to beat the SQL when collective measurements are used.

We shall consider a model where both the system  $S$  and the ancillae  $A_n$  are resonant qubits, described by  $H_S = \Omega\sigma_Z^S/2$  and  $H_{A_n} = \Omega\sigma_Z^{A_n}/2$ , respectively. Additionally, the interaction between the  $S$  and the environment  $E$  is described through the GKLS [228] master equation:

$$\frac{d\rho_S}{dt} = \mathcal{L}(\rho_S) = \gamma(\bar{n} + 1)\mathcal{D}[\sigma_-^S] + \gamma\bar{n}\mathcal{D}[\sigma_+^S], \quad (4.38)$$

where  $\mathcal{D}[L] = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}$  is the dissipator,  $\gamma$  is the coupling strength and  $\bar{n} = 1/(e^{\Omega/T} - 1)$  is the Bose-Einstein occupation. Therefore, we can write the  $SE$  interaction as described by the map  $\mathcal{E}(\rho_S) = e^{\tau_{SE}\mathcal{L}}(\rho_S)$ . The system-ancilla interaction, on the other hand, is chosen to be a partial-swap [229]: <sup>12</sup>

<sup>12</sup>This choice is not unique. For instance, Refs. [230] and [231] employ a different interaction, where

$$U_{SA_n} = \exp \left\{ -i\tau_{SA}g(\sigma_+^S\sigma_-^{A_n} + \sigma_-^S\sigma_+^{A_n}) \right\}. \quad (4.39)$$

By choice, all ancillae are initialized in the ground-state  $\rho_A^0 = |0\rangle\langle 0|$ . Additionally, the coupling strengths  $\gamma_{TSE}$  and  $g_{TSA}$  the free parameters of our model: they can be tuned in order to improve the QFI.

Moving forward, we formalize the repeated interaction that we have described before. The system undergoes what is called a *stroboscopic* evolution, which consists in an alternating application of the two maps above,

$$\rho_S^n = \text{tr}_{A_n} \{ \mathcal{U}_{SA_n} \circ \mathcal{E}(\rho_S^{n-1} \otimes \rho_A^0) \} := \Phi(\rho_S^{n-1}), \quad (4.40)$$

where  $n = 1, 2, 3, \dots$  labels the collisions. Here  $\mathcal{U}_{SA_n}(\bullet) = U_{SA_n} \bullet U_{SA_n}^\dagger$  and  $\circ$  denotes map composition. We will only consider states in the steady-state regime, in the sense of the stroboscopic dynamics from Eq. (4.40). In other words, several ancillae are let to interact with the system, which eventually reaches a fixed point  $\rho_S^* = \Phi(\rho_S^*)$ . This regime eliminates any transient effects and introduces a translational invariance with respect to the collisions, greatly simplifying the analysis.

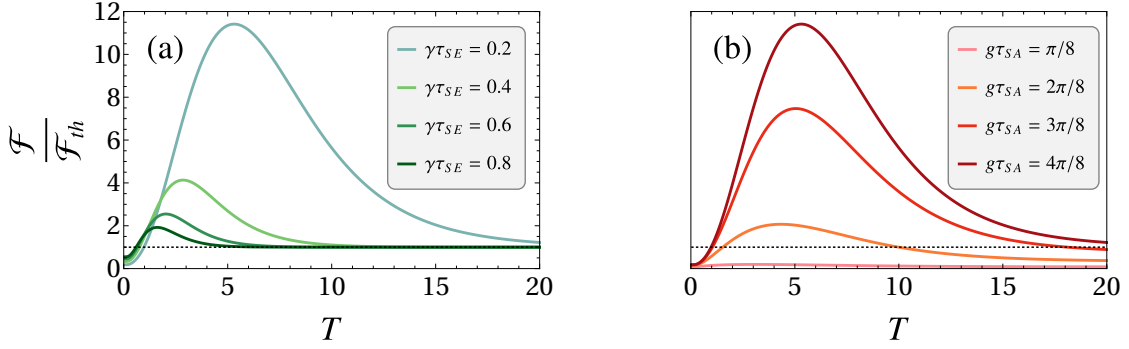
Now that the dynamics is set, we are able to discuss the limits of thermometric precision for the protocol. By following the procedures above, we calculate the steady-state of the *system* to be:

$$\rho_S^* = \begin{pmatrix} \frac{1}{(1+e^{\Omega/T})\left(1+\frac{\sin^2(g\tau_{SA})}{e^{\Gamma-1}}\right)} & 0 \\ 0 & 1 - \frac{1}{(1+e^{\Omega/T})\left(1+\frac{\sin^2(g\tau_{SA})}{e^{\Gamma-1}}\right)} \end{pmatrix}, \quad (4.41)$$

where  $\Gamma := \gamma(2\bar{n} + 1)\tau_{SE}$  is what we define as the thermal (or dynamic) relaxation rate. Note however that this is not our main object of interest. We perform measurements on the ancillae, not on the system. So our objective is to calculate the QFI (and the FI) associated with the ancillae after they interact with the system in the steady state above.

---

very analogous results can be obtained.



**Figure 4.9:** QFI in the collisional thermometry scheme. We plot the ratio between the quantum Fisher information and the thermal Fisher information for (a) full swaps and for (b)  $\gamma\tau_{SE} = 0.2$ . The dotted line corresponds to  $\mathcal{F} = \mathcal{F}_{th}$ .

That is, we are interested in the state

$$\rho_{A_n}^* = \text{tr}_S\{\mathcal{U}_{SA_n} \circ \mathcal{E}(\rho_S^* \otimes \rho_A^0)\} \quad (4.42)$$

which, from Eq. (4.41), explicitly evaluates to

$$\rho_{A_n}^* = \begin{pmatrix} \frac{1}{(1+e^{\Omega/T})(\csc^2(g\tau_{SA}) + e^{\frac{1}{\Gamma-1}})} & 0 \\ 0 & 1 - \frac{1}{(1+e^{\Omega/T})(\csc^2(g\tau_{SA}) + e^{\frac{1}{\Gamma-1}})} \end{pmatrix}. \quad (4.43)$$

Of special interest is the case of the full-swap, with  $g\tau_{SA} = \pi/2$ , where the ancilla and the system completely exchange states. In this case the ancillae state reduces to:

$$\rho_{A_n}^* \xrightarrow[\text{Swap}]{\text{Full}} \begin{pmatrix} \frac{1-e^{-\Gamma}}{1+e^{\frac{\Omega}{T}}} & 0 \\ 0 & 1 - \frac{1-e^{-\Gamma}}{1+e^{\frac{\Omega}{T}}} \end{pmatrix}, \quad (4.44)$$

which will be our case of interest from hereafter. One of the main results from Ref. [155] was to show that the QFI associated with the state (4.43) actually *surpasses* the TFI which we calculated in Eq. (4.27). More explicitly, they obtained the ratio

$$\frac{\mathcal{F}}{\mathcal{F}_{th}} = \frac{(\bar{n} + 1)(e^\Gamma + 2\bar{n}\Gamma - 1)^2}{e^{2\Gamma}(\bar{n} + 1) - e^\Gamma - \bar{n}}, \quad (4.45)$$

in the case of full-swaps. This can be done if we apply the technique from Eq. (4.12) to the state (4.44). We plot these results, including the numerical solutions for configurations

besides the full-swap, in Fig. 4.9.

The most striking aspect in this figure is that the non-equilibrium protocol may offer a very significant improvement over the default experiment. By comparing Figs. 4.7 and 4.9, the improvement is largest for intermediate temperatures (around  $2 \lesssim T \lesssim 10$ ), which is precisely the region where the TFI quickly falls off. This improvement can be attributed to the terms containing the parameter  $\Gamma$ , appearing in Eqs. (4.41) - (4.44). Notice that these terms vanish for fully thermalized ancillae. In other words, we can say that the thermal relaxation parameter  $\Gamma$  helps to better encode the temperature in those states, furnishing us with information beyond what the populations of the thermal state typically offers [155]. Another important aspect, mentioned in Ref. [155], is that for large  $\Gamma$ , the ratio in Eq. (4.45) behaves as  $\frac{\mathcal{F}}{\mathcal{F}_{th}} \approx 1 + \Gamma \bar{n} e^{-\Gamma} > 1$ . Thus, even though the ratio inevitably tends to unity for large temperatures, it is *always* advantageous to opt for this scheme, as long as an advantage over the thermal sensitivity is concerned.

Furthermore, the particular equations that we have obtained for the steady-state, and the corresponding QFI, are conditioned on the choice of ancillae which are initially on the ground-state. Other choices, such as ancillae prepared in the excited state, would yield different results. Notwithstanding, the qualitative conclusions and general considerations about the model are the same regardless of most input states that we choose for the probes. A careful consideration on this point has been given in Ref. [230], which further expands a few aspects from Ref. [155]. In particular, the authors obtain the optimal ancillae state with respect to the parameters of the model (including the temperature  $T$  itself). One of their conclusions is that ancillae initialized in the ground state are either optimal, or at least nearly-optimal, for a wide-range of regimes. For that reason, in the upcoming chapter we will only consider ground-state ancillae and the analytical results above, since other choices offer no substantial differences.

# Chapter 5

## Bayesian quantum thermometry

Throughout this thesis we have become well aware of the conceptual and practical relevance of the quantum Fisher information as a central concept in quantum metrology and quantum estimation theory. Our concerns with optimality in chapter 4 revolved entirely around the calculation of the QFI for different models. In particular, we were interested in the suitability of a collisional model as a platform for quantum thermometry [155] and how its non-equilibrium configuration could be used to provide an advantage over the standard probe-based schemes, which was mainly quantified through the QFI. This type of examination is useful due to its universality: they provide important benchmarks and a global view of the problem, independent of the particular choice of estimators.

Nonetheless, the post-processing aspects are complementary and equally important. Those are vital when discussing concrete thermometry protocols and, as discussed earlier in chapter 2, maximum likelihood methods constitute the default choice as far as concrete estimators are concerned [232–234]. Moreover, further layers of complexity are introduced whenever physical constraints or symmetries of the problem are considered [119, 235].<sup>1</sup> In the grounds of quantum thermometry, concrete implementations and practical choices of estimators have been a concern in recent works and, as we discussed in Chapter 3, the Bayesian theory provides a very powerful framework in this sense [176, 236–238].<sup>2</sup> To the best of our knowledge, the first use of Bayesian estimation tailored

---

<sup>1</sup>As a practical example, consult the Sec. 6 of Ref. [128], where they discuss different problems in quantum metrology, such as the phase-estimation in a quantum harmonic oscillator. This example also shows how insisting on unbiased estimators might result in unphysical scenarios.

<sup>2</sup>The reader may consult Ref. [239] for a very thorough discussion on the differences and the application of frequentist and Bayesian statistics in quantum metrology, albeit with some particular considerations for

for quantum thermometry has been provided in Ref. [240]<sup>3</sup>, with many important and adjacent investigations following shortly after [24, 27, 28, 242–244].

Our objective in this chapter is to build a bridge between collisional quantum thermometry and bayesian inference. Bayesian estimation (BE) provides a very natural framework for a sequential updating schemes, where one is able to continuously update the state of knowledge about the temperature distribution in a experimentally friendly manner. From these, it is very simple to implement concrete estimators with the aid of the Bayesian tool set [24, 240]. This chapter concludes our investigations about estimation theory, quantum metrology and thermometry, encompassing original results of ours published in [245].

## 5.1 Ancilla-ancilla correlations

In Sec. 4.4 we summarized many of the physical features for collisional thermometry provided in Refs. [155] and [230]. However, an important point which was left off from the discussion in Sec. 4.4, pertains the non-local aspects of the model and the correlations between the ancillae. The interaction of the ancillae  $A_n$  with the intermediate system  $S$  will typically introduce correlations between them in varying degrees. To examine this point in more detail, let us consider a block  $A_i \dots A_{i+n}$  of  $n$  ancillae which interact with  $S$  after it reaches its steady state  $\rho_S^*$ . That is, in analogy with Eq. (4.42) for a single ancilla, consider instead the collective state:

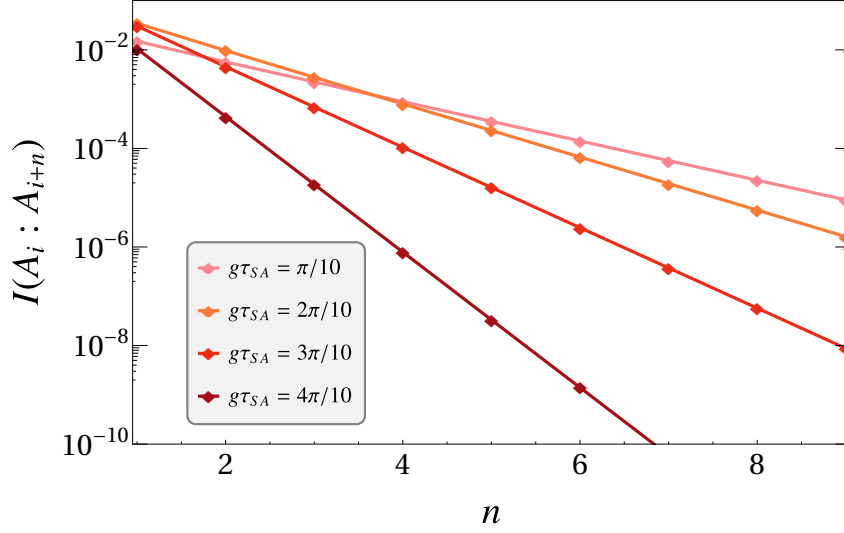
$$\rho_{A_i \dots A_{i+n}} = \text{tr}_S \left\{ \mathcal{U}_{SA_{i+n}} \circ \mathcal{E} \circ \dots \circ \mathcal{U}_{SA_i} \circ \mathcal{E} (\rho_S^* \otimes \rho_A^0 \otimes \dots \otimes \rho_A^0) \right\}. \quad (5.1)$$

Remember that  $\rho_{A_i \dots A_{i+n}}$  is independent of  $i$  due to the translational invariance of steady-state regime. Here, we have the freedom of choosing joint measurements of several ancillae, or even the whole block, at once. However, besides the experimental challenges, such type of measurement inevitably make the theoretical analysis more involved. For that reason, we focus on single-ancilla measurements instead. If we denote the set of POVMs

---

the problem of phase estimation. See also Ref. [235] for a more general review on Bayesian statistics and its applications in physics.

<sup>3</sup>In the case of *classical* thermometry, Prosper was able to employ Bayesian estimation in his seminal work, putting forth temperature estimation as a scale estimation problem, in an early study [241].



**Figure 5.1:** A log-linear plot for the correlation between an ancilla and its  $n$ -th neighbor for different  $SA$  coupling strengths. We take  $T/\Omega = 2$  and  $\gamma\tau_{SE} = 0.2$ .

by  $M_X$ , with possible outcomes  $X = 0, 1$ , we can associate the pdf

$$P(X_n, \dots, X_1|T) = \text{tr}\{M_{X_n} \dots M_{X_1} \rho_{A_1 \dots A_n}\}. \quad (5.2)$$

with the state (5.1). For that reason, we use the mutual information, defined as:

$$I(A_i : A_{i+n}) = S(\rho_{A_i}) + S(\rho_{A_{i+n}}) - S(\rho_{A_i A_{i+n}}), \quad (5.3)$$

where  $S(\rho) = -\text{tr}(\rho \ln \rho)$  is the von Neumann entropy, to quantify the correlations which can arise between different ancillae. We show the result in Fig. 5.1, plotting  $I(A_i : A_{i+n})$  as a function of  $n$ . We can observe that the correlations decay exponentially with each neighbor. Additionally, the rate at which they decay depends on the  $SA$  decoupling, decaying faster for ancillae near the full-swap regime (where the correlations eventually vanish). Therefore, the pdf associated with the measurement of the ancillae can be written, to a good approximation, as

$$P(X_n, \dots, X_1|T) \approx P(X_n|T) \dots P(X_1|T), \quad (5.4)$$

where

$$P(X_i|T) = \text{tr}(M_{X_i} \rho_{A_i}). \quad (5.5)$$



We stress that such an approximation is not an requirement from a Bayesian point of view, but merely a simplification of the problem. For a brief discussion on how the correlations should be computed, consult Appendix B.

## 5.2 Numerical analysis

We can now introduce numerical apparatuses which will allow us to employ the Bayesian updating scheme for the collisional thermometry. The procedure here is very analogous to the examples discussed in chapter 3. Given the outcomes  $\mathbf{X} = (X_1, \dots, X_n)$ , which are obtained by measuring the ancillae  $A_1, \dots, A_n$ , we can update the state-of-knowledge about the temperature though the Bayes theorem:

$$P(T|\mathbf{X}) = \frac{P(\mathbf{X}|T)P(T)}{P(\mathbf{X})}. \quad (5.6)$$

The prior is usually chosen either by mathematical convenience or by experimental-motivations. An instance of the latter case is when one might be aware of typical experimental intervals, such as in, e.g., Bose-Einstein condensates [246], or even previous iterations of similar experiments. Since we are interested in a purely theoretical analysis, we shall focus on nearly-flat priors, which we shall define later on. In terms of figures of merit, we shall use the MSE and the BMSE, defined in Eqs. (2.8) and Eq. (3.20), respectively. For convenience, we rewrite them here, in terms of the temperature, as

$$\epsilon(\hat{T}(\mathbf{X})|T) = \int (T - \hat{T})^2 P(\mathbf{X}|T) d\mathbf{X}, \quad (5.7)$$

and

$$\epsilon_B(\hat{T}(\mathbf{X})) = \int \epsilon(\hat{T}(\mathbf{X})|T) P(T) dT. \quad (5.8)$$

Likewise, we will also pick the Bayesian average from Eq. (3.21) as our estimator of choice, which minimizes the BMSE (3.20), in the mean-square sense:

$$\hat{T}(\mathbf{X}) = \int T P(T|\mathbf{X}) dT \quad (5.9)$$

We further stress that the posterior loss (3.34) is sometimes regarded as the fully Bayesian

error, due to the fact that it is conditioned on the actual realization of the experiment, so this is a quantity which might be of great interest for the experimenter whenever concrete data is available. We also mention the MAP estimator (3.23) as an equally simple alternative, or even the median estimator, defined in Eq. (3.23), as a further option. However, the problem with the latter is that the numerical computation for the median is less straightforward than the other two choices. It is not our objective here to dwell into such technical discussions regarding the estimators themselves. A very thorough investigation regarding how different Bayesian estimators - and also different priors - can be used in the collisional thermometry scheme has been done in [24]. The authors also give a very special care to *uninformative* priors in this context.

While more efficient methods might be available, we describe our procedures for numerically simulating Eqs. (5.6) - (5.8). Our objective is to sequentially compute the posterior (5.6) from a set of  $n$  random outcomes  $\mathbf{X}$ , with the possibility of easily updating the distribution as new data arrives [236, 247]. Additionally, we also need to be able to compute any other corresponding quantities, such as the moments of the distribution. However, two complications pop up in regard to a naive numerical implementation: we have to deal both with a large number of multiplying probabilities in  $P(X_1|T) \dots P(X_n|T)$  and with the computation of the normalization factor  $P(\mathbf{X}) = \int P(\mathbf{X}|T)P(T)dT$  from Eq. (5.6).

We start by discretizing the interval of interest  $[T_{\min}, T_{\max}]$  into  $N_T$  points  $T_k$ . Here we denote the lower and upper cutoffs by  $T_{\min}$  and  $T_{\max}$ , respectively. By doing so, the prior distribution is discretized into a distribution  $P_k$  into an analogous manner. Similarly, we can discretize the log-likelihood as

$$L_{kn} = \sum_{i=1}^n \ln P(X_i|T_k). \quad (5.10)$$

Using the log-likelihood here instead of the standard likelihood provides a stable countermeasure against the increasingly smaller probabilities. For practical purposes, we can also visualize  $L_{kn}$  as a matrix (or a *grid*) of dimensions  $N_T \times n$ . For instance,  $L_{k,3}$  describes the log-likelihood of the string  $X_1, X_2, X_3$ . Also notice how  $L_{kn} = L_{k,n-1} + \log P(X_n|T_k)$ , which can be incremented sequentially with each new piece of data. In this discretized

form, the Bayesian updating from Eq. (5.6) can now be written as

$$P_{k|n} = \frac{e^{L_{kn}} P_k}{\sum_q e^{L_{qn}} P_q}, \quad (5.11)$$

where  $P_{k|n}$  is a shorthand for  $P(T_k|X_1 \dots X_n)$ .

Furthermore, by defining and using the max of the log-likelihood at each  $n$ , as  $L_n^{\max} = \max_k L_{kn}$ , we circumvent the numerical instabilities. Namely, we rewrite Eq. (5.11) as

$$P_{k|n} = \frac{e^{L_{kn} - L_n^{\max}} P_k}{\sum_q e^{L_{qn} - L_n^{\max}} P_q}. \quad (5.12)$$

This modification ensures that the most likely events will have the best numerical precision. By doing so, Eq. (5.12) assumes the form

$$P_{k|n} = \frac{P_{kn}}{\sum_q P_{qn}}, \quad (5.13)$$

where  $P_{kn} = e^{L_{kn} - L_n^{\max}} P_k$  can be interpreted as a grid of dimensions  $N_T \times n$ , which is readily constructed from the log-likelihood matrix  $L_{kn}$  and the vector  $P_k$  which describes the discretized prior.

This procedure allows us to readily compute any interest quantity of interest from the numerical posterior. In particular, the BA (5.9) is given by:

$$\hat{T}_n = \sum_k T_k P_{k|n}. \quad (5.14)$$

We can obtain the MSE [Eq. (5.7)] by sample averaging the square error by  $(\hat{T}_n - T_0)^2$  over multiple realizations, where we draw the vector  $\mathbf{X}$  according to the likelihood  $P(\mathbf{X}|T)$ , that is:  $\mathbf{X} \sim P(\mathbf{X}|T) = P(X_n|T) \dots P(X_1|T)$ . Meanwhile, we obtain the BMSE (5.8) in a similar manner, but with an extra step where we generate the data  $\mathbf{X}$  from the likelihood  $P(\mathbf{X}|T)$  by randomly sampling the temperature  $T$  from the prior  $P(T)$ .

### 5.3 Bayesian estimation and collisional thermometry

We are finally equipped with all the tools necessary for applying the Bayesian framework into collisional thermometry. As we have mentioned, choosing an appropriate prior is a very extensive topic in Bayesian inference [58], specially when physical considerations or further constraints are taken in to account [24, 235, 240, 241]. Since we are interested in generic non-equilibrium scenarios and in numerical simplicity, we choose near-flat prior over the interval  $[T_{\min}, T_{\max}]$ . More specifically, we employ [239]

$$P(T) = \frac{1}{(T_{\max} - T_{\min})} \lambda_{\alpha} \left( \frac{T - T_{\min}}{T_{\max} - T_{\min}} \right), \quad (5.15)$$

where

$$\lambda_{\alpha}(\theta) = \frac{e^{\alpha \sin^2(\pi\theta)} - 1}{e^{\alpha/2} I_0\left(\frac{\alpha}{2}\right) - 1}, \quad (5.16)$$

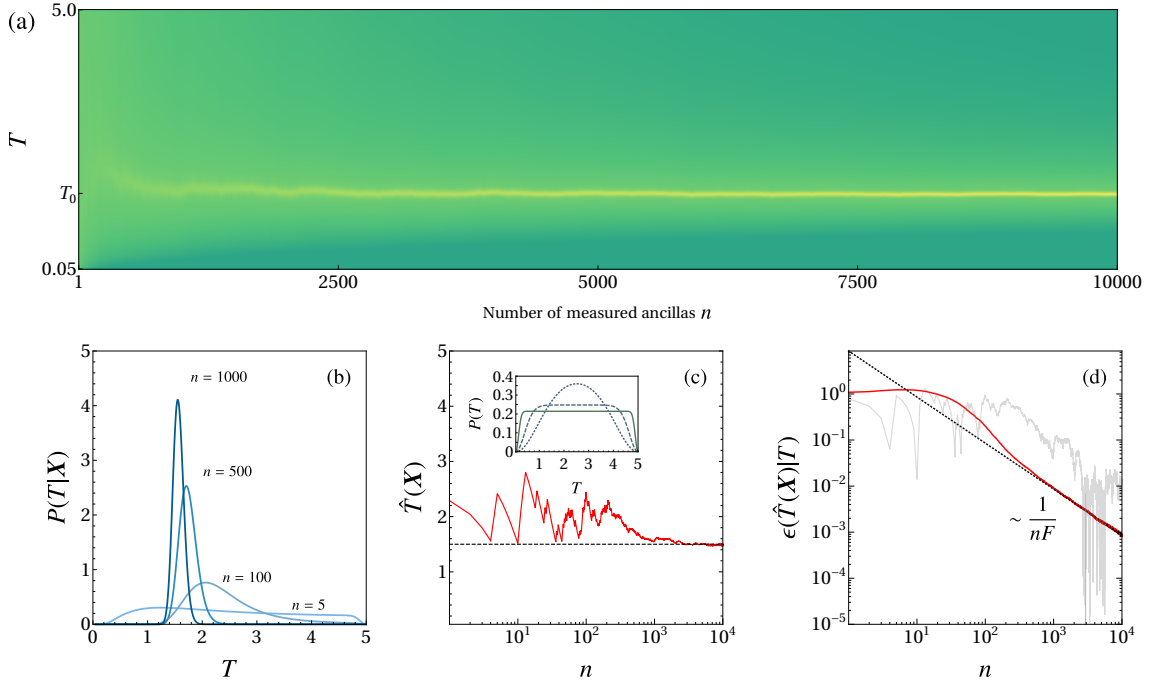
and  $I_0$  is the modified Bessel function of the first kind. A major reason for choosing this distribution lies in a technicality concerning the VTSB: the bound, in its standard form, does not hold for priors which do not vanish at the endpoints, such as the flat prior itself. Choosing the distribution (5.15) avoids this problem [112, 121]. Moreover, this prior very conveniently allows us to smoothly interpolate between flat and sharp distributions, without altering the endpoints. This can be done by changing  $\alpha$ , as shown on the inset of Fig. (c). The distribution is flat-like for  $\alpha$  large and *negative*. Conversely, taking  $\alpha > 0$  results in increasingly sharper distributions.

We summarize the basic results in Fig. 5.2. We start by fixing the value of the true temperature  $T_0$ , which is then used to generate a string of random outcomes  $X_i$  from the likelihood  $P(X_i|T = T_0)$ . We consider the system to be in the full-swap regime  $g\tau_{SA} = \pi/2$ , so we obtain the steady from state (4.44). In this case, the likelihood reads

$$P(X_i = 1|T) = \frac{1 - e^{-\Gamma}}{1 + e^{\frac{\Omega}{T}}} \quad (5.17)$$

when we perform measurements in the computational basis, which conveniently happen to be optimal for this class of states, as we discussed in Sec. 4.4. This also implies that the FI  $F(T)$  associated with the likelihood above coincides with the QFI given in (4.45).

In Fig. 5.2 (a) we plot the posterior distribution  $P(T|X_1 \dots X_n)$  [Eq. (5.12)]. The



**Figure 5.2:** Bayesian estimation for collisional thermometry. The true temperature was chosen as  $T_0/\Omega = 1.5$ . (a) Density plot of the posterior  $P(T|X_1 \dots X_n)$  [Eq. (5.12)]. (b) Same, but as a function of  $T$ , for select values of  $n$ . In both figures, the distribution is clearly seen to converge towards  $T_0$  as  $n$  increases. (c) Random realization of the BA (5.9), as a function of  $n$ . Inset: prior distribution (5.15), for  $\alpha = -1$  (dotted),  $\alpha = -10$  (dashed) and  $\alpha = -100$  (solid). (d) MSE [Eq. (5.7)] for a single stochastic realization (gray), and averaged over multiple realizations (red). For large  $n$  it converges to  $1/nF(T_0)$  (dotted), which saturates the CRB. All curves were plotted using  $\gamma_{\tau_{SE}} = 0.4$ ,  $g_{\tau_{SE}} = \pi/2$  and  $\alpha = -100$ . The temperature was discretized in steps of  $N_T = 500$ , from  $T_{\min} = 0.05$  to  $T_{\max} = 5$ . Caption from [245].

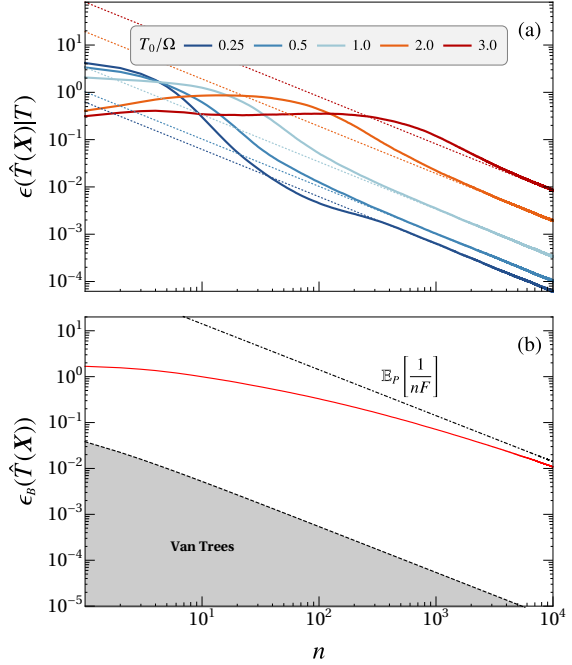
temperature is represented along the vertical axis, while the horizontal axis shows the number of measured ancillae. A similar visualization scheme for this sequential updating strategy has been given in Refs. [247] and [236], for problems in quantum metrology and open quantum systems. Meanwhile, Fig. 5.2 (b) plots the same results as a function of  $T$  and for a few selected values of  $n$ . We can explicitly see how it gradually peaks around the true value of the parameter. Additionally, we can also verify that it indeed starts to resemble a Gaussian for large  $n$ , with variance  $1/nF(T)$ , as predicted by the Bernstein–von Mises theorem, discussed in Sec. 3.3 [Eq. (3.47)]. Similarly, in Fig. 5.2 (c) we explicitly plot a stochastic realization of the estimator. The error associated with this particular realization of  $\hat{T}(\mathbf{X})$  in is plotted as a gray curve in Fig. 5.2 (d). From these panels we can see how the estimator converges to the true value of the temperature as

the number of measured ancillae increases (notice the log scale). Its average over multiple realizations  $\mathbf{X}$  results in the MSE (5.7), plotted as the solid red line in Fig. 5.2 (d). We can explicitly observe that the estimator is *asymptotically* efficient, as the MSE eventually converges to  $1/nF(T_0)$ , shown by the dotted line, saturating the CRB.

Another very important aspect is how well a protocol fares for different temperatures. Thus, in Fig. 5.3 (a) we plot the MSE for different values of  $T_0$ . The dashed lines highlight the asymptotic limit  $1/nF(T_0)$ . We can see that the estimation is more accurate as the temperature decreases, which is attributed to the larger sensitivity on  $T$  in the likelihood Eq. (5.5).

As we have thoroughly discussed in chapters 2 and 3, the MSE and the FI often depend on the very parameter being estimated. In the same manner, the MSE in Fig. 5.3 (a) depends on the *unknown* value of the temperature  $T_0$ . This is one of the situations where the Bayesian framework shines as a convenient tool. The BMSE  $\epsilon_B$  from Eq. (5.8) gives us a more concise and global view of the problem, averaging the accuracy of the protocol over the range  $[T_{\min}, T_{\max}]$  of temperature, which is independent of particular value of the parameter being estimated. This is shown in Fig. 5.3 (b).

We also compare this curve for the BMSE with the VTSB (3.46), depicted by the gray region in Fig. 5.3 (b). Notice, for reasons pointed in Sec. 3.3, that the bound is not saturated, even in the asymptotic limit. Instead, the BMSE converge to the average



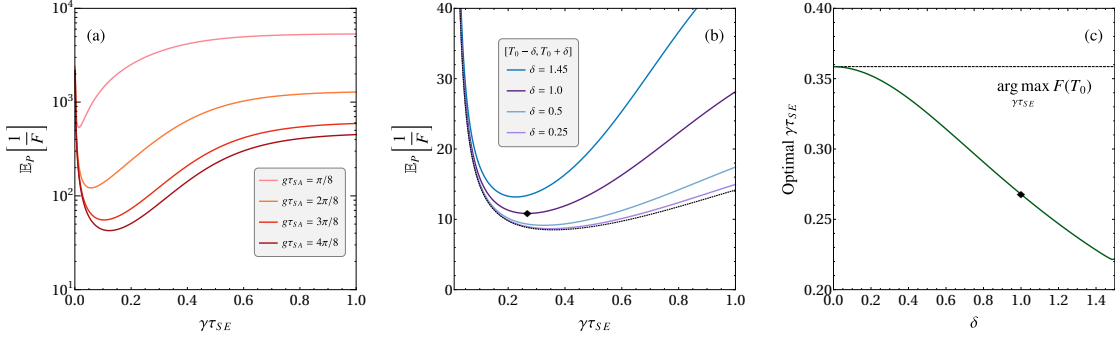
**Figure 5.3:** (a) MSE [Eq. (5.7)] for different values of  $T$ , averaged numerically over 3000 different trajectories. The dotted lines correspond to the CRB for the different temperatures shown in the legend. (b) We repeat the procedure in (a), averaging the MSE over 500 trajectories and also over the prior distribution (5.15), obtaining the BMSE [Eq. (5.8)]. The integral over the prior is also performed numerically through a temperature discretization with  $N_T = 150$ . We also show the asymptotic limit defined in (3.48) (dot-dashed). Other parameters are the same as in Fig. 5.2. Caption from [245].

(over the prior) of the reciprocal FI, per Eq. (3.48). In other words, the dashed line in Fig. 5.3 (b) to which the BMSE converges is simply an average of the lines depicting the CRB in Fig. 5.3 (a), but averaged over the prior. We could say that the asymptotic value of the *Bayesian* MSE is, in a sense, described as an "expected" Cramér-Rao bound. Therefore, computing the asymptotic BSME will give a succinct and sufficiently global view of the aspects of this scheme henceforth. More specifically, we can investigate the behavior of  $\epsilon_B$  for large  $n$  as we change the system parameters.

With that in mind, we plot the asymptotic value of the BMSE from Eq. (3.48) in Figs. 5.4 (a) and (b). In the panel (a) we describe the asymptotic accuracy of the protocol as a function of  $\gamma_{\mathcal{T}_{SE}}$  for different values of the  $SA$  coupling. Naturally, since we are employing a Bayesian figure of error here, this plot is *independent* of the temperature, and what we have here is an *average* behavior (with respect to the prior). What is relevant in this case is the temperature interval, that is, the cutoff in the temperatures  $T_{\min}$  and  $T_{\max}$ , and also the prior. A smaller value of  $\mathbb{E}_P[1/nF]$  results in a more accurate estimation, as seen in the asymptotic behavior of the BMSE, described in Eq. (3.48). Very interestingly, this plot shows us that (i) increasing  $g\tau_{SA}$  always yields better results and also that (ii) the accuracy ultimately depends on the value of  $\gamma_{\mathcal{T}_{SE}}$ . The latter has also been briefly discussed in Ref. [24].

As we mention above, the achievable precision - in the Bayesian sense - depends on the prior and the chosen interval. Here we try to establish a relationship between the parameters of the model and temperature cutoff. In Fig. 5.4 (b) we plot  $\mathbb{E}_P[1/F]$  as a function of the  $SE$  coupling, considering a symmetric interval from  $T_{\min} = T_0 - \delta$  to  $T_{\max} = T_0 + \delta$ , centered at  $T_0/\Omega = 1.5$  for different values of  $\delta$ . That is,  $2\delta$  can be interpreted as the "interval width".

We can see that in either panel, the optimal value of  $\gamma_{\mathcal{T}_{SE}}$  depends on the other parameters of the model, such as  $g\tau_{SA}$  and the interval in consideration. In particular, we can notice from this plot that larger intervals are more sensitive to  $\gamma_{\mathcal{T}_{SE}}$ : the optimal regime is narrower, and the error quickly increases with sub-optimal choices. For larger intervals, we should tune  $\gamma_{\mathcal{T}_{SE}}$  to smaller values in order to achieve the optimal regime. Conversely, we can notice that as the interval narrows, we recover the results found for the a deterministic temperature  $T_0$ , in the same spirit of Fig. 4.9 (a) from the previous



**Figure 5.4:** (a) We plot the expectation in the RHS of Eq. (3.48) as a function of  $\gamma\tau_{SE}$  considering the same temperature range  $[T_{\min}, T_{\max}]$  from Fig. 5.2. (b) We fix the SA coupling in the full swap regime  $g\tau_{SA} = \pi/2$  and plot Eq. (3.48) for different temperature intervals. Note how the minima shifts to the left as  $\delta$  increases. (c) We plot the optimal value of  $\gamma\tau_{SE}$  as we increase the size of the interval  $[T_0 - \delta, T_0 + \delta]$ . We also show the optimal value (dashed) for  $F(T_0)$  at the particular temperature  $T_0/\Omega = 1.5$ . The diamond symbol highlights the minimum for  $\delta = 1$ . The prior used here is also given by Eq. (5.15), with  $\alpha = -100$ , but the endpoints are changed as described above. Caption from [245].

chapter. For instance, notice from Fig. 5.4 (c) how the optimal parameters continuously decrease as  $\delta$  increases. We also mention that it is known that the optimal value of  $\gamma\tau_{SE}$ , in terms of the QFI and the FI, actually depends on the temperature itself [24], so this is one of the scenarios where an adaptive strategy could shine: by continuously performing measurements and updating the state-of-knowledge about the temperature  $T$ , the coupling parameter  $\gamma\tau_{SE}$  could also be changed accordingly. In that sense, the work done in Ref. [231] provides a very clever implementation. There, the authors employ a *stochastic* interaction time  $\tau_{SE}$  between the system and the environment.<sup>4</sup> By doing so, one can obtain a better regime of operation for a broader range of parameters. This might also avoid the necessity of being certain of the particular value of the parameter. Thus, while there might be some loss with respect to a deterministic QFI, by employing random collision times we average out, in a sense, the sub-optimal effect of being unaware of the exact temperature of the bath and the best choice of parameters.

The current discussion showcases how the BMSE might be helpful in the optimization of thermometry protocols. As we have made clear in Chapters 2 and 3, and also with Fig. 5.3 in the current problem, the Fisher information and the CRB both depend on the

<sup>4</sup>Curiously, it has been shown recently that it is possible to link such a scenario of stochastic waiting-times to collisional models where ancilla-ancilla correlations are present [248]. This type of implementation is often associated with non-Markovian behaviour [249, 250].



unknown temperature of the bath. Moreover, as shown in Fig. 5.4, this temperature dependence is also true for the optimal values of  $\gamma_{\tau_{SE}}$  and  $g_{\tau_{SA}}$ . This leads us to a conundrum, where the optimal configuration of the protocol depends on the very parameter we are trying to estimate. This is a hurdle which was similarly made clear in Ref. [230] in regard to the optimization of the probe states. The Bayesian framework avoids this problem with an alternative formulation, focusing on an entire range of temperatures, whose likelihood is quantified by the prior  $P(T)$ . In this sense, we can also see the prior as quantifying a certain "degree of importance" to different temperatures: when calculating the BMSE and the optimal parameters, greater consideration is given to temperatures in which the system is more likely to be found. If the system is unlikely to be found at a certain interval of temperatures, there is no reason to take it into much consideration when performing any type of optimization. Additionally, this scenario also makes clear why the BMSE might also be called a *preposterior* error: it is a quantity of special importance *before* any experimental run is performed (hence, when the posterior distribution and any information which might come with it, is unavailable). When the parameter is unknown and no data is available, it makes sense to use the BMSE: it is a figure of merit which averages out over all the possible experiments, by integrating over all the possible values of  $T$  (weighted by the prior) and also over all the possible stochastic realizations (weighted by their likelihood). This is a scenario where we are yet to acquire any experimental data and where no knowledge about  $T$ , besides what is contained within the prior, is present. Thus, this type of analysis might be very valuable from an experiment design point of view and for optimization purposes (in the Bayesian sense), as also discussed in Refs. [240] and [24]. Hence, by focusing on the asymptotic BMSE ( $n \rightarrow \infty$ ), as compared to the asymptotic MSE  $1/nF(T)$ , we were able to show in Fig. 5.4 how the BMSE in Fig. 5.3 (b) can be optimized over  $\gamma_{\tau_{SE}}$  and  $g_{\tau_{SA}}$ , to yield a strategy which is good for the entire temperature range. Therefore, even though the temperature is in principle unknown, the Bayesian approach provides us with a strategy which, while possibly sub-optimal for a given temperature, will work well *on average*, avoid more situations where the estimation is too imprecise.

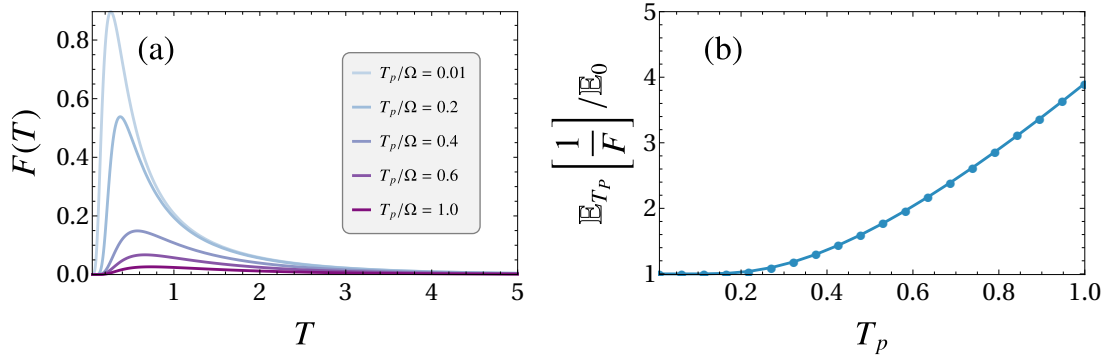
## 5.4 Effect of noisy probes

We conclude our investigations by discussing the effect of noisy ancillae in the protocol. Initially we assumed an ideal scenario where one has perfect control over the probe state. Now we further generalize our approach to the context where the observer cannot always initialize the ancilla in the desired state. For concreteness, we first begin with a scenario where we assume that the ancillae are initialized in a thermal state, investigating how the temperature of the probes affect the asymptotic precision of the estimation, in the same spirit of Fig. 5.4. In this more general scenario, the likelihood now assumes the form

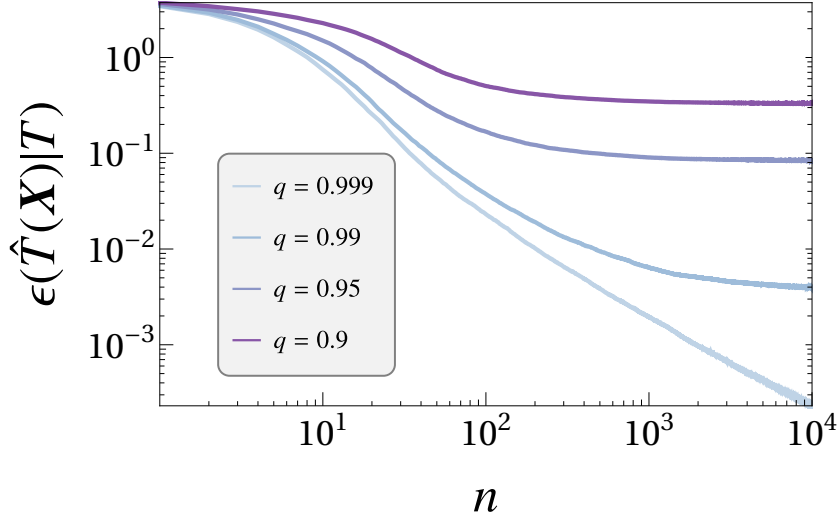
$$P_{T_p}(X_i = 1|T) = \frac{e^{-\Gamma}}{1 + e^{\frac{\Omega}{T_p}}} + \frac{1 - e^{-\Gamma}}{1 + e^{\frac{\Omega}{T}}} \quad (5.18)$$

instead, replacing Eq. (5.17). As a consequence of the linearity of the stroboscopic channel [Eq. (4.40)], the resulting likelihood for the thermalized probe is a convex combination of the likelihood associated with ancillae initialized in the states  $|0\rangle\langle 0|$  and  $|1\rangle\langle 1|$ , weighted by their Gibbs probabilities [Eq. (4.36)]. Furthermore, the convexity of the FI [251] implies that the resulting precision will be smaller for noisy probes, when compared to ideal probes initialized in the ground state.

A quantitative description can be given in terms of the asymptotic value of the Bayesian error given by Eq. (3.48), in the same vein of Fig. 5.4. We first calculate the asymptotic



**Figure 5.5:** (a) We plot the Fisher information of the likelihood in Eq. (5.18) for different values of  $T_p$ . (b) We calculate the ratio between the asymptotic Bayesian risk (3.48) obtained by integrating the FI in (a), from  $T_{\min} = 0.1$  to  $T_{\max} = 5$  and the asymptotic Bayesian risk for  $T_p = 0$ . All the other parameters are the same as in Fig. 5.2. Caption from [245].



**Figure 5.6:** Bayesian MSE (5.8) calculated for outcomes generated from the likelihood in Eq. (5.19) for different values of  $q$ . This type of bias introduces a systematic error; the Bayesian risk initially decreases with a  $1/n$  scaling but eventually saturates, since the estimation converges to a wrong value of temperature. All the other parameters in the simulation are the same as in Fig. 5.2. Caption from [245].

error for ground-state ancillas, which we denote by  $\mathbb{E}_0$ . This quantity can be obtained from the ideal FI given by Eq. (4.45). In Fig. 5.5 (a) we show how the Fisher information  $F(T)$  depends on the temperature  $T_p$  of the probe. Meanwhile, in panel (b) we show how much precision is lost for non-ideal probes, i.e., we calculate the ratio between the asymptotic factors  $\mathbb{E}_{T_p}[1/F]$  and  $\mathbb{E}_0$  as a function of the probe temperatures.

We finish by analyzing a second, but similar, scenario: we are interested in what happens when there is an uncertainty, or systematic error, in the preparation of the ancillae. Namely, the ancillae prepared either in the states  $|0\rangle\langle 0|$  and  $|1\rangle\langle 1|$  with classical probabilities  $q$  and  $1 - q$ , respectively, but the experimenter is unaware of their value. Mathematically, this scenario is described by the likelihood

$$P(X_i|T) = qP(X_i|T, \rho_{A,0}) + (1 - q)P(X_i|T, \rho_{A,1}), \quad (5.19)$$

where  $\rho_{A,k} = |k\rangle\langle k|$ , with  $k = 0, 1$ . Nevertheless, the estimation is still performed with respect to the ideal model from Eq. (5.17). In other words, what we are describing here is a scenario where we perform inference using an *incorrect*, or imprecise, model. This introduces a persistent error into the estimation, as we show in Fig. 5.6. Since the

experimenter is not using the proper model for the likelihood, the resulting estimation diverges from the correct parameter. The error eventually saturates at a certain value, which will correspond to the difference between the true value  $T_0$  of the temperature and the temperature one would get from the ideal likelihood (5.17) for a given realization  $\mathbf{X} = (X_1, \dots, X_n)$ .

## **Part II**

# **Time-optimal holonomic quantum computation**

## Chapter 6

# Geometrical phases in quantum mechanics

During the mid-twentieth century, Ehrenberg and Siday [252, 253], and later on Aharonov and Bohm [254, 255], were responsible for laying the groundwork [256] for some of the most influential results on non-local aspects of quantum theory during the earlier days of QM. These pioneering investigations were fundamental to Berry's seminal work [257], where he showed the physical consequences of adiabatic geometrical phases (GP) and how they arise in quantum mechanical systems. <sup>1</sup> Further advancements were made by Wilczek and Zee, in a generalization to the degenerate case, showing how a non-Abelian structure emerges. A few years later, important contributions by Aharonov and Anandan [262, 263] also came out, further expanding upon the non-Abelian geometric phase from Wilczek and Zee [264] to the non-adiabatic case. In this section, we briefly review the results by Berry, Aharonov and Anandan. Later on, their results will be important for us in order to motivate the construction of geometric quantum gates. For a review on GPs in quantum information, see [265, 266]. A rigorous formulation of GPs in terms of differential geometry is given in [267].

---

<sup>1</sup>An earlier description of the phenomenon was given by Pancharatnam [258] for polarized light in terms of the Poincaré sphere [259]. T. Kato also made significant observations about the adiabatic theorem and its geometrical aspects during the early 50's [260]. However, the physical relevance of the geometrical phases as we know today still went unnoticed by him at the time. A thorough discussion and the historical considerations of his work can be found in [261].

## 6.1 The Berry phase

The Berry phase is a geometrical phase which can emerge when the system evolves adiabatically, i.e., when the Hamiltonian changes very slowly. By doing so, the evolution in the Hilbert space follows the instantaneous eigenstates of the Hamiltonian. We discuss an example of this regime in Fig. 6.1.

More concretely, let us consider a Hamiltonian  $H(\mathbf{R}(t))$  which depends on a vector of parameters  $\mathbf{R}(t) = (R_1, R_2, \dots)$ . Initially, we also consider a non-degenerate case. The eigenstates  $|n(\mathbf{R}(t))\rangle$  of this Hamiltonian will naturally depend on the parameters:

$$H(\mathbf{R}(t)) |n(\mathbf{R}(t))\rangle = \epsilon_n(\mathbf{R}(t)) |n(\mathbf{R}(t))\rangle. \quad (6.1)$$

According to the adiabatic theorem, a state which is initially the  $n$ -th eigenstate of  $H$ ,

$$|\psi_n(0)\rangle = |n(\mathbf{R}(0))\rangle, \quad (6.2)$$

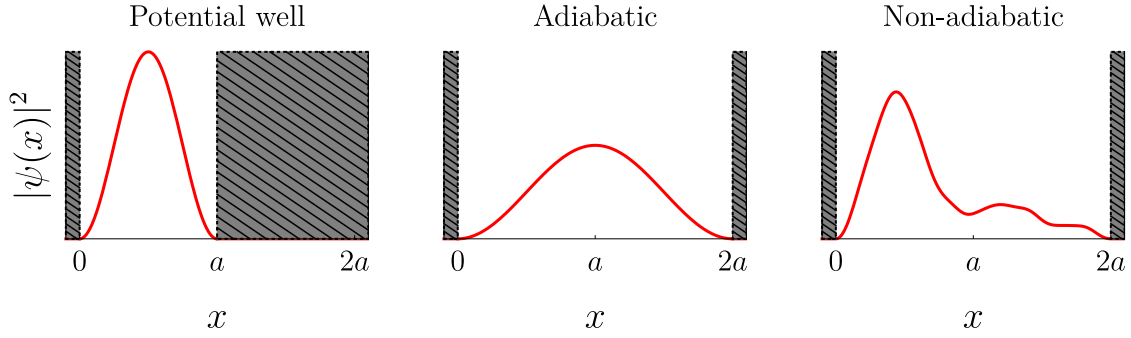
will evolve into the instantaneous eigenstate of  $H$  at a later time  $t$ , that is

$$|\psi_n(t)\rangle = c_n(t) |n(\mathbf{R}(t))\rangle, \quad (6.3)$$

given that the evolution is sufficiently slow. In this sense, the time scale is defined by the inverse of the energy gap between the energy of the  $n$ -th state and the neighboring states; a valid adiabatic regime requires an increasingly slower process the smaller the gap. In this regime, we can establish a useful relationship between the path in the parameter space and a path in the Hilbert space of the wave functions. The coefficient in Eq. (6.3) can be written as

$$c_n(t) = e^{i\gamma_n(t)} \exp\left\{ \left[ -i \int_0^t dt' \epsilon_n(t') \right] \right\}, \quad (6.4)$$

which is the main result of the adiabatic theorem [268]. The second factor in the equation above is called *the dynamical phase* [268]. Our object of interest, however, is precisely  $\gamma_n$ , which is known as the *geometric phase*. We will show why it takes this name very shortly. By plugging the equation above and Eq. (6.3) into the Schrödinger equation



**Figure 6.1:** Adiabatic and non-adiabatic dynamics of an infinite potential well. In this example we illustrate how the wave function evolves under two different scenarios for an infinite potential well by doubling its width. Consider that the w.f. is initially in the ground-state. When the width  $a$  of the well is increased *adiabatically*, the w.f. follows the instantaneous ground-state of the well, preserving its trigonometric form. In the non-adiabatic scenario where the width suddenly changes, we get a much more complicated solution, which is a combination of several eigenstates.

$i\hbar \frac{d}{dt} |\psi_n\rangle = H(\mathbf{R}(t)) |\psi_n\rangle$ , we get, after taking the inner product with  $\langle \psi(t) |$  [269]:

$$\gamma_n(t) = i \int_0^t dt' \langle n(\mathbf{R}(t')) | \frac{d}{dt'} |n(\mathbf{R}(t'))\rangle. \quad (6.5)$$

The geometrical aspect of the above equation becomes evident if we make one further modification. By using the chain rule to write

$$\frac{d}{dt} |n(\mathbf{R}(t))\rangle = \nabla_{\mathbf{R}} |n(\mathbf{R}(t))\rangle \cdot \frac{d\mathbf{R}}{dt}, \quad (6.6)$$

we may recast Eq.(6.5) as

$$\gamma_n(t) = \int_C \mathcal{A}^n(\mathbf{R}) \cdot d\mathbf{R}, \quad (6.7)$$

where

$$\mathcal{A}^n(\mathbf{R}) = i \langle n(\mathbf{R}) | \nabla_{\mathbf{R}} |n(\mathbf{R})\rangle \quad (6.8)$$

is called the *Berry connection*. This result is remarkable because it reveals a novel interpretation of  $\gamma_n$ . This phase is not an explicit function of the time, but rather, it depends only on the path  $C$  of  $\mathbf{R}$  in the parameter space, irrespective of the dynamical details of the evolution (see Fig. 6.2). For instance, it is irrelevant, from the point of view of the GP, the rate at which the path  $C$  in the parameter space is traversed, as long as the adiabatic regime is valid. Hence, this is the origin of the term *geometric phase*.



One important detail is that this phase is, in general, not gauge invariant. If we change the eigenstate by a phase  $\delta(\mathbf{R})$ , a change in the Berry connection (6.8) also incurs:

$$|n(\mathbf{R})\rangle \rightarrow e^{i\delta(\mathbf{R})} |n(\mathbf{R})\rangle \implies \mathcal{A}^n(\mathbf{R}) \rightarrow \mathcal{A}^n(\mathbf{R}) - \nabla_{\mathbf{R}}\delta(\mathbf{R}). \quad (6.9)$$

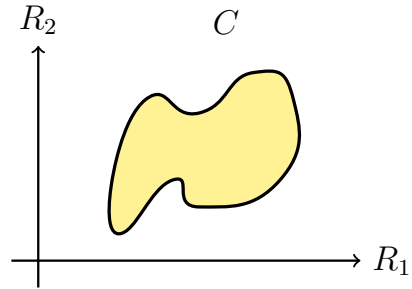
This results in a phase change which depends on the endpoints of this gauge, in other words:

$$\gamma_n \rightarrow \gamma_n + \delta(\mathbf{R}(0)) - \delta(\mathbf{R}(t)). \quad (6.10)$$

A important observation which can be made here [257] is that *cyclic* adiabatic processes in the *parameter* space, i.e., evaluations for which  $\mathbf{R}(t_{\text{end}}) = \mathbf{R}(0)$ , dispel the gauge-dependence in the geometrical phase, unambiguously defining a Berry connection for this cyclic evolution:

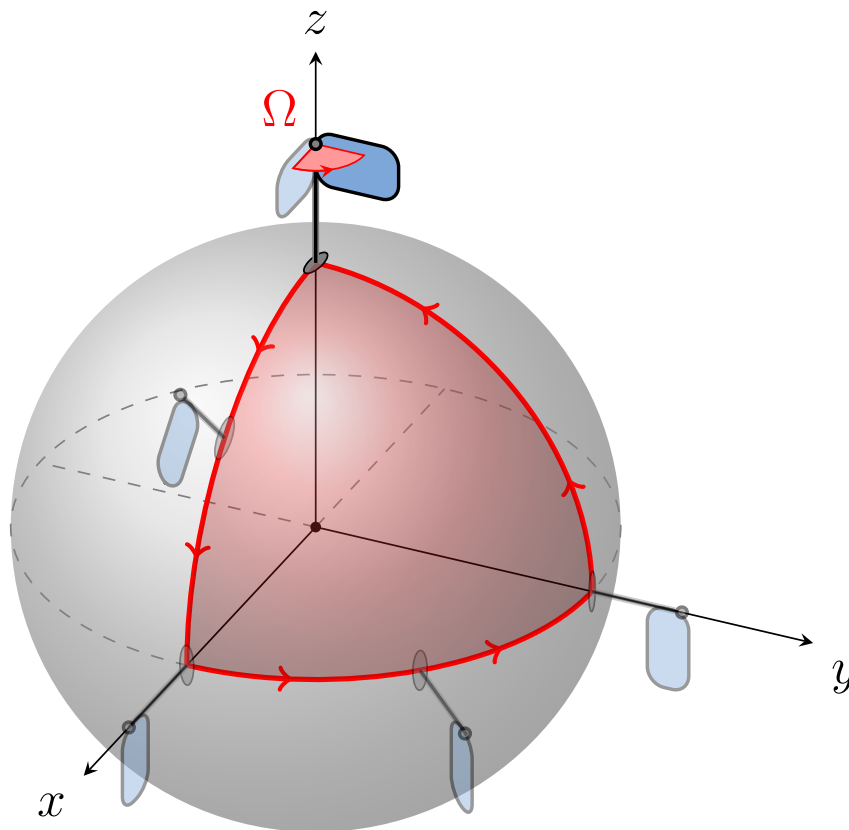
$$\gamma_n(t) = \oint \mathcal{A}^n(\mathbf{R}) \cdot d\mathbf{R}. \quad (6.11)$$

The geometrical interpretations of this result are even more far reaching. For instance, one may notice that the Berry connection, due to its gauge properties, is strongly analogous to the vector potential of a magnetic field. It is also possible to define a Berry curvature, which is a tensor constructed from the Berry connection. This allows us to use Stokes' theorem to express the geometrical phase (6.11) in terms of the area enclosed by the loop in the parameter space [269].



**Figure 6.2:** A closed path  $C$  in the parameter space.

In Fig. 6.3 we show a purely geometrical analogy of this phenomenon. For that, we can picture the following scenario: a flag is put up very close to the north pole of a sphere, pointing south. Afterwards, we carry this flag along a closed path on the sphere, always making sure that it points south. More rigorously, we say that the object is being parallel transported, i.e., we perform no local rotations [42]. Upon returning to the initial point in the north pole, the flag will acquire an azimuthal angle which is proportional to the solid angle  $\Omega$  enclosed by the path. Since no local rotations were performed throughout the process, the resulting rotation is an effect which is completely due to the curvature of the



**Figure 6.3:** Illustration of the geometrical consequences of the transport of a vector on curved surfaces.

surface.

Geometrical phases in quantum mechanics are typically discussed in terms of more abstract and higher dimensional spaces, so this description should be seen as a motivation for the problem and as a pedagogical analogy with a clear geometric visualisation of the effect, and not as a complete or rigorous description [270]. An analogous formulation for classical physics, which is more akin to this illustration, was given in terms of action-angle variables by J. H. Hannay [271], as an extension to Berry's result. There, as an example, he discusses a spinning symmetric top, in order to explain the distinction between the two contributions to the rotation: one of them being due to the angular velocity of the top, and the other one being due to the geometrical contribution of the curved surface of the earth. Nevertheless, in the next session we will see a result in quantum mechanics for a two-level system in terms of the Bloch sphere, which, in the same spirit as Fig. 6.3, can also be easily visualized.

## 6.2 Non-adiabatic geometrical phases

A natural question which arises from Berry's framework is whether a similar result holds for a non-adiabatic evolution. Built upon Berry's work and also on the Wilczek-Zee phase, which is a previous generalization of the BP for the degenerate case, it was shown that it is possible to obtain a non-adiabatic non-Abelian geometrical phase [262, 263] under certain conditions. This result will be vital for the construction of quantum gates later on, as we shall see.

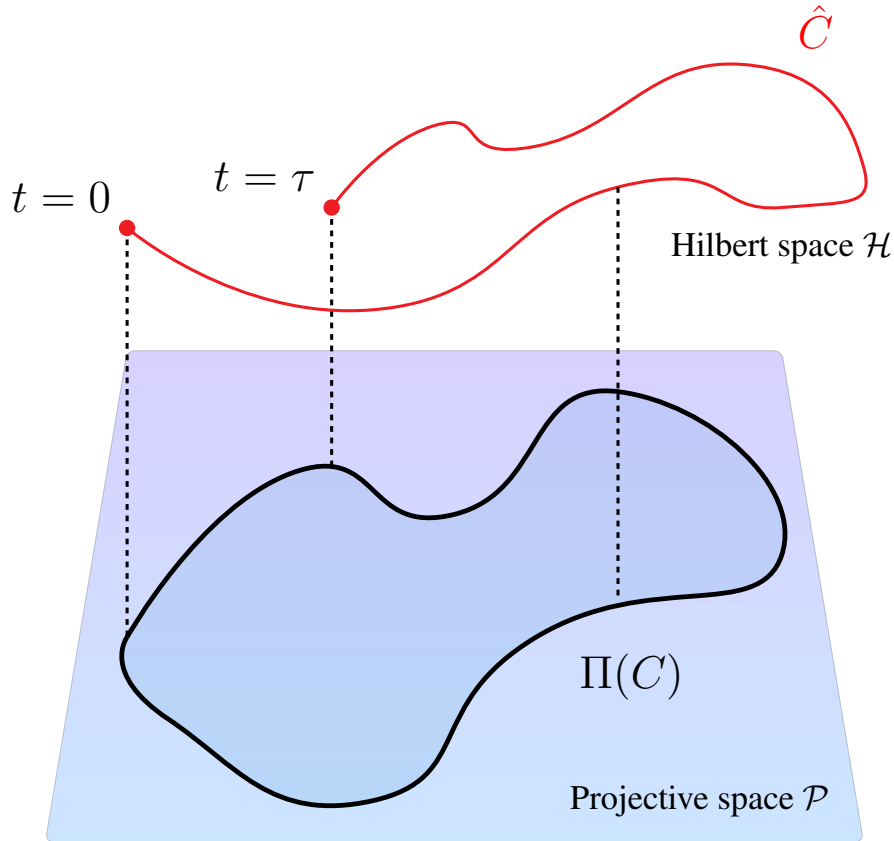
Our objective in this section is to follow the approaches in [262] and [263] to obtain a closed form expression for the desired non-adiabatic non-Abelian geometrical phases. We begin by addressing the non-adiabatic (but Abelian) phase studied in [262]. This approach is slightly more general in the sense that we consider the Hilbert space  $\mathcal{H}$  (with  $\dim \mathcal{H} = n$ ) and its projective space of rays, i.e. the set of rays  $\mathcal{P}$  (with  $\dim \mathcal{P} = n - 1$ ) defined by the projection map  $\Pi : \mathcal{H} \rightarrow \mathcal{P}$  with  $\Pi(|\psi\rangle) = \{|\psi'\rangle : |\psi'\rangle = c|\psi\rangle, \text{ s.t. } c \in \mathbb{C}\}$ , as we illustrate in Fig. 6.4. In short, what Aharonov and Anandan did was to show that it is possible to compute a geometrical phase for *all* cyclic evolutions, not only the adiabatic ones [262, 263]. Therefore, this framework encompasses other well known special cases, such as the Berry phase, periodic dynamics (such as the precession of a particle in a constant magnetic field) and so on.

Trying to analyze Berry's phase and the trajectory of the state in the parameter space for a system with a degenerate subspace is one of the examples which illustrate why this new approach can be seen as more general. Curves in the parameter space, presented in the previous section, can show a certain redundancy in this case, since for a degenerate part of the eigenspace the curve in the projective Hilbert space will simply be a point in the parameter space [263]. This dynamics nevertheless still describes a proper loop in the space of rays  $\mathcal{P}$ , which is associated to a non-trivial geometrical phase.

More concretely, let us start by considering a normalized state  $|\psi(t)\rangle \in \mathcal{H}$  which evolves according to the Schrödinger equation

$$H(t) |\psi(t)\rangle = i\hbar \frac{d}{dt} |\psi(t)\rangle \quad (6.12)$$

up to a time  $\tau$ , which satisfies  $|\psi(\tau)\rangle = e^{i\phi} |\psi(0)\rangle$ . This evolution defines an arbitrary



**Figure 6.4:** Schematic diagram of the Hilbert space and the ray space  $\mathcal{P}$ . The open path  $\hat{C}$  in the Hilbert space  $\mathcal{H}$  and the closed loop  $C = \Pi(\hat{C})$  in the projective space are both shown in the figure.

curve  $\hat{C}$  in the Hilbert space  $\mathcal{H}$ . In the ray space  $\mathcal{P}$  however this represents a closed loop  $C = \Pi(\hat{C})$ .

Now, define a state  $|\bar{\psi}(t)\rangle = e^{-if(t)} |\psi(t)\rangle$  with  $f(\tau) - f(0) = \phi$ . The evolution will be cyclic in the ray space since  $|\bar{\psi}(\tau)\rangle = |\bar{\psi}(0)\rangle$ , and the SE yields:

$$\frac{df}{dt} = \langle \bar{\psi}(t) | i \frac{d}{dt} | \bar{\psi}(t) \rangle - \frac{1}{\hbar} \langle \psi(t) | H | \psi(t) \rangle. \quad (6.13)$$

By defining

$$\beta \equiv \phi + \frac{1}{\hbar} \int_0^\tau \frac{1}{\hbar} \langle \psi(t) | H | \psi(t) \rangle dt \quad (6.14)$$

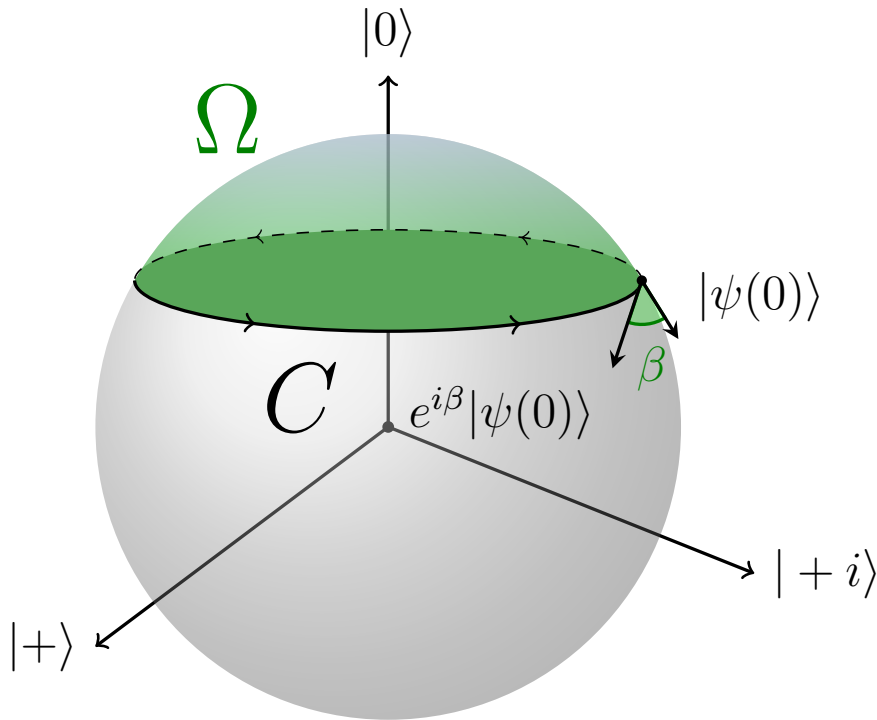
we can integrate Eq. (6.13) to obtain:

$$\beta = \int_0^\tau \langle \bar{\psi}(t) | i \frac{d}{dt} | \bar{\psi}(t) \rangle dt. \quad (6.15)$$

This identity shows that the phase  $\beta$  is purely geometrical and depends only on the curve  $C$  in  $\mathcal{P}$ :  $\beta$  is independent of both  $\phi$  and  $H$ . Moreover,  $H(t)$  can even be chosen in an alternative way such that the dynamical term in (6.14) vanishes. Note how no approximations were made here, and the expression for the phase  $\beta$  is exact: the evolution needs not to be slow neither the state  $|\psi(t)\rangle$  needs to be an eigenstate of  $H(t)$ . As a side note, we can also check that by taking  $|\bar{\psi}(t)\rangle \approx |n(t)\rangle$  we recover Berry's result (6.5), where  $|n(t)\rangle$  corresponds to an instantaneous eigenstate of  $H(t)$ .

If one is interested in measuring the geometrical phase (6.14), a few strategies are possible. One may tune  $H$  in such a way that the second term in Eq. (6.14) is zero, i.e. effectively eliminating the dynamical contribution from the phase, so that  $\phi$  becomes purely geometrical. Another possible strategy is to evolve two different states such that their dynamical part is the same, so the geometrical phase (difference) can be measured through  $\phi_1 - \phi_2 = \beta_1 - \beta_2$ .

An interesting example is the periodic evolution of a qubit subject to a constant magnetic field in the  $z$ -direction, given by  $H_B = -\Omega\sigma_z$ , where  $\Omega = \mu B$ , with  $\mu$  as the



**Figure 6.5:** Loop  $C$  on the Bloch sphere for the periodic evolution of a qubit under a constant magnetic field. The geometrical phase acquired by the qubit corresponds to half the value of the solid angle  $\Omega$  associated with the loop.

magnetic moment and  $B$  as the strength of the magnetic field. Consider an initial state of the form  $|\psi(t)\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$ . At a later time  $t$  the wave function will be  $|\psi(0)\rangle = e^{i\Omega t/\hbar} \cos(\theta/2)|0\rangle + e^{-i\Omega t/\hbar} \sin(\theta/2)|1\rangle$ . The period of the evolution is  $\tau = \pi\hbar/\Omega$ . By direct evaluation we can see that  $|\psi(\tau)\rangle = -|\psi(0)\rangle$  and  $\phi = \pi$ . Thus, by solving the integral in Eq. (6.14) we get:

$$\beta = \pi(1 - \cos\theta). \quad (6.16)$$

A noteworthy property of this example is that the geometrical phase corresponds to half the angle enclosed by the loop on the Bloch sphere [257, 262]. This example is thus clarifying because this description in terms of the Bloch state helps us visualize what is happening geometrically. However, as we have stressed before, what is really important is the loop in the ray space  $\mathcal{P}$ , and not the path on the Bloch sphere itself, as this intuitive interpretation in terms of the Bloch sphere is a very particular property of our model and does not hold for all cyclic evolutions (just as the loop in the parameter space for the Berry phase is also just a particular case of a more general result).

### 6.3 Non-adiabatic non-abelian geometrical phases

We can now extend this result to the non-Abelian case, following the steps from [263]. Consider a  $n$ -dimensional subspace  $V_n(t)$  of  $\mathcal{H}$ . We shall consider that evolution is cyclic so this subspace is the same for the endpoints of the dynamics, i.e.,  $V_n(0) = V_n(\tau)$ . Now, consider a decomposition  $H = V_n(t) \oplus V_m(t)$  of the Hilbert space into two subspaces of dimension  $n$  and  $m$ , respectively. As we will see later, this type of division can be really natural. For example, in several applications, such as one we will choose in subsequent sections, the computational space of the qubits is just a subspace of a larger Hilbert space, which may include auxiliary states into the implementation of a quantum gate. Another common application occurs when  $V_n$  is a degenerate eigenspace of  $H$ , such as it happens in the Wilczek-Zee phase [264].

Now, consider two orthonormal bases. First a basis  $\{|\bar{\psi}_a(t)\rangle, a = 1, \dots, n\}$  of  $V_n$  which satisfies the cyclic condition  $|\bar{\psi}_a(\tau)\rangle = |\bar{\psi}_a(0)\rangle$ . And second, a basis  $\{|\psi_a(t)\rangle, a =$

$1, \dots, n\}$  which follows the SE:

$$i\hbar \frac{d}{dt} |\psi_a(t)\rangle = H |\psi_a(t)\rangle. \quad (6.17)$$

Both bases initially coincide:  $|\bar{\psi}_a(0)\rangle = |\psi_a(0)\rangle$ . These two bases will be related by a unitary matrix:

$$|\psi_a(t)\rangle = \sum_{b=1}^n U_{ba}(t) |\bar{\psi}_b(t)\rangle. \quad (6.18)$$

By inserting the previous equation into Eq. (6.17) we obtain an explicit form for the unitary, given by:

$$U(t) = \mathcal{T} \exp \left\{ \left( \int_0^t i(\mathbf{A} - \mathbf{K}) dt \right) \right\}. \quad (6.19)$$

Here  $\mathcal{T}$  is the time-ordering operator,  $K_{ab} = (1/\hbar) \langle \bar{\psi}_a | H | \bar{\psi}_b \rangle$  corresponds to the dynamical part of evolution and  $A_{ab} = i \langle \bar{\psi}_a | d/dt | \bar{\psi}_b \rangle$  corresponds to the geometrical part of the evolution. To see this, note that  $A_{ab}$  does *not* depend on  $H$ , but rather, only on the structure of the Hilbert space, i.e., it can be computed entirely from the basis  $|\bar{\psi}_a(t)\rangle$ .

Investigating how  $A_{ab}$  transforms further clarifies its nature. By choosing a different basis  $|\bar{\psi}'\rangle = \Omega |\bar{\psi}\rangle$ , where  $\Omega$  is a unitary, we can see that the two matrices transform as:

$$\mathbf{A} \rightarrow i\Omega^\dagger \dot{\Omega} + \Omega^\dagger \mathbf{A} \Omega, \quad \mathbf{K} \rightarrow \Omega^\dagger \mathbf{K} \Omega, \quad (6.20)$$

showing once again that  $\mathbf{A}$  transforms as a vector potential and  $\mathcal{A} = i \langle \bar{\psi}_a | d | \bar{\psi}_b \rangle$  is a matrix-valued connection one-form [267]. In this sense, we can say that  $\mathbf{A}$  is a holonomy matrix for non-adiabatic evolutions [45].

This observation is important because it dictates the type of model and evolution we will be interested in. Namely, we are interested in loops for which the dynamical part  $\mathbf{K}$  vanishes. This guarantees that evolution is purely geometric for the reasons we have presented before. In this case, from Eq. (6.19) we can write the unitary implemented by a loop  $C$  as

$$U(C) = \mathcal{P} \exp \left\{ \left( i \oint_C \mathcal{A} \right) \right\}, \quad (6.21)$$

where  $\mathcal{P}$  is the path-ordering operator. And finally, we are interested in the non-Abelian property of these phases, in other words, we should be able to obtain two different loops  $C$

and  $C'$  for which the corresponding unitaries (6.21) do not commute. One motivation for this is in universal quantum computing, which obviously requires non-commuting gates. As we will see in the next section, the  $\Lambda$ -type systems are a promising platform which satisfy both of these requirements.



# Chapter 7

## Non-adiabatic holonomic quantum computing

The use of geometrical phases in quantum computing, while robust against certain types of noise and error [272–274], still suffers from decoherence and other open quantum system effects [46]. This calls for strategies that are able to circumvent this type of problem [275]. Non-adiabatic holonomic quantum computing (NHQC) [45] has shown promise in performing this task [276]. The considerable speed-up in the operation time makes the system much more robust against decoherence. This configuration has been used before in a couple of different experimental systems, such as nitrogen-vacancy centers in diamond [277, 278] and superconducting qubits [279]. Several extensions or alternative formulations of the original proposal have already been investigated, such as implementations which shorten the original protocol by employing single-loop holonomies [280–283] and a generalization to discrete holonomies [284]. In this section we review the original implementation of NHQC using  $\Lambda$ -type systems [45] and its robustness against noise.

### 7.1 The $\Lambda$ -type system

Here we follow the original setup for NHQC in the  $\Lambda$ -type system from Ref. [45]. In this system, we couple two states  $|0\rangle$  and  $|1\rangle$  to an auxiliary excited state  $|e\rangle$ , while  $|0\rangle$  and  $|1\rangle$ , which span the qubit space, are uncoupled between themselves. Thus, the system

acquires a  $\Lambda$ -like structure, depicted in Fig. (7.1).

The starting point to describe the model is the Hamiltonian:

$$H(t) = H_0 + \boldsymbol{\mu} \cdot \mathbf{E}(t), \quad (7.1)$$

where  $H_0 = -f_{e0} |0\rangle \langle 0| - f_{e1} |1\rangle \langle 1|$  is the bare Hamiltonian and

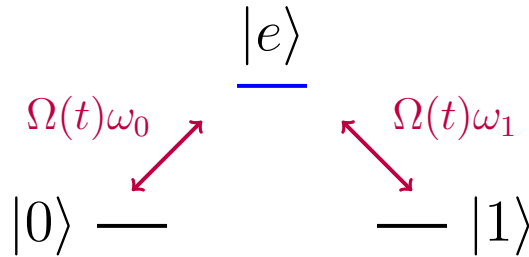
$$\mathbf{E}(t) = g_0(t) \cos(\nu_0 t) \boldsymbol{\epsilon}_0 + g_1(t) \cos(\nu_1 t) \boldsymbol{\epsilon}_1, \quad (7.2)$$

is the applied oscillating electric pulse. Here,  $g_j(t)$  and  $\nu_j$  (with  $j = 0, 1$ ) are the pulse envelope and the oscillation frequency, respectively. Additionally,  $\boldsymbol{\mu}$  is the magnetic dipole moment operator and  $\boldsymbol{\epsilon}$  is the polarization. From hereafter, we also take  $\hbar = 1$ . By moving to the interaction picture Hamiltonian  $H_I(t) = e^{-iH_0 t} H(t) e^{iH_0 t}$  we obtain:

$$\begin{aligned} H_I(t) = & \Omega_0(t) (e^{-i(f_{e0} + \nu_0)t} + e^{-i(f_{e0} - \nu_0)t}) |e\rangle \langle 0| \\ & + \Omega_1(t) (e^{-i(f_{e1} + \nu_1)t} + e^{-i(f_{e1} - \nu_1)t}) |e\rangle \langle 1| + \text{h.c.} \end{aligned} \quad (7.3)$$

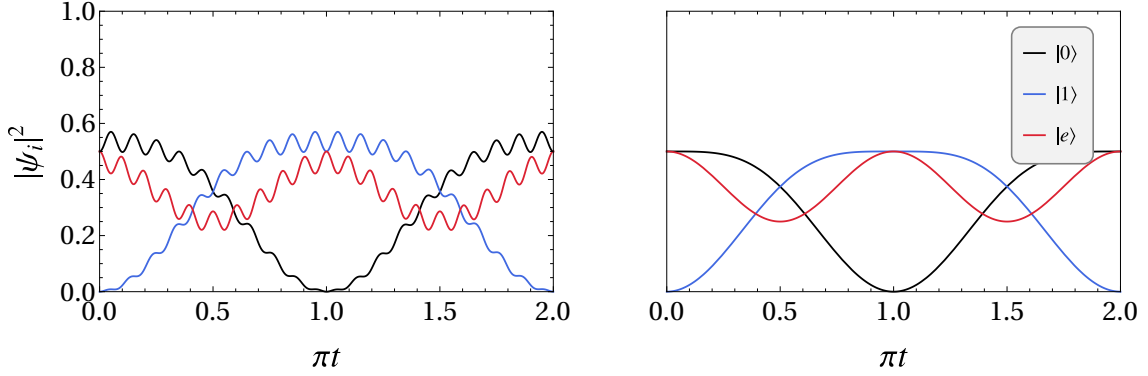
where  $\Omega_j = \langle e | \boldsymbol{\mu} \cdot \boldsymbol{\epsilon} | j \rangle g_j(t) / 2$  are transition frequencies which depend only on the parameters of the applied field. In a final step, we tune the frequencies  $\nu_j$  so they are resonant with the bare transition frequencies  $f_{ej}$ , i.e.  $\nu_j = f_{ej}$ . By doing so one finds

$$H_I(t) = \Omega_0(t) (1 + e^{-2if_{e0}t}) |e\rangle \langle 0| + \Omega_1(t) (1 + e^{-2if_{e1}t}) |e\rangle \langle 1| + \text{h.c.}, \quad (7.4)$$



**Figure 7.1:** Basic setup for the  $\Lambda$ -type system.

Note how the bare Hamiltonian introduces counter-rotating terms of the type  $(1 + e^{-2if_{ej}t})$ . These terms are a source of nonideality and the focus of our research, therefore, in the next chapter we will be concerned with how to handle these terms and how they



**Figure 7.2:** Effects of the counter-rotating terms in the dynamics of the  $\Lambda$ -system. We compare the solutions of Eq. (7.4). We plot the populations (left) in the presence of counter-rotating terms, where  $f_{e0} = f_{e1} = 10$  and  $\Omega_0(t) = \Omega_1(t) = 1/\sqrt{2}$ , and (right) in the RWA regime. In both cases we take the initial state to be  $|\psi_0\rangle = (|0\rangle + |e\rangle)/\sqrt{2}$ . As the frequencies  $f_{ei}$  increase, the smaller the ripples on the non-RWA solution get in comparison to overall period of the dynamics.

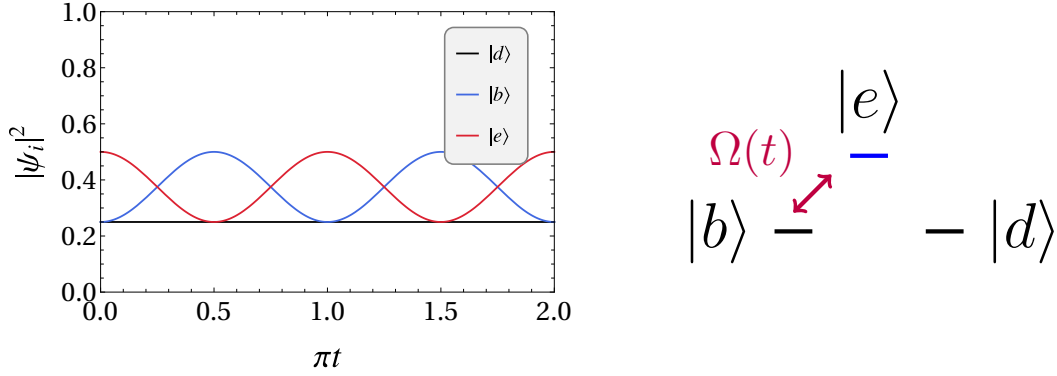
affect the quantum gates. However, for now we will be interested in the regime of the rotating-wave approximation (RWA). The argument here is that whenever  $f_{ei}$  is much bigger the other typical frequencies of the system,  $e^{\pm 2if_{ej}t}$  become rapidly oscillating terms which average out to zero.<sup>1</sup> In this scenario the counter-rotating terms and their effects are negligible. This allows us to describe the system and the implementation of NHQC in an ideal setting. This can be seen in more detail in Fig. (7.2). The counter-rotating terms introduces "ripples" in the solution. Meanwhile, when we average out these terms, we get the RWA Hamiltonian

$$H_I^{\text{RWA}}(t) = \Omega_0(t) |e\rangle \langle 0| + \Omega_1(t) |e\rangle \langle 1| + \text{h.c.}, \quad (7.5)$$

which leads to a smooth solution, seen on Fig. (7.2)(b). Thus, it is possible to smooth out the dynamics of the system by increasing the energy gap of the bare states, approaching the RWA regime.

To further comprehend the dynamics of the system, we can perform a very illustrative analysis on the Hamiltonian in Eq. (7.5) in terms of its eigenstates. Here we introduce a frequency envelope  $\Omega(t)$  and the relative amplitudes  $\omega_0$  and  $\omega_1$ , which allows us rewrite the transition frequencies as  $\Omega_0(t) = \omega_0\Omega(t)$  and  $\Omega_1(t) = \omega_1\Omega(t)$ , where we assume that

<sup>1</sup>One way to justify this claim is by employing the Magnus expansion [285, 286] for the unitary evolution. We get leading correction terms of order  $O(\Omega/f_{ej})$ , which are negligible whenever  $\Omega/f_{ej} \ll 1$ .



**Figure 7.3:** Evolution of the amplitudes of the state  $\psi_0$  defined in Fig. (7.2) in terms of the bright and dark states (left). We also depict the  $\Lambda$ -type-system in this case (right).

$|\omega_0|^2 + |\omega_1|^2 = 1$ . We then define the dark state as  $|d\rangle = -\omega_1 |0\rangle + \omega_0 |1\rangle$  and the bright state as  $|b\rangle = \omega_0^* |0\rangle + \omega_1^* |1\rangle$ . By doing so, we can rewrite Eq. (7.5) as:

$$H_I^{\text{RWA}}(t) = \Omega(t)(|e\rangle \langle b| + |b\rangle \langle e|). \quad (7.6)$$

Thus, we can see the convenience of this change of basis; the state  $|d\rangle$  is decoupled from the evolution, and we effectively get a two-level system with oscillations between the bright state  $|b\rangle$  and the auxiliary state  $|e\rangle$ . In Fig. (7.3) we depict the Rabi oscillations between these two states. Note how the amplitude of  $|d\rangle$  remains unchanged throughout the evolution.

## 7.2 Implementing single-qubit gates

Now we can finally make use of the formalism from the previous chapter to show how the geometrical properties of the  $\Lambda$ -type system can be used in order to implement universal single-qubit gates. In our implementation we are interested in the qubit subspace  $M(0) = \text{span}\{|0\rangle, |1\rangle\}$ , while the state  $|e\rangle$  plays the important auxiliary role. This space evolves into  $M(t)$ , which is spanned by:

$$|\psi_k(t)\rangle = \exp\left(-i \int_0^t H_I^{\text{RWA}}(t') dt'\right) |k\rangle = \mathcal{U}(t, 0) |k\rangle, \quad (7.7)$$

where  $k = 0, 1$  and  $\mathcal{U}(t, 0)$  is the time-evolution operator. Notice that in the case of the RWA Hamiltonian (7.5) it is not necessary to include a time-ordering operator. It is also possible to get some insight into the dynamics by writing these expressions in terms of the bright-dark basis. The unitary matrix becomes

$$\mathcal{U}_{bd}(t, 0) = |d\rangle\langle d| + \cos(\Phi)(|b\rangle\langle b| + |e\rangle\langle e|) - i \sin(\Phi)(|e\rangle\langle b| + |b\rangle\langle e|) \quad (7.8)$$

where

$$\Phi = \int_0^t \Omega(t') dt' \quad (7.9)$$

is the area enclosed by the pulse. Consequently, the bright and dark states evolve as [281]:

$$\begin{aligned} |\psi_d(\Phi)\rangle &= |d\rangle, \\ |\psi_b(\Phi)\rangle &= \cos(\Phi) |b\rangle - i \sin(\Phi) |e\rangle. \end{aligned} \quad (7.10)$$

In particular, we are interested in pulses which satisfy

$$\Phi = \Phi_C := \pi. \quad (7.11)$$

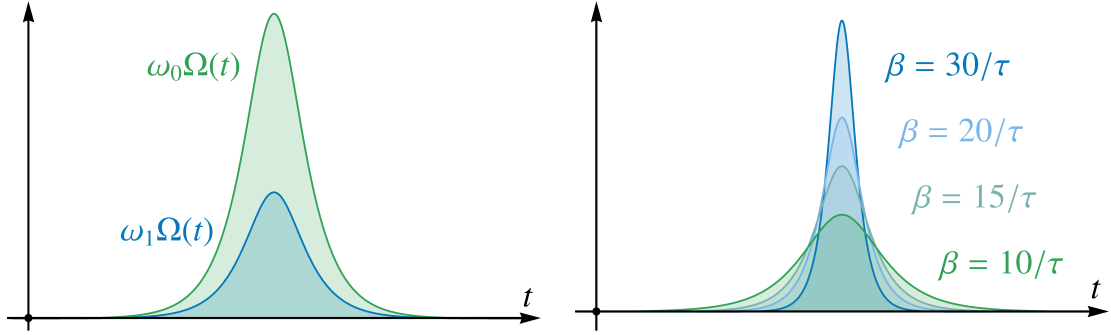
In this case, we can see that the states evolve as

$$\begin{aligned} |\psi_d(\Phi_C)\rangle &= |d\rangle, \\ |\psi_b(\Phi_C)\rangle &= -|b\rangle, \\ |\psi_e(\Phi_C)\rangle &= -|e\rangle. \end{aligned} \quad (7.12)$$

The effect of this geometrical evolution whenever  $\Phi = \pi$  is to implement a holonomy matrix  $Z_{bd}$  which acts by flipping the sign of  $|b\rangle$  and  $|e\rangle$  in the bright-dark basis. In other words:

$$\mathcal{U}_{bd}(C) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (7.13)$$

An explicit calculation shows that this matrix, in the computational basis, becomes, after properly projecting it back onto the qubit space:



**Figure 7.4:** (a) A representation of the hyperbolic secant pulse from Eq. (7.16). (b) Representation of the pulse for different values of  $\beta$ . Note how  $1/\beta$  can be seen as the length of the pulse. Here  $\tau$  is an arbitrary time scale.

$$U(C) = \mathcal{U}_{bd}(C)\mathbb{P} = \begin{pmatrix} \cos \theta & e^{-i\phi} \sin \theta \\ e^{i\phi} \sin \theta & -\cos \theta \end{pmatrix} = \mathbf{n} \cdot \boldsymbol{\sigma} \quad (7.14)$$

where  $\mathbf{n} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$  and  $\mathbb{P} = |0\rangle\langle 0| + |1\rangle\langle 1|$ .<sup>2</sup> Here, we have parametrized the frequencies  $\omega_0$  and  $\omega_1$  as  $\omega_0 = \sin(\theta/2)e^{i\phi}$  and  $\omega_1 = -\cos(\theta/2)$ . Besides the convenient representation for the unitary, this also guarantees that  $|\omega_0|^2 + |\omega_1|^2 = 1$ . Thus, this process implements a  $\pi$  rotation around  $\mathbf{n}$  on the Bloch sphere.<sup>3</sup> This unitary, however, is not universal; Eq. (7.14) only implements traceless operations. By employing a second loop  $C_m$  we can implement the universal gate:

$$U(C) = U(C_m)U(C_n) = \mathbf{n} \cdot \mathbf{m} - i\boldsymbol{\sigma} \cdot (\mathbf{n} \times \mathbf{m}). \quad (7.15)$$

This transformation has a clear geometrical meaning as well. The universal gate  $U(C)$  above corresponds to a rotation in the plane spanned by  $\mathbf{n}$  and  $\mathbf{m}$  by an angle of  $2 \cos^{-1}(\mathbf{n} \cdot \mathbf{m})$ . Therefore, any single-qubit gate can be obtained by properly choosing the appropriate pulses, which will determine  $\mathbf{n}$  and  $\mathbf{m}$ .

Under these considerations, we are going to choose a very convenient type of pulse, namely, pulses which have the shape of a hyperbolic secant:

$$\Omega(t) = \beta \operatorname{sech}(\beta t). \quad (7.16)$$

<sup>2</sup>This is a slight abuse of notation, since  $U(C)$  is still a  $3 \times 3$  matrix and proper projection operators should be rectangular matrices for our intended purpose. Nevertheless,  $U(C)|e\rangle = 0$  so, unless explicitly stated, we write all the unitaries and relevant operators in the subspace of the qubit states for simplicity.

<sup>3</sup>To see this, note that Eq. (7.14) can be alternatively written as  $U(C) = ie^{-i\frac{1}{2}\pi(\mathbf{n} \cdot \boldsymbol{\sigma})}$  [45, 281].

This parametrization is convenient because it guarantees that condition (7.11) is satisfied for whichever  $\beta$  is chosen. This choice gives us freedom of using different pulse lengths just by changing the parameter  $\beta$ . As shown in Fig. (7.4),  $1/\beta$  can be interpreted as the length of the pulse: pulses are sharper for large values of  $\beta$  and wider when  $\beta$  is small. Moreover, by appropriately choosing the angles  $\theta$  and  $\phi$  we can determine the vector  $\mathbf{n}$  (and by choosing a different set of angles for the second pulse we determine  $\mathbf{m}$ ). Pulses of this type are experimentally friendly and exhibit desirable properties in quantum optics and quantum control [287–290]. Of course, other pulses can also be used, such as Gaussian and square pulses [280, 291].

As a practical example, we consider the Hadamard gate  $H$  and the  $S$  gate, which are defined as

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad (7.17)$$

respectively. For the Hadamard gate we need only a single pulse, choosing  $\theta = \pi/4$  and  $\phi = 0$ . For the  $S$ -gate the protocol is slightly more complicated, since we need the two pulses. In this case, we can take  $\theta_0 = \pi/2$  and  $\phi_0 = \pi/2$  for the first pulse and  $\theta_1 = 0$  and  $\phi_1 = \pi/4$  for the second pulse.

The drawback of gates which require two loops, such as the  $S$ -gate, is that the exposure time to decoherence effects is longer. We will see this explicitly in the next chapter. Protocols which implement non-adiabatic quantum gates for single-loops which can mitigate this effect can be found in [280] and [281].

On a closing note, we comment on the geometrical aspect of the model, checking for the condition discussed in Sec (6.3) for a fully geometrical evolution. We can see, from Eqs. (7.6) and (7.10), that the dynamical contribution is  $\langle \psi_k(t) | H | \psi_l(t) \rangle = 0$  for  $k, l = b, d$ . Since  $|d\rangle$  decouples from the Hamiltonian in Eq. (7.6), and  $|\psi_d(t)\rangle = |d\rangle$ , it is easy to see that  $\langle \psi_k(t) | H | \psi_l(t) \rangle$  vanishes for any inner product involving the dark state. Finally, we can show that  $\langle \psi_b(t) | H | \psi_b(t) \rangle = 0$  by direct calculation, since

$$\langle \psi_b(t) | H | \psi_b(t) \rangle = \Omega(t)(\cos \Phi \langle e | + i \sin \Phi \langle b |)(|e\rangle\langle b| + |b\rangle\langle e|)(\cos \Phi |e\rangle - i \sin \Phi |b\rangle) = 0.$$

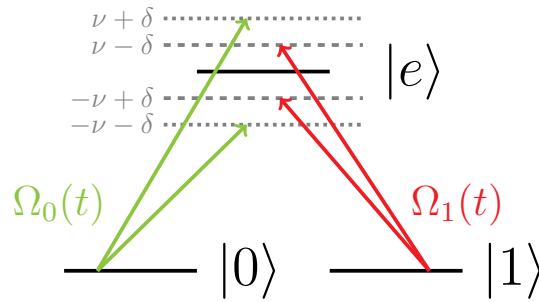
This proves the geometric nature of the gate: the dynamical contributions in  $K_{kl} =$

$\langle \psi_k(t) | H | \psi_l(t) \rangle$  in Eq. (6.19) vanish, and we are left with the geometrical contributions in Eq. (6.21) only [45].

### 7.3 Implementing two-qubit gates

In this section we once more follow the original proposal in [45], which also includes a protocol based on the Sørensen–Mølmer scheme [292] for implementing two-qubit gates (an adiabatic implementation is presented in [41]. See also Ref. [293] for a multi-qubit generalization of the NHQC scheme). The two-qubit gate can be constructed with two ions in the same three-level  $\Lambda$  configuration. The transition  $0 \rightarrow e$  ( $1 \rightarrow e$ ) is driven by a laser with detuning  $\nu \pm \delta$  [ $\mp(\nu \pm \delta)$ ], where  $\nu$  is the phonon frequency of a vibrational mode [41, 294] and  $\delta$  is an additional detuning [Fig. (7.5)]. Moreover, two extra conditions which the setup should satisfy are: the Lamb-Dicke criterion  $\eta \ll 1$ , where  $\eta$  is the Lamb-Dicke parameter, and  $|\Omega_i(t)| < \nu$  in order to suppress the off-resonant couplings [292]. The parameter  $\eta$  is related to the zero-point spread of the ion [295] and the coupling strength between its motional states and internal degrees of freedom [296]. The effective Hamiltonian describing this interaction assumes the form (see Ref. [297] for a derivation):

$$H^{(2)} = \frac{\eta^2}{\delta} (|\Omega_0(t)|^2 \sigma_0(\phi, t) \otimes \sigma_0(\phi, t) - |\Omega_1(t)|^2 \sigma_1(-\phi, t) \otimes \sigma_1(-\phi, t)), \quad (7.18)$$



**Figure 7.5:** Setup for the ions in the two-qubit gate. The green (red) arrows and the dotted (dashed) lines correspond to the  $\nu \pm \delta$  ( $\nu \mp \delta$ ) detuning of the  $0 \rightarrow e$  ( $1 \rightarrow e$ ) transition. Both ions have the same configuration.



where

$$\begin{aligned}\sigma_0(\phi, t) &= e^{i\phi/4}(1 + e^{-2if_{e0}t}) |e\rangle \langle 0| + \text{h.c.}, \\ \sigma_1(-\phi, t) &= e^{-i\phi/4}(1 + e^{-2if_{e1}t}) |e\rangle \langle 1| + \text{h.c.}\end{aligned}\quad (7.19)$$

After eliminating off-resonant couplings of the singly excited states  $|0e\rangle$ ,  $|e0\rangle$ ,  $|1e\rangle$  and  $|e1\rangle$  and performing the RWA the Hamiltonian reads:

$$H^{(2),\text{RWA}}(t) = \frac{\eta}{\delta^2} \sqrt{|\Omega_0(t)|^4 + |\Omega_1(t)|^4} \left( H_0^{(2),\text{RWA}}(t) + H_1^{(2),\text{RWA}}(t) \right), \quad (7.20)$$

with

$$H_0^{(2),\text{RWA}}(t) = \sin \frac{\theta}{2} e^{i\phi/2} |ee\rangle \langle 00| - \cos \frac{\theta}{2} e^{-i\phi/2} |ee\rangle \langle 11| + \text{h.c.} \quad (7.21)$$

and

$$H_1^{(2),\text{RWA}}(t) = \sin \frac{\theta}{2} |e0\rangle \langle 0e| - \cos \frac{\theta}{2} |e1\rangle \langle 1e| + \text{h.c.} \quad (7.22)$$

The single and two-qubit gates share similar criteria: the phase  $\phi$  should be kept constant throughout the evolution, while the frequencies satisfy  $|\Omega_0(t)|^2/|\Omega_1(t)|^2 = \tan(\theta/2)$ . Additionally, we should once again respect the criterion (7.11) for the pulse area:

$$\frac{\eta^2}{\delta} \int_0^\tau \sqrt{|\Omega_0(t)|^4 + |\Omega_1(t)|^4} dt = \pi. \quad (7.23)$$

Moreover, one extra observation is in order; since  $H_0^{(2)}(t)$  and  $H_1^{(2)}(t)$  commute, it is possible to decompose the evolution of the total Hamiltonian (7.20) as

$$\exp \left\{ \left( -i \int_0^\tau H^{(2),\text{RWA}}(t) dt \right) \right\} = \exp \left\{ \left( -i\pi H_0^{(2),\text{RWA}} \right) \right\} \exp \left\{ \left( -i\pi H_1^{(2),\text{RWA}} \right) \right\}.$$

The Hamiltonian  $H_1^{(2)}(t)$  however acts trivially on the relevant computational subspace  $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$  [45], making  $H_0^{(2)}(t)$  the relevant term in the evolution. Due to this fact, by redefining some parameters in Eq. (7.20),  $H^{(2)}(t)$  can effectively be seen as a  $\Lambda$ -type-like Hamiltonian [293]

$$H_{\text{eff}}^{(2),\text{RWA}} = \Omega^{(2)}(t) (\omega_0^{(2)} |00\rangle \langle ee| + \omega_1^{(2)} |11\rangle \langle ee|) + \text{h.c.} \quad (7.24)$$

with  $\Omega^{(2)}(t) = \sqrt{|\Omega_0(t)|^4 + |\Omega_1(t)|^4}$ ,  $\omega_1^{(2)} = e^{i\frac{\phi}{2}} \sin(\theta/2)$  and  $\omega_0^{(2)} = -e^{-i\frac{\phi}{2}} \cos(\theta/2)$ . Thus, by analogy to the single qubit gate, as done in Eqs. (7.7)-(7.14), we get the corresponding unitary:

$$U^{(2)}(C_n) = \cos\theta |00\rangle\langle 00| + e^{-i\phi} \sin\theta |00\rangle\langle 11| + e^{i\phi} \sin\theta |11\rangle\langle 00| - \cos\theta |11\rangle\langle 11| + |01\rangle\langle 10| + |10\rangle\langle 10|. \quad (7.25)$$

This unitary acts just like a single qubit gate in the space  $\{|00\rangle, |11\rangle\}$  and it leaves the components in the space  $\{|01\rangle, |10\rangle\}$  invariant.

By choosing  $\theta = 0$  we construct a CZ gate

$$U_{CZ}^{(2)} = |00\rangle\langle 00| + |01\rangle\langle 10| + |10\rangle\langle 10| - |11\rangle\langle 11|, \quad (7.26)$$

which is an entangling gate and can form a universal set together with other single-qubit gates [298].

Finally, if we take the counter-rotating terms into account, Eq. (7.21) and Eq. (7.22) become

$$H_0^{(2)}(t) = (1 + e^{-2if_{e0}t})^2 \sin\frac{\theta}{2} e^{i\phi/2} |ee\rangle\langle 00| - (1 + e^{-2if_{e1}t})^2 \cos\frac{\theta}{2} e^{-i\phi/2} |ee\rangle\langle 11| + \text{h.c.} \quad (7.27)$$

and

$$H_1^{(2)}(t) = 4 \cos^2(f_{e0}t) \sin\frac{\theta}{2} |e0\rangle\langle 0e| - 4 \cos^2(f_{e1}t) \cos\frac{\theta}{2} |e1\rangle\langle 1e| + \text{h.c.}, \quad (7.28)$$

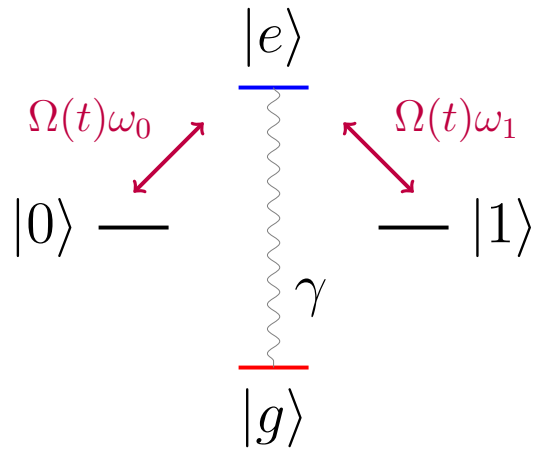
respectively.

## 7.4 Gate robustness

Throughout this work, we have mentioned the robustness of the geometric quantum gates several times. Our objective in this section is to make this claim more precise, explicitly showing, by numerical means, how the NHQC implementation in the  $\Lambda$ -system is sufficiently robust against open quantum systems effects. We discuss the impact of amplitude

damping in the system, as it was originally done in Ref. [45]. The interested reader can find an investigation on the robustness of the model against a few other error sources in Ref. [276]. For instance, it was shown by the authors that contributions due to errors in the pulse envelope and relative pulse amplitude are all of *second* order and, more strikingly, independent of the pulse duration. We omit these results here, since we will not consider this type of error in our analysis in the next chapter.

We consider the following model: in the  $\Lambda$ -type system, the excited state  $|e\rangle$  is an unstable state which undergoes dissipation, while the computational states  $|0\rangle$  and  $|1\rangle$  can be regarded as stable ground states. Thus, a physical description of the model will look like Fig. (7.6). In our formulation we assume that the excited state decays to an auxiliary ground state  $|g\rangle$  which is *not* coupled to any other states through any type of unitary dynamics. This means that whenever the ground state is populated, these excitations are "lost" from the point of view of the computational subspace. This will result in a non-ideal mixed state at the end of the computation, which



**Figure 7.6:** Basic setup for the  $\Lambda$ -type system when an open quantum system approach is considered. The excited state decays with a rate  $\gamma$ . There is no coupling or decay with the computational subspace itself. This means that dissipation effects only occur while the pulse is being applied and the excited state is populated.

will decrease the overall fidelity of the gate. Our purpose is precisely to investigate how to mitigate this effect.<sup>4</sup>

We will model the decay with an amplitude-damping jump operator given by  $L = |g\rangle\langle e|$ . Under a purely dissipative evolution this term will push the excitations in  $|e\rangle$  towards the ground state  $|g\rangle$  (and, of course, this process will also damp the coherences of the system, together with the populations). Initially considering the RWA case, the

<sup>4</sup>Of course, other formulations are also possible. For instance, in transmon qubits the state  $|0\rangle$  is already encoded into the ground state and there is no further low-lying level  $|g\rangle$ . Thus, in [279] decoherence is modeled by considering the dephasing dissipators  $D_\rho(|1\rangle\langle 1| - |0\rangle\langle 0|)$  and  $D_\rho(|e\rangle\langle e| - |0\rangle\langle 0|)$ .

dynamics of the system will then be given by

$$\frac{d\rho}{dt} = i[\rho, H_I^{\text{RWA}}(t)] + \gamma D_\rho(|g\rangle\langle e|), \quad (7.29)$$

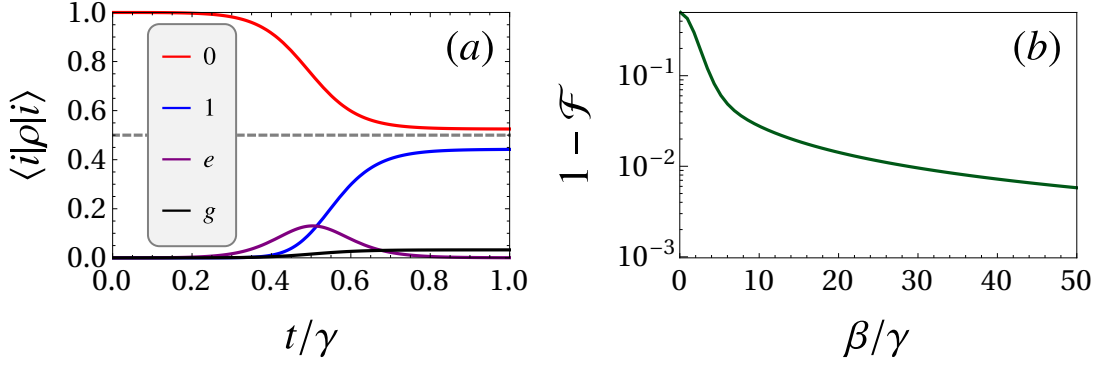
where  $H_I^{\text{RWA}}(t)$  is the RWA Hamiltonian in Eq. (7.5). Robustness against this type of noise has been shown in [45]. Meanwhile, the effect of dephasing is qualitatively similar and has also been investigated in [276]. The takeaway message in this case was that the inverse pulse length should be much larger than the typical coupling strength  $\gamma$ , i.e. we should have  $\beta \gg \gamma$ . By increasing  $\beta$  the fidelity also monotonically increases, approaching unity. This is precisely one of the advantages of this non-adiabatic scheme: by removing constraints on the operation time one can use shorter pulses, diminishing the effects of decoherence.

As a sanity check, we can, by turning off the unitary part in Eq. (7.29), also understand how the dissipative part acts on the system. For a pure initial state of the type  $|\psi\rangle = c_0|0\rangle + c_1|1\rangle + c_e|e\rangle$ , the density matrix evolves as:

$$\rho(t) = \begin{pmatrix} |c_0|^2 & c_0c_1^* & e^{-\gamma t/2}c_0c_e^* & 0 \\ c_1c_0^* & |c_1|^2 & e^{-\gamma t/2}c_1c_e^* & 0 \\ e^{-\gamma t/2}c_e c_0^* & e^{-\gamma t/2}c_e c_1^* & e^{-\gamma t}|c_e|^2 & 0 \\ 0 & 0 & 0 & |c_e|^2(1 - e^{-\gamma t}) \end{pmatrix}. \quad (7.30)$$

We can see that the process does not interfere with the computational subspace, but it dampens the population of the excited state and destroys the coherences. In the steady state limit  $\gamma\tau \rightarrow \infty$ , all the population from the excited state is transferred to the ground state. Hence, what happens when we turn off the unitary interaction is that the dissipative process occurs only during the pulse application, since there is no coupling between the computational subspace and the ground state. Thus, by shortening the pulse we also shorten the time during which the excited state is occupied and we minimize the errors due to dissipation. This is of course consistent with what we expected when choosing the jump operators to model the dissipation in this setup, so this result should not be seen as surprising.

We can now turn on the unitary part of interaction and try to quantify how well this



**Figure 7.7:** We plot (a) the populations of the density matrix for the implementation of the Hadamard gate and (b) the infidelity  $1 - \mathcal{F}$  of the gate as a function of the pulse duration  $\beta$ . In both cases the input state is  $|0\rangle$ . The gray dashed line highlights the point  $\langle i|\rho|i\rangle = 0.5$ .

protocol performs. For that, we will take the initial state  $|\psi_0\rangle = |0\rangle$ . Our gate of choice will be the Hadamard gate, whose implementation was discussed in Sec. (7.2). Ideally, the application of the Hadamard gate upon  $|0\rangle$  should yield  $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ .

The figure of merit for gate performance will be the fidelity, defined for two arbitrary mixed states  $\rho$  and  $\sigma$  as [299, 300]

$$F(\rho, \sigma) = \left( \text{tr} \left\{ \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right\} \right)^2. \quad (7.31)$$

The fidelity essentially quantifies how similar two states are, being 0 for orthogonal states and 1 when both states are equal. Since we want to compare our output state with an expected result  $|\psi\rangle$  which is always pure, the fidelity assumes a much simpler form. In this case, it will be

$$\mathcal{F} = F(\rho, |\psi\rangle\langle\psi|) = \langle\psi|\rho|\psi\rangle, \quad (7.32)$$

where  $|\psi\rangle = U(C)|\psi_0\rangle$  is the ideal output upon the application of the ideal quantum gate  $U(C)$  and  $\rho$  is the density matrix obtained by evolving the system under Eq. (7.29) for the whole duration of the pulse. We may also use the infidelity  $1 - \mathcal{F}$  whenever it is convenient.

Basic results are shown in Fig. (7.7). On panel (a) we can see that the application of the pulse temporarily populates the excited state  $|e\rangle$ , transferring excitations from  $|0\rangle$  to  $|1\rangle$ . Meanwhile, the ground state is also slightly populated due to these dissipative effects,

and some of these excitations are "lost", decaying to the ground state  $|g\rangle$ . And while we did not plot the coherences here, they are also affected by this dissipative process. At the end of the protocol we get a mixed state which deviates from the pure target state. We can quantify this effect by investigating how the end result changes depending on the (inverse) duration  $\beta$  of the pulse. In Fig. (7.7)(b) we plot these results, showing the infidelity  $1 - \mathcal{F}$  as a function of the inverse pulse length relative to the dissipation strength. We can see that it monotonically decreases as we increase  $\beta/\gamma$ . Thus, as long as the RWA is valid, a decrease in the pulse length results in a monotonically increasing fidelity, approaching unity asymptotically.

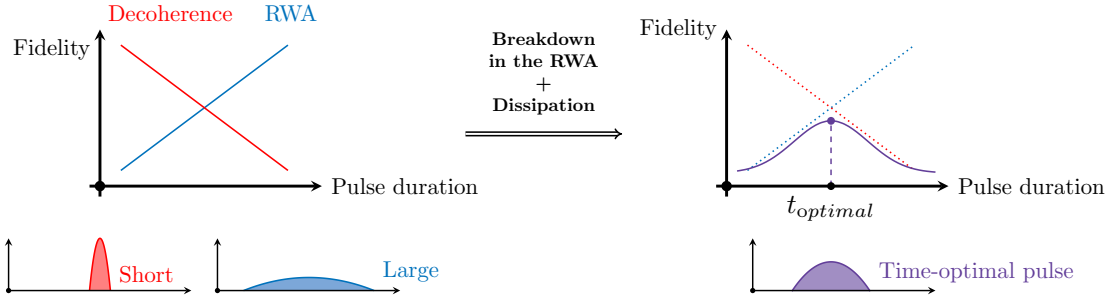
# Chapter 8

## Time-optimality

In the previous section we considered an ideal implementation of the  $\Lambda$ -type system where the counter-rotating frequencies from Eq. (7.4) are rapidly oscillating. This yields the RWA Hamiltonian from Eq. (7.5) whose robustness under amplitude noise we analyzed in Sec. 7.4. In this ideal scenario the fidelity can be arbitrarily improved simply by shortening the pulse duration as much as necessary. However, we are interested in what happens in a regime where the RWA is not valid.

An extension of the original proposal in Ref. [45] for the non-RWA case has been made in the past [291], albeit for the dissipationless model. There, the authors show that the RWA starts to break down at very small operation times due to the comparable time scale between the pulse and counter-rotating frequencies. Unfortunately, this constrains how fast these gates can operate in a scenario where one is trying to mitigate decoherence, making it impossible to employ increasingly shorter pulses beyond a certain limit. Instead, it is also important to account for the validity of the RWA; while it is necessary to use sufficiently short pulses in order to avoid dissipative losses, the pulses should also be long enough so as not to introduce significant dynamical corrections due to the counter-rotating terms. Our contribution here is to consider both effects at once, investigating the interplay between the counter-rotating oscillations and the dissipative phenomena in the system. We show how the combination of the breakdown in the RWA and the open quantum system effects introduce a trade-off in the pulse length, limiting the maximum achievable fidelity of the gate. This is illustrated in Fig. 8.1.

In this chapter we perform a series of numerical investigations in order to calculate the



**Figure 8.1:** A diagram illustrating the competition between decoherence and the counter-rotating contributions. By *increasing* the pulse length we approach the RWA regime, mitigating the effects of the counter-rotating terms. On the other hand, by *decreasing* the pulse duration the dissipative losses are diminished. When both effects are considered at once the optimal pulse duration lies in an interval in-between these two scenarios.

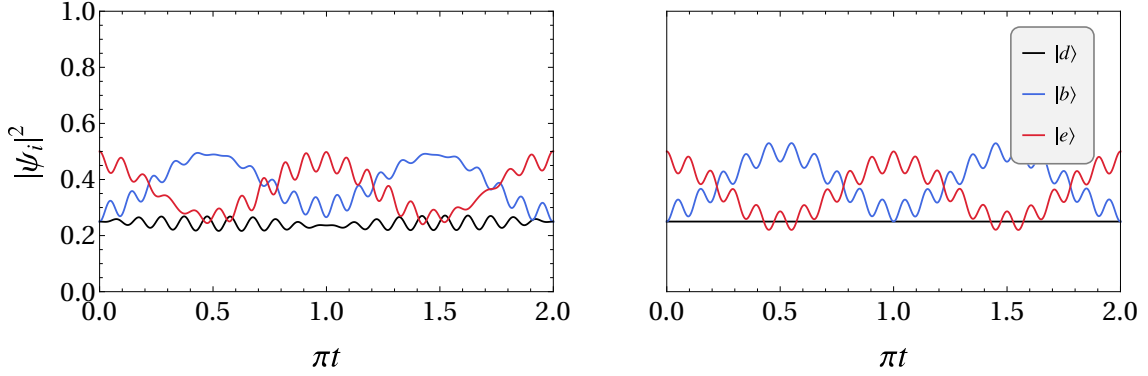
optimal pulse duration and the corresponding fidelity in a few cases of interest. Moreover, we also analyze the Hamiltonian in the bright-dark basis when counter-rotating corrections are considered. Surprisingly, these terms introduce a coupling between the dark and the excited state which, differently from what is seen in Eq. (7.6), is not present in the RWA regime, further hampering the protocol. Finally, we also investigate the effect of heterogeneous counter-rotating frequencies in the system. We show that they may, very slightly, improve the fidelity of two-loop gates. This chapter constitutes on our original results, discussed in Ref. [301]. Note that some similar investigations in a different context also exist. In such cases, the trade-off occurs between decoherence and the (non-ideal) finite run-time of adiabatic implementations instead [302, 303].

## 8.1 Counter-rotating effects in the bright-dark basis

During this section we will, once again, briefly disregard decoherent effects and go back to the Hamiltonian in Eq. (7.4). It is meaningful to ask ourselves what happens when we rewrite this full Hamiltonian in terms of the dark and the bright states. When we do that, we get

$$\begin{aligned}
 H_{bd}(t) = & \Omega(t)(1 + |\omega_0|^2 e^{-2if_{e0}t} + |\omega_1|^2 e^{-2if_{e1}t}) |e\rangle \langle b| \\
 & + \Omega(t)\omega_0\omega_1(e^{-2if_{e1}t} - e^{-2if_{e0}t}) |e\rangle \langle d| + \text{h.c.}
 \end{aligned} \tag{8.1}$$





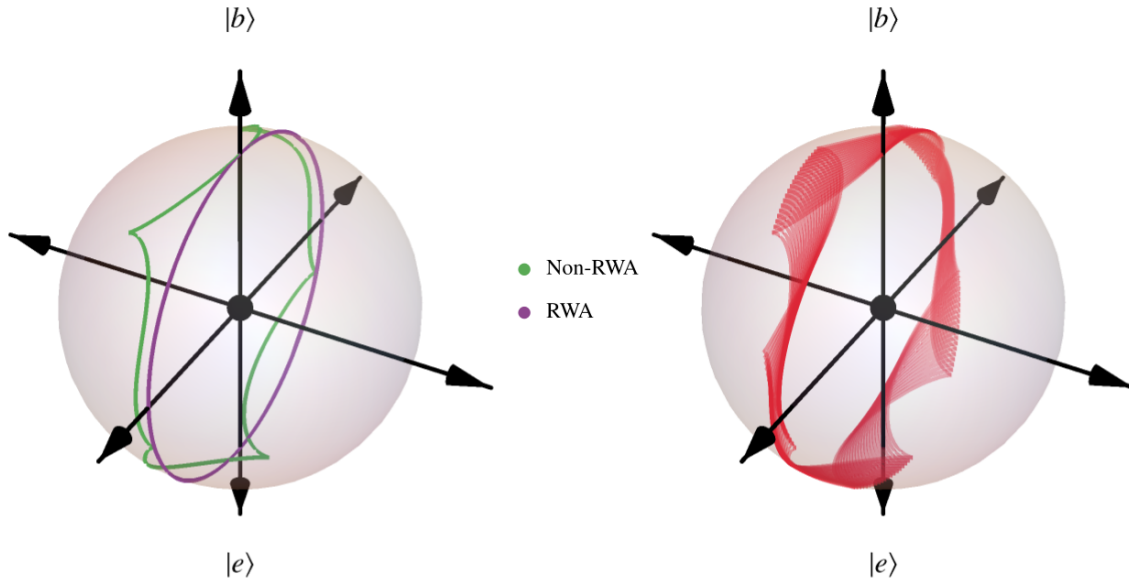
**Figure 8.2:** Dynamics of the amplitudes of the bright and dark states in the  $\Lambda$ -system, according to Eq. (8.1), with  $f_{0e} = 10$  and  $f_{1e} = 11$  (left) and Eq. (8.2) with  $f_{0e} = f_{1e} = f = 10$  (right). The initial state is  $|\psi_0\rangle = (|0\rangle + |e\rangle)/\sqrt{2}$  and the pulse is specified by the parameters  $\Omega(t) = 1$  and  $\omega_0 = \omega_1 = 1/\sqrt{2}$ .

Notice from the second line that the counter-rotating terms, differently from RWA case, couple the dark state with the rest of the system. Thus, we lose this interesting property. As we will see later on, this will play a large role for the single-qubit gates. However, the expression above assumes a particularly simple form when  $f_{e1} = f_{e0} = f$ , which is:

$$H_{bd}(t) = \Omega(t)(1 + e^{-2ift}) |e\rangle \langle b| + \text{h.c.} \quad (8.2)$$

When the two counter-rotating frequencies are the same this extra term cancels out and the dark state coupling vanishes once more. We show this effect concretely through numerical simulations in Fig. 8.2. It may also be elucidating to examine what happens when we plot the trajectory of the state on the Bloch sphere when the Hamiltonian (8.2) is considered. If we initialize the system in the state  $|\psi_0\rangle = |b\rangle$  the amplitude of the dark state remains zero at all times, so it is possible to depict the evolution of the wave function on the Bloch sphere with poles  $|b\rangle$  and  $|e\rangle$ . Results are shown in Fig. 8.3 for  $\omega_0 = \omega_1 = 1/\sqrt{2}$ .

We can make a few observations about this picture. As one would expect, the trajectory for the RWA is just a great circle around one of the axis, since the Hamiltonian just implements a rotation in the Bloch sphere. Meanwhile, the counter-rotating terms introduce a wobbling movement around the ideal path. These corrections introduce a slight deviation from the starting point after the end of the evolution. In Fig. 8.3 we can see that the wave function does not return to the initial state. Instead, what we observe is a new trajectory with a slight offset in comparison to the previous "cycle". If we run the simu-



**Figure 8.3:** (Left) Trajectory of RWA and non-RWA evolution on the Bloch sphere for the state  $|\psi_0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ . (Right) Trajectory on the Bloch sphere over a long time. Other parameters are the same as in Fig. 8.2.

lation for several cycles it is possible to see, in the right panel of Fig. 8.3, the cumulative effect of these slight deviations. Thus, even when the dark state coupling is eliminated by choosing equal counter-rotating frequencies, we do not, in general, return to the initial subspace.

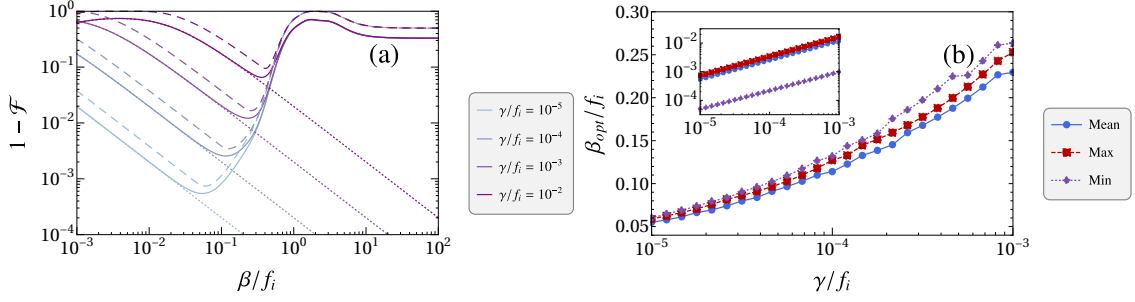
## 8.2 Single-qubit gates

Now that we have gained some further intuition on the effect of the counter-rotating contributions to the system, we are in a position to discuss the time-optimal regime of the  $\Lambda$ -Hamiltonian. In this section we investigate the performance of single-qubit gates when both counter-rotating and open quantum system effects are present. For that, we compute the fidelity

$$\mathcal{F} = \langle \psi_0 | U(C)^\dagger \rho U(C) | \psi_0 \rangle, \quad (8.3)$$

where  $\rho$  evolves according to the master equation

$$\frac{d\rho}{dt} = i[\rho, H_I] + \gamma D(|g\rangle \langle e|), \quad (8.4)$$



**Figure 8.4:** Time-optimal configuration of the S gate. (a) We plot the mean (maximum) infidelity  $1 - \mathcal{F}$  with the solid (dashed) lines for different values of  $\gamma/f_i$ . Results are shown as a function of the ratio  $\beta/f_i$  between the inverse pulse length and the counter-rotating frequency. The dotted lines show the RWA solution from Eq. (7.29), i.e., we assume that  $\beta/f_i \rightarrow 0$ . We avoid any significant overlap between the pulses by choosing a spacing of  $\Delta t = 10/\beta$  between them. (b) Plot of the optimal (inverse) pulse-length  $\beta_{\text{opt}}$  as a function of  $\gamma/f_i$ . The corresponding infidelity is shown in the inset. The grid in this case is a sample space of 1000 points uniformly distributed in the interval  $[0.03, 0.3]$ . In both panels the average infidelity was calculated for 100 input states uniformly distributed over the Bloch sphere, as described in Appendix C.

defined in Eq. (4.38). Here,  $|\psi_0\rangle$  refers to the input state. Meanwhile, the unitary  $U(C)$  is the ideal gate unitary, given by Eq. (7.15). Therefore,  $U(C)|\psi_0\rangle$  represents the ideal output. Hence, in Eq. (8.3) we calculate the fidelity between the ideal outcome and the density matrix  $\rho$  of the non-ideal process, which is obtained by solving the master equation in Eq. (8.4). The Hamiltonian  $H_I$  is, as we saw, given by Eq. (7.4) and contains the counter-rotating terms. Thus, we can see that Eq. (8.4) accounts for both the dissipative and counter-rotating contributions.

Note however that these steps yield the fidelity for a particular input state  $|\psi_0\rangle$ . A final step in our simulations is to compute the *average* fidelity for several different states which are uniformly distributed in the Bloch sphere, in order to get a more representative picture. We describe the algorithm for uniformly choosing states in the Bloch sphere in Appendix C. Moreover, for better visualization we shall focus on the infidelity  $1 - \mathcal{F}$  when plotting the figures.

We start by performing simulations for the  $S$  gate. More generally, the phase-shift gate  $|k\rangle \rightarrow e^{2ik\pi(\phi' - \phi)} |k\rangle$  can be implemented through two unit vectors:  $\mathbf{n} = (\cos \phi, \sin \phi, 0)$  and  $\mathbf{m} = (\cos \phi', \sin \phi', 0)$ , which corresponds to the choice  $\theta = \theta' = \pi/2$ , as defined in Eq. (7.15). Thus, as we mention in Sec. 7.2, the  $S$  gate can be constructed by choosing

$\phi = \pi/4$  and  $\phi' = \pi/2$ . First results are shown in Fig. 8.4(a), relating the infidelity  $1 - \mathcal{F}$  to the inverse pulse length  $\beta$  (in units of  $f_i$ ). In Fig. 8.4(b) we also plot the optimal inverse pulse length  $\beta_{\text{opt}}$ , which minimizes the infidelity from panel (a). For now, we consider the counter-rotating frequencies to be homogeneous, i.e.,  $f_{0e} = f_{1e} = f_i$ , as described by the Hamiltonian in Eq. (8.2). Besides the average infidelity, we also include the maximum and minimum infidelities (relative to the input states). Their qualitative behavior in panel (a), and also the optimal value of  $\beta$  in Fig. 8.4(b), are largely unchanged. Therefore, from hereafter we shall focus solely on the mean infidelity in our analysis, as it captures the optimal configuration and the general behaviour of the gates well enough.

We can observe the optimal regime of the system for intermediate values of  $\beta/f_i$ . If possible, one should tune the parameters of the system such that the relation  $f_i > \beta \gg \gamma$  holds. That is, the inverse pulse length  $\beta$  should be much larger than the dissipation rate  $\gamma$ , and simultaneously, the counter-rotating frequencies  $f_i$  should be sufficiently larger than  $\beta$ . By doing so, one can achieve the minimum infidelity shown on the inset of panel (b) for curves such as the ones seen on panel (a).

From Fig. 8.4(b), for example, we can observe that the ratio between the inverse pulse length and the counter-rotating frequencies should be of the order of  $\beta/f_i \approx 0.1$ , for  $\gamma/f_i \approx 10^{-4}$ , which is of the same order as parameters found in the literature [279]. The precise value of the optimal  $\beta/f_i$ , however, will depend on how strong the decoherence is. Note, in Fig. 8.4(a), how the minimum shifts depending on the value of  $\gamma/f_i$ . In particular, as  $\gamma/f_i$  decreases, we achieve the minimum infidelity by using larger pulses (smaller  $\beta/f_i$ ). This result, of course, agrees with intuition: if dissipation plays a lesser role, then it makes sense to use larger pulses in order to improve the validity of the RWA. That is why, following this idea, in Fig. 8.4(b) we plotted the optimal (inverse) pulse length  $\beta_{\text{opt}}$  as a function of  $\gamma/f_i$ . The results seem to agree with what we just described. Another interesting observation is that, from the inset in panel (b), we can see that the minimum achievable infidelity scales with a power of  $\gamma/f_i$ .

We can also discuss two other relevant regimes by analyzing Fig. 8.4(a). We begin with the regime of small  $\beta/f_i$ , describing longer pulses. In particular, we first focus on the limiting case of  $\beta/f_i \rightarrow 0$ . In this situation, all the curves should converge to the same fidelity, regardless of the value of  $\gamma/f_i$ . The reason is that the open quantum system

effects become dominant in this scenario, hampering any population transfer between the qubit states and the excited state  $|e\rangle$ . Thus, what happens in this case is that the qubit state remains largely unchanged, and the output state is approximately the same as the input state. This is not visible in the figure due to the fact that this result would require unfeasibly long simulations. Instead, we will provide an analytical calculation to support this claim near the end of the current section. We will also include a better visual picture of this result when discussing two-qubit gates in the next section.

We used dotted lines to plot the numerical solution for the RWA Hamiltonian in the purely dissipative scenario (see the caption in Fig. 8.4). As one would expect, under this approximation the infidelity decreases monotonically as we shorten the duration of the pulses. Moreover, the solution of Eq. (8.4), plotted with solid lines, agrees very well with the RWA solution, evidencing the fact that dissipation is much more relevant than the counter-rotating dynamics in this regime. We can also clearly see the breakdown of the RWA [276] in this figure. As we increase  $\beta/f_i$ , the non-RWA solution eventually starts to diverge from the RWA results around the minimum of the plot. This change of behavior marks the point where the RWA starts to fail.

In the same spirit, we eventually reach the limiting case  $\beta/f_i \gg 1$  of very short pulses. In this other regime, the counter-rotating corrections in Eq. (7.4) become  $1 + e^{-2if_it} \approx 2$ . Thus, the physical consequence is that the system undergoes a cyclical evolution twice. In other words, the bright state evolves as  $|b\rangle \rightarrow -|b\rangle \rightarrow |b\rangle$ , per Eq. (7.12). Therefore, the net effect of the protocol is simply to implement the identity gate, which, just like the strongly dissipative regime, leaves the input state (approximately) unchanged.

It is possible to quantify the infidelity in both of these cases, and the calculation happens to be the same, as we have just explained. The fidelity of a single input state is given by the overlap  $|\langle\psi_0|U(C)|\psi_0\rangle|^2$  between the target state  $U(C)|\psi_0\rangle$  and the input state  $|\psi_0\rangle$ . If we use the usual parametrization in the Bloch sphere

$$|\psi_0\rangle = \cos\left(\frac{\alpha}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\alpha}{2}\right)|1\rangle \quad (8.5)$$

for the input states, where  $\alpha$  and  $\varphi$  describe the polar and the azimuthal angle in the sphere, respectively, the mean infidelity can be obtained through an integration over the

Bloch sphere for all input states:

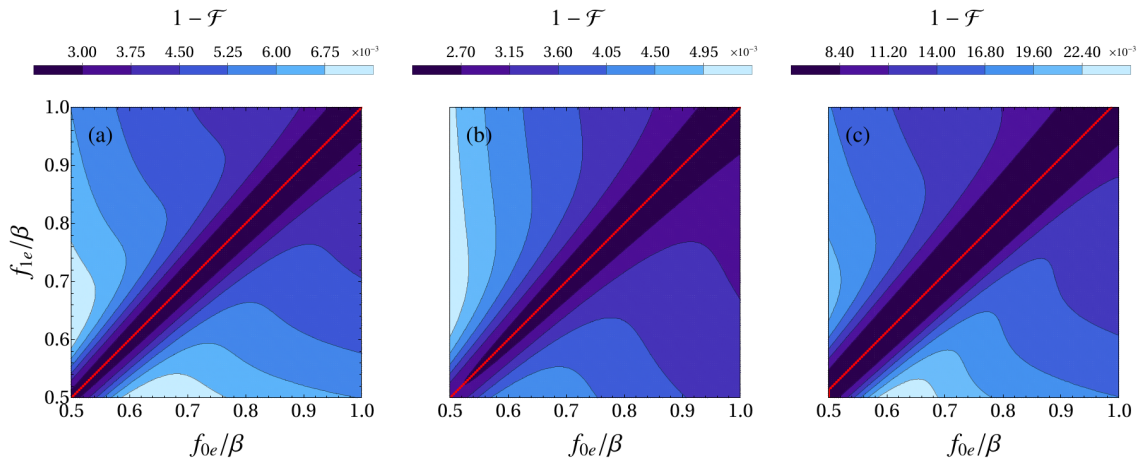
$$1 - \mathcal{F}_\infty = \frac{1}{4\pi} \int_0^{2\pi} d\varphi \int_0^\pi d\alpha |\langle \psi_0 | U(C) | \psi_0 \rangle|^2. \quad (8.6)$$

For the  $S$  gate, a direct calculation yields  $1 - \mathcal{F}_\infty = 2/3$ . As we mentioned, this is not visible for  $\beta/f_i \ll 1$ . However, we can visually verify it for  $\beta/f_i \gg 1$  in Fig. 8.4(a).

### 8.3 Heterogeneous frequencies

Now we move to the case where the two counter-rotating frequencies appearing in Eq. (8.1) are different. As we have mentioned before, whenever  $f_{e0} \neq f_{1e}$  there is a coupling between the dark and the excited state, as seen in the last term of Eq. (8.1). We will be interested in what role this difference might play in comparison to the homogeneous case.

In Fig. 8.5 we show a contour plot for the infidelity  $1 - \mathcal{F}$  as a function of  $f_{0e}/\beta$  and  $f_{1e}/\beta$ . We show the results for the  $X$ ,  $H$  and  $S$  gates, respectively. This plot can give us a few important insights. First, the  $S$  gate is more susceptible to decoherence due to the fact that it requires two pulses, as we discussed in the paragraph above Eq. (7.15). Additionally, it is also possible to observe that Fig. 8.5(a) is symmetric, while Fig. 8.5(b)



**Figure 8.5:** Contour plot of the infidelity for heterogeneous frequencies. Average infidelity  $1 - \mathcal{F}$  as a function of the counter-rotating frequencies  $f_{0e}$  and  $f_{1e}$  for  $\gamma/\beta = 0.02$ . Results are shown for (a) the  $X$  gate, (b) the  $H$  gate, and (c) the  $S$  gate. Other details and parameters are chosen to be the same as those in Fig. 8.4. The simulations were done for a grid of  $150 \times 150$  frequencies.

and (c) are not. This is due to the fact that we have  $\omega_0 = \omega_1$  for the  $X$ . Meanwhile, we have different amplitudes for the H and  $S$  gates.

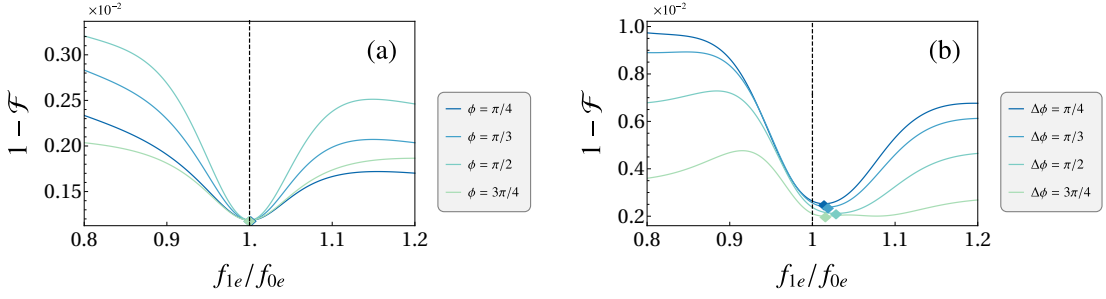
Another question we can try to answer is how  $f_{0e}$  and  $f_{1e}$  should be chosen in order to maximize the fidelity. We used a red line to plot the optimal  $f_{1e}$  for a given  $f_{0e}$ . The results shown in panels (a) and (b) show the infidelity is minimized by taking  $f_{0e} = f_{1e}$ . These results indicate that single-pulse gates are optimal for homogeneous frequencies. This possibly happens due to the decoupling between the dark and the excited states which occurs in this scenario. A further evidence of this is that we can also observe, in Figs. 8.4(a) and (b), that increasing one of the frequencies beyond the point where  $f_{0e} = f_{1e}$  actually *decreases* the fidelity of the gate. Therefore, the coupling between the  $|e\rangle$  and  $|d\rangle$  seems to play an important role in the performance of the protocol.

Surprisingly, in Fig. 8.5(c) we obtain a slightly different behavior for the  $S$  gate. The optimal regime occurs when the frequencies  $f_{0e}$  and  $f_{1e}$  are *different*. The red line shows that ideally one should increase the frequencies at a constant ratio, which is, however, different from unity. We believe that the relevant mechanism here is the non-abelian feature of the gates. Since the two unitaries in Eq. (7.15), which are necessary for implementing two-pulse gates, are non-commuting, they might be responsible for this behavior. This is supported by the fact that the fidelity of noncommuting gates is lower than the product of their fidelities. This has been verified in Ref. [291].

Another point of interest is whether other gates display the same behavior as the ones we have just seen. We start by further comparing how different frequencies affect the single and two-loop gates, plotting the infidelity as a function of the ratio  $f_{1e}/f_{0e}$  for a fixed  $f_{0e}$ . Results are shown in Fig. 8.6. We implement single and a two-loop gates in panels (a) and (b), respectively. In Fig. 8.6(a) we consider a class of traceless single-qubit gates given by Eq.(7.14). In our simulations we fix  $\theta = \pi/4$ , repeating the simulation for different values of  $\phi$ . For the second gate, shown in panel (b), we consider the phase-shift gate described in the previous section, with  $\Delta\phi = 2(\phi' - \phi)$ .

We are interested in investigating the influence of gate parameters in the optimal ratio between the two frequencies. We do this by varying  $\phi$  and  $\phi'$ . We find that single-pulse gates are optimal when the two frequencies are the same. Meanwhile, panel (b) tell us that the optimal configuration occurs when the two frequencies are *slightly* different. We

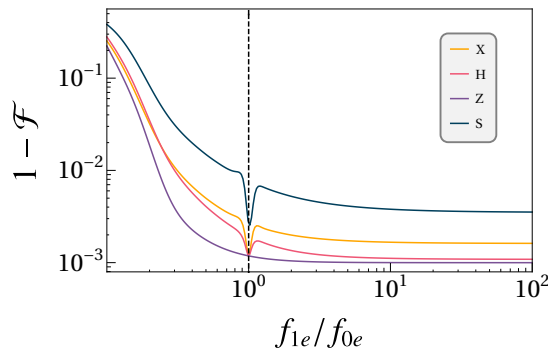




**Figure 8.6:** Infidelity as a function of the ratio  $f_{1e}/f_{0e}$  for single-qubit gates. We show the results for (a) a single-pulse gate with  $\theta = \pi/4$  and for different values of  $\phi$ . (b) we plot the results for the phase-shift gate, with  $\theta = \theta' = \frac{\pi}{2}$  and  $\Delta\phi = 2(\phi' - \phi)$ . The coupling strength is given by  $\gamma/\beta = 10^{-3}$ , the spacing between the pulses is  $\Delta t = 20/\beta$  and the frequency  $f_{0e}$  is fixed and given by  $f_{0e}/\beta = 10$ . We use a sampling space of 250 frequencies and 100 input states, as described in Fig. 8.4.

can see, however, that this gain is very marginal. This indicates that a small coupling between the dark state and the excited state is actually the most important mechanism in reducing the infidelity, and that small differences between the two frequencies are not so important. The discussion in the next paragraph makes this point more clearer, where we plot the infidelity as a function of the ratio  $f_{1e}/f_{0e}$  for different gates in Fig. 8.7.

Simulations for the  $X$ ,  $H$ , and  $S$  gates show a sharp decrease in the infidelity when  $f_{0e} \approx f_{1e}$ . For the  $X$  and  $S$  gates we have a global minimum at  $f_{e0} \approx f_{1e}$ , while for the Hadamard gate this point is only a local minimum. For the  $Z$  gate on the other hand, we observe no critical point at all; the infidelity is monically decreasing in this



**Figure 8.7:** Relevance of the  $|e\rangle \leftrightarrow |d\rangle$  coupling in the  $\Lambda$ -system. We plot the infidelity as a function of the ratio  $f_{1e}/f_{0e}$  for the NOT gate (yellow), the Hadamard gate (magenta), the  $Z$  gate (purple), and the  $S$  gate (dark blue). The decay rate, the sampling space, and  $f_{0e}$  are the same as those in Fig. 8.6. The dashed line highlights the point where  $f_{1e}/f_{0e} = 1$ . We also highlight the minimum, in each plot with a diamond mark.



case. So, whether  $f_{0e} = f_{1e}$  is the optimal gate configuration will depend on the gate parameters. Nevertheless, for all the gates shown in Fig. 8.7 we can see that homogeneous frequencies result in infidelities which are very close to the minimum. Thus, taking  $f_{0e} \approx f_{1e}$  seems like a good rule of thumb, and even for the H and Z gates, going to regime where  $f_{e1} \gg f_{0e}$  does not seem to yield much improvement.

The behavior of these curves is very intuitive if we look into the transition amplitudes for each gate. Among the single-loop gates, the X gates displays the most drastic drop. This is due to the fact that both transition frequencies have the same modulus, so this maximizes the last term in Eq. (8.1), which introduces the  $|e\rangle \leftrightarrow |d\rangle$  coupling. For that reason, the X gate (or any other gate which differs by a complex phase in the transition amplitudes), is the most sensitive to this effect. Meanwhile, in the Z gate we have  $\omega_0 = 0$  and  $\omega_1 = 1$ . Hence,  $f_{0e}$  plays no role here and we do not observe any coupling between the dark and the excited states. For that reason we get a monotonically decreasing curve in Fig. 8.7. Finally, we should observe an intermediate behavior for other gates, such as the Hadamard gate. Whether taking  $f_{0e} = f_{1e}$  or taking  $f_{1e} \gg f_{0e}$  is truly optimal will depend on the ratio between the two transition amplitudes  $\omega_0$  and  $\omega_1$ .

## 8.4 Two-qubit gates

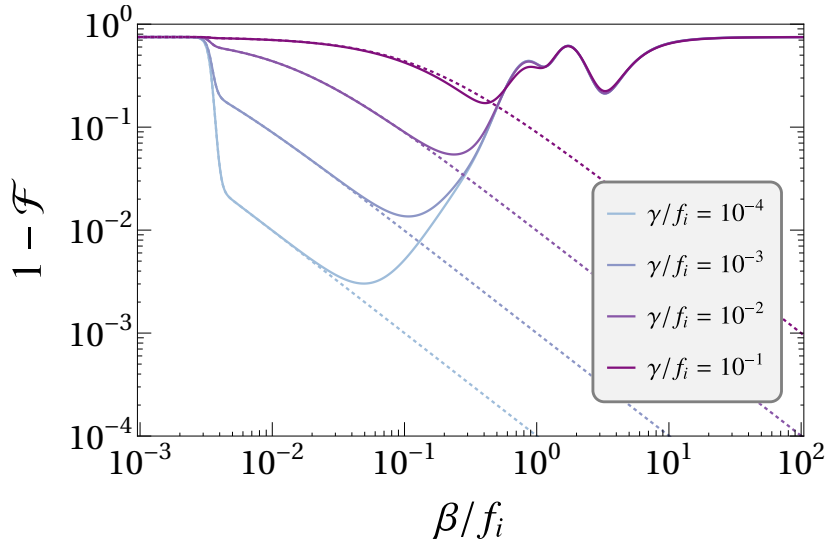
We conclude our analysis by investigating a two-qubit gate. Our findings are quite analogous to the single-qubit results. Here we consider the Hamiltonians from Eq. (7.27). Just like in the RWA case, here we also have  $[H_0^{(2)}(t), H_1^{(2)}(t)] = 0$ , so  $H_1^{(2)}(t)$  only introduces a trivial phase in the relevant subspace  $\text{span}\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ . Moreover, we only consider local noise (we suppose that there is no collective decay). The master equation in this case reads

$$\frac{d\rho^{(2)}}{dt} = i[\rho^{(2)}, H_0^{(2)}(t)] + \gamma_1 D(|g\rangle\langle e| \otimes I) + \gamma_2 D(I \otimes |g\rangle\langle e|), \quad (8.7)$$

where  $\rho^{(2)}$  is the density matrix for the two-qubit state. For simplicity we also consider that  $\gamma_1 = \gamma_2 = \gamma$ . We perform simulations for the CZ gate, defined as

$$CZ = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_z. \quad (8.8)$$

It can be constructed by choosing  $\theta = 0$ , and we show the results in Fig. 8.8. In this case, we opt for a much simpler space of input states. A uniform sampling of states for two-qubits is much harder than in the single-qubit scenario [304]. Therefore, we focus on a few cases of interest. In particular, we consider four different input states:  $|+\rangle|+\rangle$ ,  $|+\rangle|-\rangle$ ,  $|-\rangle|+\rangle$ , and  $|-\rangle|-\rangle$ . These states are useful because they yield maximally entangled states upon the application of the CZ gate. For instance,  $CZ|+\rangle|+\rangle = (|0\rangle|+\rangle + |1\rangle|-\rangle)/\sqrt{2}$ . Note how the application of the Hadamard gate in the second qubit results in the Bell state  $(|00\rangle + |11\rangle)/\sqrt{2}$ .



**Figure 8.8:** Infidelity of the CZ gate as a function of  $\beta/f_i$ . The dotted lines represent the RWA solution given by the Hamiltonian in Eq. (7.21). The infidelity is averaged over the input states  $|+\rangle|+\rangle$ ,  $|+\rangle|-\rangle$ ,  $|-\rangle|+\rangle$ , and  $|-\rangle|-\rangle$ . Other details are the same as those in Fig. 8.4.

Qualitatively, the results are very similar to what we have obtained for the single-qubit gates, demonstrating that two-qubit gates are also reasonably robust against this type of noise. Note however that their performance is much more similar to the  $S$  gate in Fig. (8.4), which is a two-loop gate. So they are not as robust as the single-loop gates, such as the  $X$  or  $H$  gates. A useful characteristic of this gate, however, is that the optimal inverse pulse length is very similar to the single-qubit scenario. This might be a useful experimental feature, since the the optimal pulse duration is of similar magnitude for both the single and two-qubit gates. Finally, note how the behavior in the limiting cases of strong dissipation or of very small counter-rotating frequencies is similar to what we

described at the end of Sec. 8.2 for the single-qubit gates. Moreover, note how we can visually observe, in this plot, how all the curves converge to the same value. As we have thoroughly discussed in Sec. 8.2; the strong decoherence spoils the protocol, and the final state is essentially unchanged. Thus, we can see that the infidelity converges to a value which is independent of  $\beta/f_i$ .

# Chapter 9

## Conclusion

In the first part of this dissertation we incorporated, through the Bayesian lens, concrete aspects of estimation theory into the collisional thermometry scheme. Our results showcased how the Bayesian framework provides an alternative point of view to thermometry, yielding further insights as a thermometric tool and also some practical advantages [240, 244]. The first aspect of this approach was the possibility of sequentially updating the temperature distribution based on the measurement outcomes. The estimators were then derived from those distributions, whose performance was then assessed through the Bayesian mean-square error. The Cramér-Rao bound and its Bayesian counterpart, the van Trees-Schützenberger inequality, were then used as some of the main benchmarks for the estimator performance. Thus, by investigating the asymptotic value of the Bayesian MSE, we performed a global analysis of protocol independently of particular values of the temperature, obtaining some of the optimal configurations for the model among the regimes which we have considered. Further generalizations, such as the use of collective measurements, are also possible. In that sense, we could also employ adaptive strategies [24, 28, 128, 231]; allowing us to optimize the coupling parameter  $\gamma_{TSE}$  between the system and the bath, the probe state [230] or the basis (in the scenario of collective measurements). An extension of the model to continuous variables systems [305], and also the use of machine learning tools [306], are also promising research venues.

Naturally, our discussion only scratches the surface of what Bayesian thermometry can offer. A very natural direction to follow is the investigation of different estimators and prior distributions. Much progress has been made in Refs. [24, 240, 244] on this

topic. Additionally, Bayesian estimation very often helps dealing with difficulties which are particular or specific to the problem at hand, such as constraints, symmetries and even experimental challenges, by means of heuristic strategies [127, 164, 176, 179, 235, 240, 307].

Likewise, investigations which dwell into further and more general aspects of quantum estimation theory in a broader sense than presented here, such as alternative bounds to the CRB and the Van-Trees Schützenberger bound, are equally important [146]. See Refs. [308] and [309] and the discussion therein, which serve as quantum formulations of the Ziv-Zakai and Weiss-Weinstein bounds from classical estimation theory [310, 311]. The recent results in Refs. [307] and [312] also provide some illuminating ideas. A bigger focus on situations where data might be scarce is an additional scenario where Bayesian estimation might shine. Some important contributions have been made in, e.g., Refs. [23, 313, 314]. Finally, what should be clear at this point is that the global treatment is not limited either to the Bayesian and frequentist techniques, nor are the two frameworks mutually exclusive [119]. Some classical works [146, 148, 149] provide important instances of a fully Bayesian treatment. Similarly, the frequentist approach is just as useful when considering practical applications. At the end of the day, the advantages and the obstacles in each approach shifts in a case-by-case scenario, and it is left to the experimenter to evaluate their clarity and usefulness in each different context. It is also important to stress, as we have seen many times, that these tools are not restricted to thermometry. Rather, Bayesian estimation has found widespread success in the quantum metrology community for decades, in a plethora of different problems, albeit it has been far less employed in quantum thermometry, in part due to the very recent history of the field.

In the second part of this work we studied a time-optimal approach to non-adiabatic holonomic quantum computation based on  $\Lambda$ -type three-level systems. More specifically, we have further investigated the validity of the rotating-wave approximation in this model and its conjunction with open quantum system effects. We showed that there is a trade-off in the gate operation time; shorter pulses can be used to protect the system against decoherence, but they are more fragile against imperfections brought from the RWA. Additionally, we were able to obtain the optimal parameters for typical experimental configurations of the  $\Lambda$ -system, allowing for an optimization of the gate performance in terms

of the pulse length.

Furthermore, we also explored the role of asymmetric frequencies in the performance of single-qubit gates. We found that different counter-rotating frequencies introduce a non-ideal coupling between the dark and the excited states of the  $\Lambda$ -Hamiltonian. For that reason, similar frequencies suppress the effect of this unwanted coupling, resulting in a sudden improvement of the gate performance in comparison to the asymmetrical configuration. For two-loop single-qubit gates on the other hand, we found that different counter-rotating frequencies result in a very slight improvement. We believe that this might be a result of the non-Abelian character of the geometrical phases used in the construction of these gates. Nevertheless, in all scenarios, approximate (but not necessarily identical) frequencies resulted in a significant improvement. These aspects are a discussion of interest as far as real experiments are considered, since, besides their own idiosyncrasies and limitations, the platform of choice might deeply constrain how much these frequencies can be fine-tuned.

A few other important aspects of this model have been left out, and might be useful subjects of future investigations. One of these directions is how well other pulse shapes might fare, such as square and Gaussian pulse. The latter, for instance, has been previously used in real experiments for NHQC [279]. It might be interesting to ask what pulse shape yields the maximum achievable fidelity when both dissipation and the breakdown of the RWA are considered. In particular, as far as we are aware, it is an open question whether certain pulse shapes are more robust against any of these effects. Some similar investigations already exist [315, 316], and they might be an useful aid in this direction of study.

The use of time-dependent perturbation theory using, e.g., the Magnus expansion, might also provide some useful insight into the model for the dissipationless case. Unfortunately, we could not find much success in the use of perturbation theory in the scope of this dissertation, in a large part due to the complexity that Sech pulses introduce into these methods, resulting in unwieldy expressions. It is, however, possible to apply these techniques at lower orders for square and trigonometric pulses. This might provide some further insight into a few parts of this work, such as the optimality of asymmetric frequencies for two-loop single-qubit gates. Nevertheless, applying an analytical approach when

both counter-rotating contributions and open quantum system effects are considered is still a challenge.

Finally, among generalizations and future investigations, one can include extensions of our study into different implementations of the NQHC protocol for  $\Lambda$ -systems, such as the off-resonant scheme [280, 317], the single-loop scheme [281], environment-assisted implementations [318], path shortening [283], and time-optimal-control implementations [319, 320]. It would also be possible to include other type of open quantum system effects, such as dephasing and collective noise, or even other type of errors, such as imperfections on the pulse amplitude itself [276]. Extensions to the multi-qubit gates [321, 322] and strategies which possibly do not rely on RWA [323] also provide interesting research venues.

# Appendix A

## Vectorization

In this appendix we discuss the vectorization procedure. See Ref. [324] for an operational approach. In the field of quantum information, the concept is known as the Choi–Jamiołkowski isomorphism [325–328]. Under this isomorphism we can very conveniently define an operation known as the Hilbert-Schmidt inner product:

$$\mathrm{tr}\{A^\dagger B\} = \mathrm{vec}(A)^\dagger \mathrm{vec}(B), \quad (\text{A.1})$$

which, as written in this form, is based on a process known as *vectorization*, where one "stacks" the rows of a matrix into a single column:

$$\mathrm{vec} \begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}. \quad (\text{A.2})$$

In order to prove Eq. (A.1), we can write the trace of the product  $A^\dagger B$  as:

$$\mathrm{tr}\{A^\dagger B\} = \sum_i (A^\dagger B)_{ii} = \sum_{ij} A_{ij}^\dagger B_{ji} = \sum_{ij} (A_{ji}^*)^\dagger B_{ji}. \quad (\text{A.3})$$

Note however that if we can write the matrix  $A$  as  $A = (\vec{A}_1 \ \vec{A}_2 \ \dots \ \vec{A}_n)$ , where we define  $\vec{A}_i$  as the  $i$ -th column of the matrix  $A$ , we can write the element  $A_{ji}$  as  $A_{ji} = (\vec{A}_j)_i$ .



Thus, the previous equality becomes:

$$\text{tr}\{A^\dagger B\} = \sum_j (\vec{A}_i^*)_j (\vec{B}_i)_j = \text{vec}(A^*) \cdot \text{vec}(B) = \text{vec}(A)^\dagger \text{vec}(B), \quad (\text{A.4})$$

as we wanted to show.

Another very important identity is the following:

$$\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B). \quad (\text{A.5})$$

As a particular case, we can take either  $A = I$  or  $C = I$  to show that the product between two matrices can be expressed as:

$$AB = (I \otimes A) \text{vec}(B) = (B^T \otimes I) \text{vec}(A). \quad (\text{A.6})$$

To see this, note that the  $k$ -th row of this product can be written as:

$$(\overrightarrow{ABC})_k = AB\vec{C}_k = A \sum_i \vec{B}_i C_{ik}. \quad (\text{A.7})$$

However we can split this sum in a more explicit way:

$$(\overrightarrow{ABC})_k = \sum_i (AC_{ik}) \vec{B}_i = \begin{pmatrix} C_{1k}A & C_{2k}A & \dots \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \vdots \end{pmatrix}.$$

The row vector however is simply  $(\vec{C}_k^T \otimes A)$ , and the second term can readily be identified as  $\vec{B}$ . Therefore, by stacking all the  $(\overrightarrow{ABC})_k$  together we arrive at:

$$A\vec{B}C = \begin{pmatrix} \vec{C}_1^T \otimes A \\ \vec{C}_2^T \otimes A \\ \vdots \end{pmatrix} \vec{B} = (C^T \otimes A) \text{vec}(B), \quad (\text{A.8})$$

which is the identity we wanted to prove.

# Appendix B

## Bayesian inference for correlated ancillae

In this appendix we show how Eq. (5.4) can be decomposed in a more general manner. From (5.2), we can write:

$$\begin{aligned} P(X_1, \dots, X_n|T) &= P(X_n|T, X_1, \dots, X_{n-1}) \\ &\quad \times P(X_{n-1}|T, X_1, \dots, X_{n-2}) \\ &\quad \times \dots \\ &\quad \times P(X_2|T, X_1) \\ &\quad \times P(X_1|T). \end{aligned} \tag{B.1}$$

Each of these transition probabilities can all be calculated directly from the model. For instance, we can write  $P(X_2|X_1) = \text{tr}(M_2 M_1 \rho_{A_1 A_2}) / \text{tr}(M_1 \rho_{A_1 A_2})$ .

Focusing on the case where  $T$  is discretized into steps  $T_k$ , we can now generalize Eq. (5.10) to

$$L_{kn} = \sum_{i=1}^n \ln p(X_i|T_k, X_1, \dots, X_{i-1}). \tag{B.2}$$

By using Eq. (B.2) in the place of Eq. (5.10), the construction from chapter 5 remains valid, even when considering dependent outcomes. In other words, the formalism itself remains the same, and we should just be careful to simply incorporate the correlations into the likelihood.

Since the results of Fig. 5.1 show that the mutual information decays exponentially with the distance between the ancillas, one does not need, in practice, to retain the full hierarchy of distributions in Eq. (B.2). It is possible, instead, to perform a truncation at a finite Markov order. For the sake of the example, let us consider the case where only nearest-neighbor correlations are important. We may then approximate:

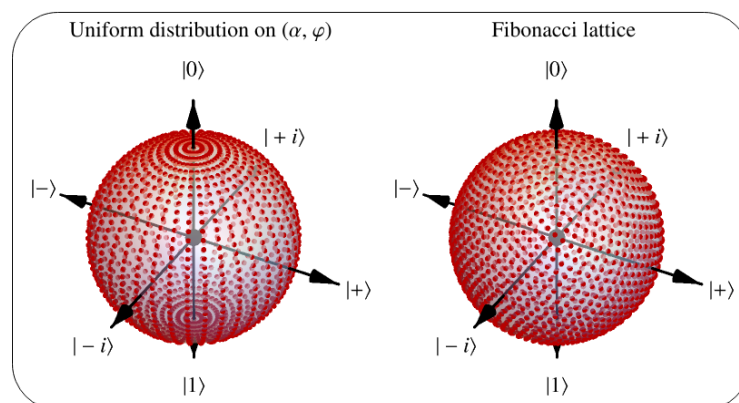
$$L_{kn} \simeq \sum_{i=1}^n \ln P(X_i|T_k, X_{i-1}), \quad (\text{B.3})$$

where  $P(X_i|T_k, X_{i-1})$  forms essentially a Markov chain. This feature is very useful, since it is hard to handle the theoretical model due to the exponentially-increasing dimensions of the Hilbert space as extra ancillae are considered. Meanwhile, dealing with a small Markov order is analytically/numerically tractable.

# Appendix C

## Uniform sampling over the Bloch sphere

In this section we describe an algorithm to approximately distribute points on a sphere in a uniform manner. A common approach to this problem consists in randomly sampling these points with respect to the Haar measure [276]. However, since computing the fidelity for a large number of points is computationally expensive, we turn instead to a simple algorithm called Fibonacci lattice (or nodes), which is a reasonable approximation to deterministically distribute states over the Bloch sphere. This approach is used to map points from a Fibonacci lattice in a square onto the surface of a sphere through a cylindrical equal area projection [329]. This is advantageous due to the fact that it is possible to get a uniform distribution even for a small number of points, while random methods usually require a larger sample size.



**Figure C.1:** Uniform distribution of points using the Fibonacci nodes algorithm. We denote the polar and azimuthal angles by  $\alpha$  and  $\varphi$ , respectively. We plot (left) the points sampled from a uniform distribution on  $(\alpha, \varphi)$  and (right) the points sampled by using the algorithm.

This type of procedure is motivated by the fact that in the naive approach, where one uniformly samples the spherical angles, the density of points is strongly skewed towards the poles. This can be seen in Fig. (C.1), where we compare the uniform sampling with the the algorithm. The procedure for a set of  $N$  points works as follows:

1. Uniformly sample the  $z$ -coordinates  $z_n$  from the interval  $[-1, 1]$
2. Distribute the azimuthal angle according to  $\varphi_n = 2\pi\Phi n$ , where  $\Phi = (1 + \sqrt{5})/2$  is the golden ratio
3. Take  $x_n = \sqrt{1 - z_n^2} \cos \varphi_n$  and  $y_n = \sqrt{1 - z_n^2} \sin \varphi_n$

The desired set of points will be given by the set of triples  $(x_n, y_n, z_n)$  for  $n = 1, \dots, N$ . See an implementation below, in the Table C.1:

```

FibonacciNodes[np_] := Module[{x, y, z, r, alpha},
  z = Table[1 - 2  $\frac{n-1}{np-1}$ , {n, np}];
  r =  $\sqrt{1-z^2}$ ;
  alpha = (2 $\pi$  N@GoldenRatio) (Range@np);

  Transpose[{r Cos[alpha], r Sin[alpha], z}
]

```

**Table C.1:** Mathematica code for the Fibonacci nodes algorithm.

```

FibonacciStates[np_] := Module[{alphaPhiList, psi},
  psi[alpha_, phi_] := Cos[ $\frac{\alpha}{2}$ ] {1, 0} + ei phi Sin[ $\frac{\alpha}{2}$ ] {0, 1};

  (* (x, y, z) -> (Polar, Azimuthal) *)
  alphaPhiList = (ToSphericalCoordinates@FibonacciNodes[np]) [[All, {2, 3}]];

  (* Removes singularities *)
  (* Psi_1, Psi_2, ..., Psi_n *)
  psi @@@ alphaPhiList /. {Indeterminate -> 0}
]

```

**Table C.2:** Generates a set of points in the Bloch sphere from the Fibonacci nodes algorithm.

For our purposes, this procedure can be used to uniformly generate points on the Bloch sphere. The Mathematica code which generates a list containing uniformly distributed states over the Bloch sphere is shown in Table C.2.

One of the advantages of this method is that it is really simple to implement, and since the points are not distributed randomly, we can get a reasonably uniform distribution even for a small number of points. Of course, if one is interested in methods which randomly distribute the points more canonical approaches are also available [330]. For a pedagogical discussion on the Haar measure and how to apply it to quantum information theory in the random sampling of unitaries and quantum states, see Ref. [304].

# Bibliography

- [1] V. I. Goldanskii, “Chemical Reactions at Very Low Temperatures”, [Annual Review of Physical Chemistry](#) **27**, 85–126 (1976).
- [2] F. J. DiSalvo, “Thermoelectric Cooling and Power Generation”, [Science](#) **285**, 703–706 (1999).
- [3] K. Uchida, S. Takahashi, K. Harii, J. Ieda, W. Koshibae, K. Ando, S. Maekawa, and E. Saitoh, “Observation of the spin Seebeck effect”, [Nature](#) **455**, 778–781 (2008).
- [4] M. Walter, J. Walowski, V. Zbarsky, M. Münzenberg, M. Schäfers, D. Ebke, G. Reiss, A. Thomas, P. Peretzki, M. Seibt, J. S. Moodera, M. Czerner, M. Bachmann, and C. Heiliger, “Seebeck effect in magnetic tunnel junctions”, [Nature Materials](#) **10**, 742–746 (2011).
- [5] Z. Iftikhar, A. Anthore, S. Jezouin, F. D. Parmentier, Y. Jin, A. Cavanna, A. Ouerghi, U. Gennser, and F. Pierre, “Primary thermometry triad at 6 mK in mesoscopic circuits”, [Nature Communications](#) **7**, 12908 (2016).
- [6] S. Vinjanampathy and J. Anders, “Quantum thermodynamics”, [Contemporary Physics](#) **57**, 545–579 (2016).
- [7] J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk, “The role of quantum information in thermodynamics—a topical review”, [Journal of Physics A: Mathematical and Theoretical](#) **49**, 143001 (2016).
- [8] A. E. Leanhardt, T. A. Pasquini, M. Saba, A. Schirotzek, Y. Shin, D. Kielpinski, D. E. Pritchard, and W. Ketterle, “Cooling Bose-Einstein Condensates Below 500 Picokelvin”, [Science](#) **301**, 1513–1515 (2003).
- [9] P. P. Hofer, J. B. Brask, M. Perarnau-Llobet, and N. Brunner, “Quantum Thermal Machine as a Thermometer”, [Physical Review Letters](#) **119**, 090603 (2017).
- [10] M. Fujiwara, S. Sun, A. Dohms, Y. Nishimura, K. Suto, Y. Takezawa, K. Oshimi, L. Zhao, N. Sadzak, Y. Umehara, Y. Teki, N. Komatsu, O. Benson, Y. Shikano, and E. Kage-Nakadai, “Real-time nanodiamond thermometry probing in vivo thermogenic responses”, [Science Advances](#) **6**, 10.1126/sciadv.aba9636 (2020).
- [11] M. G. Paris, “Quantum estimation for quantum technology”, [International Journal of Quantum Information](#) **7**, 125–137 (2009).
- [12] T. M. Stace, “Quantum limits of thermometry”, [Physical Review A](#) **82**, 011611 (2010).

- [13] J. Goold, M. Paternostro, and K. Modi, “Nonequilibrium Quantum Landauer Principle”, [Physical Review Letters](#) **114**, 060602 (2015).
- [14] P. Kammerlander and J. Anders, “Coherence and measurement in quantum thermodynamics”, [Scientific Reports](#) **6**, 22174 (2016).
- [15] K. Micadei, J. P. S. Peterson, A. M. Souza, R. S. Sarthour, I. S. Oliveira, G. T. Landi, T. B. Batalhão, R. M. Serra, and E. Lutz, “Reversing the direction of heat flow using quantum correlations”, [Nature Communications](#) **10**, 2456 (2019).
- [16] G. T. Landi and M. Paternostro, “Irreversible entropy production: From classical to quantum”, [Reviews of Modern Physics](#) **93**, 035008 (2021).
- [17] T. A. B. P. Silva and R. M. Angelo, “Quantum mechanical work”, [Physical Review A](#) **104**, 042215 (2021).
- [18] P. Lipka-Bartosik, M. Perarnau-Llobet, and N. Brunner, *What is the temperature of a quantum state?*, Apr. 2022.
- [19] A. Rivas and S. F. Huelga, *Open Quantum Systems*, SpringerBriefs in Physics (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).
- [20] T. B. Batalhão, A. M. Souza, L. Mazzola, R. Auccaise, R. S. Sarthour, I. S. Oliveira, J. Goold, G. De Chiara, M. Paternostro, and R. M. Serra, “Experimental Reconstruction of Work Distribution and Study of Fluctuation Relations in a Closed Quantum System”, [Physical Review Letters](#) **113**, 140601 (2014).
- [21] A. M. Timpanaro, G. Guarnieri, J. Goold, and G. T. Landi, “Thermodynamic Uncertainty Relations from Exchange Fluctuation Theorems”, [Physical Review Letters](#) **123**, 090604 (2019).
- [22] M. M. Ali, W.-M. Huang, and W.-M. Zhang, “Quantum thermodynamics of single particle systems”, [Scientific Reports](#) **10**, 13500 (2020).
- [23] J. Rubio and J. Dunningham, “Bayesian multiparameter quantum metrology with limited data”, [Physical Review A](#) **101**, 032114 (2020).
- [24] J. Boeyens, S. Seah, and S. Nimmrichter, “Uninformed Bayesian quantum thermometry”, [Physical Review A](#) **104**, 052214 (2021).
- [25] F. Ciccarello, “Collision models in quantum optics”, [Quantum Measurements and Quantum Metrology](#) **4**, 53–63 (2017).
- [26] S. Seah, S. Nimmrichter, D. Grimmer, J. P. Santos, V. Scarani, and G. T. Landi, “Collisional Quantum Thermometry”, [Physical Review Letters](#) **123**, 180602 (2019).
- [27] W.-K. Mok, K. Bharti, L.-C. Kwek, and A. Bayat, “Optimal probes for global quantum thermometry”, [Communications Physics](#) **4**, 62 (2021).
- [28] M. Mehboudi, M. R. Jørgensen, S. Seah, J. B. Brask, J. Kołodyński, and M. Perarnau-Llobet, “Fundamental Limits in Bayesian Thermometry and Attainability via Adaptive Strategies”, [Physical Review Letters](#) **128**, 130502 (2022).
- [29] D. Deutsch and R. A. Jozsa, “Rapid solution of problems by quantum computation”, [Proceedings of the Royal Society of London. Series A](#) **439**, 553–558 (1992).



- [30] P. W. Shor, “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer”, [SIAM Journal on Computing](#) **26**, 1484–1509 (1997).
- [31] D. P. DiVincenzo, “The Physical Implementation of Quantum Computation”, [Fortschritte der Physik](#) **48**, 771–783 (2000).
- [32] P. W. Shor, “Scheme for reducing decoherence in quantum computer memory”, [Physical Review A](#) **52**, R2493–R2496 (1995).
- [33] A. M. Steane, “Error Correcting Codes in Quantum Theory”, [Physical Review Letters](#) **77**, 793–797 (1996).
- [34] J. Preskill, “Reliable quantum computers”, [Proceedings of the Royal Society of London. Series A](#) **454**, 385–410 (1998).
- [35] E. Joos and H. D. Zeh, “The emergence of classical properties through interaction with the environment”, [Z. Phys. B](#) **59**, 223–243 (1985).
- [36] W. G. Unruh, “Maintaining coherence in quantum computers”, [Physical Review A](#) **51**, 992–997 (1995).
- [37] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, “Elementary gates for quantum computation”, [Physical Review A](#) **52**, 3457–3467 (1995).
- [38] J. I. Cirac and P. Zoller, “Quantum Computations with Cold Trapped Ions”, [Physical Review Letters](#) **74**, 4091–4094 (1995).
- [39] A. Kitaev, “Fault-tolerant quantum computation by anyons”, [Annals of Physics](#) **303**, 2–30 (2003).
- [40] P. Zanardi and M. Rasetti, “Holonomic quantum computation”, [Physics Letters, Section A: General, Atomic and Solid State Physics](#) **264**, 94–99 (1999).
- [41] L.-M. Duan, J. I. Cirac, and P. Zoller, “Geometric Manipulation of Trapped Ions for Quantum Computation”, [Science](#) **292**, 1695–1697 (2001).
- [42] J. Zhang, T. H. Kyaw, S. Filipp, L.-C. Kwek, E. Sjöqvist, and D. Tong, “Geometric and holonomic quantum computation”, [2709](#) (2021).
- [43] W. Xiang-Bin and M. Keiji, “Nonadiabatic Conditional Geometric Phase Shift with NMR”, [Physical Review Letters](#) **87**, 097901 (2001).
- [44] S.-L. Zhu and Z. D. Wang, “Implementation of Universal Quantum Gates Based on Nonadiabatic Geometric Phases”, [Physical Review Letters](#) **89**, 097902 (2002).
- [45] E. Sjöqvist, D. M. Tong, L. Mauritz Andersson, B. Hessmo, M. Johansson, and K. Singh, “Non-adiabatic holonomic quantum computation”, [New Journal of Physics](#) **14**, 103035 (2012).
- [46] P. Shen, T. Chen, and Z.-Y. Xue, “Ultrafast Holonomic Quantum Gates”, [Physical Review Applied](#) **16**, 044004 (2021).
- [47] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, “Detection of the Free Neutrino: a Confirmation”, [Science](#) **124**, 103–104 (1956).
- [48] E. L. Kaplan and P. Meier, “Nonparametric Estimation from Incomplete Observations”, [Journal of the American Statistical Association](#) **53**, 457 (1958).

- [49] R. A. Fisher, “Statistical Methods for Research Workers”, [Biometrics](#) **27**, 1106 (1971).
- [50] D. R. Cox, “Regression Models and Life-Tables”, [Journal of the Royal Statistical Society: Series B \(Methodological\)](#) **34**, 187–202 (1972).
- [51] J. Haslett and A. E. Raftery, “Space-Time Modelling with Long-Memory Dependence: Assessing Ireland’s Wind Power Resource”, [Applied Statistics](#) **38**, 1 (1989).
- [52] N. Lange and S. L. Zeger, “Non-linear Fourier Time Series Analysis for Human Brain Mapping by Functional Magnetic Resonance Imaging”, [Journal of the Royal Statistical Society: Series C \(Applied Statistics\)](#) **46**, 1–29 (1997).
- [53] J. P. Hughes, P. Guttorp, and S. P. Charles, “A non-homogeneous hidden Markov model for precipitation occurrence”, [Journal of the Royal Statistical Society. Series C: Applied Statistics](#) **48**, 15–30 (1999).
- [54] D. Dochain, “State and parameter estimation in chemical and biochemical processes: a tutorial”, [Journal of Process Control](#) **13**, 801–818 (2003).
- [55] V. N. Vapnik and A. Y. Chervonenkis, “On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities”, in *Measures of complexity*, Vol. XVI, 2 (Springer International Publishing, Cham, 2015), pp. 11–30.
- [56] B. P. Abbott et al., “Observation of Gravitational Waves from a Binary Black Hole Merger”, [Physical Review Letters](#) **116**, 061102 (2016).
- [57] J. E. Bernhard, J. S. Moreland, and S. A. Bass, “Bayesian estimation of the specific shear and bulk viscosity of quark–gluon plasma”, [Nature Physics](#) **15**, 1113–1117 (2019).
- [58] E. T. Jaynes, *Probability Theory*, edited by G. L. Bretthorst (Cambridge University Press, Cambridge, 2003).
- [59] A. Klenke, *Probability Theory*, Universitext (Springer London, London, 2014).
- [60] P. J. Bickel and K. A. Doksum, *Mathematical Statistics* (Chapman and Hall/CRC, New York, 2015).
- [61] J. A. Nelder and R. W. M. Wedderburn, “Generalized Linear Models”, [Journal of the Royal Statistical Society. Series A \(General\)](#) **135**, 370 (1972).
- [62] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood from Incomplete Data Via the EM Algorithm”, [Journal of the Royal Statistical Society: Series B \(Methodological\)](#) **39**, 1–22 (1977).
- [63] B. Efron, “Bootstrap Methods: Another Look at the Jackknife”, [The Annals of Statistics](#) **7**, 701–726 (2007).
- [64] W. Gilks, S. Richardson, and D. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice* (Chapman and Hall/CRC, 1995).
- [65] L. Breiman, “Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)”, [Statistical Science](#) **16**, 199–215 (2001).
- [66] S. M. Stigler, *Statistics on the Table: The History of Statistical Concepts and Methods* (Harvard University Press, Cambridge, 1999).

- [67] J. Tabak, *Probability and Statistics: The Science of Uncertainty (History of Mathematics)* (Facts on File, New York, 2004).
- [68] C. R. Rao, “Information and the Accuracy Attainable in the Estimation of Statistical Parameters”, in *Breakthroughs in statistics*, Chapter 16 (1992), pp. 235–247.
- [69] H. Cramér, *Mathematical Methods of Statistics*, Repr (Princeton University Press, Princeton, 1999).
- [70] J. A. Rice, *Mathematical Statistics and Data Analysis*, 3rd ed. (Duxbury Press, Belmont, 2006).
- [71] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning* (Cambridge University Press, 2014).
- [72] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. (Springer New York, New York, 2006).
- [73] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Vol. I (Prentice Hall, Upper Saddle River, 1993).
- [74] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Texts in Statistics (Springer New York, New York, 1998).
- [75] A. E. Hoerl and R. W. Kennard, “Ridge Regression: Biased Estimation for Nonorthogonal Problems”, *Technometrics* **12**, 55–67 (1970).
- [76] A. Y. Ng, “Feature selection, L 1 vs. L 2 regularization, and rotational invariance”, in *Twenty-first international conference on machine learning - icml '04* (2004), p. 78.
- [77] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 9th ed. (Cengage Learning, Boston, 2016).
- [78] G. Clark, A. Gonye, and S. J. Miller, “Lessons from the German Tank Problem”, *The Mathematical Intelligencer* **43**, 19–28 (2021).
- [79] I. Miller and M. Miller, *John E. Freund’s Mathematical Statistics with Applications*, 8th ed. (Pearson, Harlow, 2014).
- [80] G. Casella and R. L. Berger, *Statistical Inference* (Duxbury Press, Pacific Grove, 2002).
- [81] T. Galili and I. Meilijson, “An Example of an Improvable Rao–Blackwell Improvement, Inefficient Maximum Likelihood Estimator, and Unbiased Generalized Bayes Estimator”, *The American Statistician* **70**, 108–113 (2016).
- [82] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, 2005).
- [83] A. Etz, “Introduction to the Concept of Likelihood and Its Applications”, *Advances in Methods and Practices in Psychological Science* **1**, 60–69 (2018).
- [84] A. Etz, “Technical Notes on Kullback-Leibler Divergence”, *PsyArXiv*, [10.31234/osf.io/5vhzu](https://doi.org/10.31234/osf.io/5vhzu) (2018).
- [85] O. Rioul, *Teoria da informação e da condificação* (Editora Unicamp, Campinas, 2018).

- [86] P. Zegers, “Fisher Information Properties”, *Entropy* **17**, 4918–4939 (2015).
- [87] A. Ly, M. Marsman, J. Verhagen, R. Grasman, and E.-J. Wagenmakers, “A Tutorial on Fisher Information”, *Journal of Mathematical Psychology* **80**, 40–55 (2017).
- [88] J. P. Romano and A. F. Siegel, *Counterexamples in Probability and Statistics* (Routledge, New York, 2017).
- [89] A. Host-Madsen, “On the existence of efficient estimators”, *IEEE Transactions on Signal Processing* **48**, 3028–3031 (2000).
- [90] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data* (Cambridge University Press, Cambridge, 2010).
- [91] J. Aldrich, “R.A. Fisher and the making of maximum likelihood 1912-1922”, *Statistical Science* **12**, 162–176 (1997).
- [92] S. M. Stigler, “The Epic Story of Maximum Likelihood”, *Statistical Science* **22**, 598–620 (2007).
- [93] M. Hardy, “An Illuminating Counterexample”, *The American Mathematical Monthly* **110**, 234 (2003).
- [94] P. B. Stark and L. Tenorio, “A Primer of Frequentist and Bayesian Inference in Inverse Problems”, in *Large-scale inverse problems and quantification of uncertainty* (John Wiley & Sons, Ltd, Chichester, UK, 2010), pp. 9–32.
- [95] L. Held and D. Sabanés Bové, *Applied Statistical Inference* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014).
- [96] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (IEEE, 2007).
- [97] C. P. Robert, *The Bayesian Choice*, Springer Texts in Statistics (Springer New York, New York, NY, 1994).
- [98] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics (Springer New York, New York, NY, 1985).
- [99] E. W. Weisstein, *Gamma Function*, Available at: <https://mathworld.wolfram.com/GammaFunction.html>.
- [100] K. P. Murphy, *Conjugate Bayesian Analysis of the Gaussian Distribution*, 2007.
- [101] R. Bassett and J. Deride, “Maximum a posteriori estimators as a limit of Bayes estimators”, *Mathematical Programming* **174**, 129–144 (2019).
- [102] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (John Wiley & Sons, Inc., New York, USA, 2001).
- [103] L. Rüschemdorf, *Mathematische Statistik* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014).
- [104] C. Stein, “Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution”, *Berkeley Symposium on Mathematical Statistics and Probability*, 197–206 (1956).

- [105] W James and C. Stein, “Estimation with Quadratic Loss”, in *Breakthroughs in statistics: foundations and basic theory*, edited by S. Kotz and N. L. Johnson (Springer New York, New York, NY, 1992), pp. 443–460.
- [106] J. Aitchison, “Two Papers on the Comparison of Bayesian and Frequentist Approaches to Statistical Problems of Prediction: Bayesian Tolerance Regions”, *Journal of the Royal Statistical Society: Series B (Methodological)* **26**, 161–175 (1964).
- [107] J. M. Bland and D. G Altman, “Statistics notes: Bayesians and frequentists”, *BMJ* **317**, 1151–1160 (1998).
- [108] J. Vallverdú, “The False Dilemma: Bayesian vs. Frequentist”, (2008).
- [109] F. J. Samaniego, *A Comparison of the Bayesian and Frequentist Approaches to Estimation*, Springer Series in Statistics (Springer New York, New York, NY, 2010).
- [110] J. M. Bernardo and A. F. M. Smith, eds., *Bayesian Theory*, Wiley Series in Probability and Statistics (John Wiley & Sons, Inc., Hoboken, NJ, USA, 1994).
- [111] M. P. Schützenberger, “A generalization of the Fréchet-Cramér inequality to the case of Bayes estimation”, *Bulletin of the American Mathematical Society* **63**, 142 (1957).
- [112] R. D. Gill and B. Y. Levit, “Applications of the van Trees Inequality: A Bayesian Cramér-Rao Bound”, *Bernoulli* **1**, 59–79 (1995).
- [113] L. L. Cam, *Asymptotic Methods in Statistical Decision Theory*, Springer Series in Statistics (Springer New York, New York, NY, 1986), p. 564.
- [114] A. W. van der Vaart, *Asymptotic Statistics* (Cambridge University Press, 1998).
- [115] D. V. Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2: Inference* (Cambridge University Press, Cambridge, 1965).
- [116] J. A. Hartigan, “Asymptotic Normality of Posterior Distributions”, in *A course in large sample theory* (1983), pp. 107–118.
- [117] R. W. Butler, *Saddlepoint Approximations with Applications* (Cambridge University Press, Cambridge, 2007).
- [118] L. Tierney and J. B. Kadane, “Accurate Approximations for Posterior Moments and Marginal Densities”, *Journal of the American Statistical Association* **81**, 82–86 (1986).
- [119] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms* (Cambridge University Press, Cambridge, 2002).
- [120] A. M. Walker, “On the Asymptotic Behaviour of Posterior Distributions”, *Journal of the Royal Statistical Society: Series B (Methodological)* **31**, 80–88 (1969).
- [121] R. Ramakrishna and A. Scaglione, “A Bayesian lower bound for parameters with bounded support priors”, in *2020 54th annual conference on information sciences and systems (ciss)* (2020).
- [122] L. Bacharach, C. Fritsche, U. Orguner, and E. Chaumette, “A Tighter Bayesian Cramér-Rao Bound”, in *Icassp 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (icassp)*, Vol. 2019-May (2019), pp. 5277–5281.

- [123] V. Giovannetti, S. Lloyd, and L. Maccone, “Advances in quantum metrology”, *Nature Photonics* **5**, 222–229 (2011).
- [124] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, “Quantum metrology with nonclassical states of atomic ensembles”, *Reviews of Modern Physics* **90**, 035005 (2018).
- [125] M. Yu, Y. Liu, P. Yang, M. Gong, Q. Cao, S. Zhang, H. Liu, M. Heyl, T. Ozawa, N. Goldman, and J. Cai, “Quantum Fisher information measurement and verification of the quantum Cramér–Rao bound in a solid-state qubit”, *npj Quantum Information* **8**, 56 (2022).
- [126] S. L. Braunstein and C. M. Caves, “Statistical distance and the geometry of quantum states”, *Physical Review Letters* **72**, 3439–3443 (1994).
- [127] A. H. Kiilerich and K. Mølmer, “Quantum Zeno effect in parameter estimation”, *Physical Review A* **92**, 032124 (2015).
- [128] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, “Quantum Metrology for Noisy Systems”, *Brazilian Journal of Physics* **41**, 229–247 (2011).
- [129] U. Marzolino and T. Prosen, “Quantum metrology with nonequilibrium steady states of quantum spin chains”, *Physical Review A - Atomic, Molecular, and Optical Physics* **90**, 1–10 (2014).
- [130] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell, and S. Pirandola, “Quantum-enhanced measurements without entanglement”, *Reviews of Modern Physics* **90**, 035006 (2018).
- [131] Z. Y. Ou, “Complementarity and Fundamental Limit in Precision Phase Measurement”, *Physical Review Letters* **77**, 2352–2355 (1996).
- [132] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, “Quantum algorithms revisited”, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **454**, 339–354 (1998).
- [133] J. H. Shapiro, S. R. Shepard, and N. C. Wong, “Ultimate quantum limits on phase measurement”, *Physical Review Letters* **62**, 2377–2380 (1989).
- [134] S. L. Braunstein, “Quantum limits on precision measurements of phase”, *Physical Review Letters* **69**, 3598–3601 (1992).
- [135] H. Lee, P. Kok, and J. P. Dowling, “A quantum Rosetta stone for interferometry”, *Journal of Modern Optics* **49**, 2325–2338 (2002).
- [136] A. Luis, “Phase-shift amplification for precision measurements without nonclassical states”, *Physical Review A* **65**, 025802 (2002).
- [137] L. Maccone, “Intuitive reason for the usefulness of entanglement in quantum metrology”, *Physical Review A* **88**, 042109 (2013).
- [138] K. Wang, X. Wang, X. Zhan, Z. Bian, J. Li, B. C. Sanders, and P. Xue, “Entanglement-enhanced quantum metrology in a noisy environment”, *Physical Review A* **97**, 042112 (2018).
- [139] P. Zanardi, M. G. A. Paris, and L. Campos Venuti, “Quantum criticality as a resource for quantum estimation”, *Physical Review A* **78**, 042105 (2008).



- [140] L. Garbe, M. Bina, A. Keller, M. G. Paris, and S. Felicetti, “Critical Quantum Metrology with a Finite-Component Quantum Phase Transition”, [Physical Review Letters](#) **124**, 120504 (2020).
- [141] Z. J. Ying, S. Felicetti, G. Liu, and D. Braak, “Critical Quantum Metrology in the Non-Linear Quantum Rabi Model”, [Entropy](#) **24**, 1–11 (2022).
- [142] C. M. Caves, “Quantum-mechanical noise in an interferometer”, [Physical Review D](#) **23**, 1693–1708 (1981).
- [143] G. J. Milburn, W.-Y. Chen, and K. R. Jones, “Hyperbolic phase and squeeze-parameter estimation”, [Physical Review A](#) **50**, 801–804 (1994).
- [144] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland, “Toward Heisenberg-Limited Spectroscopy with Multiparticle Entangled States”, [Science](#) **304**, 1476–1478 (2004).
- [145] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, “Quantum entanglement”, [Reviews of Modern Physics](#) **81**, 865–942 (2009).
- [146] S. Personick, “Application of quantum estimation theory to analog communication over quantum channels”, [IEEE Transactions on Information Theory](#) **17**, 240–246 (1971).
- [147] C. W. Helstrom, “Quantum detection and estimation theory”, [Journal of Statistical Physics](#) **1**, 231–252 (1969).
- [148] C. W. Helstrom, *Quantum detection and estimation theory* (Academic Press, New York, NY, 1976).
- [149] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, Pisa, 2011).
- [150] V. Giovannetti, S. Lloyd, and L. Maccone, “Quantum Metrology”, [Physical Review Letters](#) **96**, 010401 (2006).
- [151] G. Tóth and I. Apellaniz, “Quantum metrology from a quantum information science perspective”, [Journal of Physics A: Mathematical and Theoretical](#) **47**, 424006 (2014).
- [152] M. A. Taylor and W. P. Bowen, “Quantum metrology and its application in biology”, [Physics Reports](#) **615**, 10.1016/j.physrep.2015.12.002 (2016).
- [153] E. Polino, M. Valeri, N. Spagnolo, and F. Sciarrino, “Photonic quantum metrology”, [AVS Quantum Science](#) **2**, 024703 (2020).
- [154] M. Mehboudi, A. Sanpera, and L. A. Correa, “Thermometry in the quantum regime: recent theoretical progress”, [Journal of Physics A: Mathematical and Theoretical](#) **52**, 303001 (2019).
- [155] S. Seah, S. Nimmrichter, D. Grimmer, J. P. Santos, V. Scarani, and G. T. Landi, “Collisional Quantum Thermometry”, [Physical Review Letters](#) **123**, 180602 (2019).
- [156] L. A. Clark, A. Stokes, and A. Beige, “Quantum jump metrology”, [Physical Review A](#) **99**, 022102 (2019).
- [157] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, 2012).

- [158] D. Šafránek, “Simple expression for the quantum Fisher information matrix”, [Physical Review A \*\*97\*\*, 042322 \(2018\)](#).
- [159] J. Liu, H. Yuan, X. M. Lu, and X. Wang, “Quantum Fisher information matrix and multiparameter estimation”, [Journal of Physics A: Mathematical and Theoretical \*\*53\*\*, 10.1088/1751-8121/ab5d4d \(2020\)](#).
- [160] J. S. Sidhu and P. Kok, “Geometric perspective on quantum parameter estimation”, [AVS Quantum Science \*\*2\*\*, 014701 \(2020\)](#).
- [161] H. Xing and L. Fu, “Measure of the density of quantum states in information geometry and quantum multiparameter estimation”, [Physical Review A \*\*102\*\*, 062613 \(2020\)](#).
- [162] M. Tsang, “Physics-inspired forms of the Bayesian Cramér-Rao bound”, [Physical Review A \*\*102\*\*, 062217 \(2020\)](#).
- [163] W. Guo, W. Zhong, X. X. Jing, L. B. Fu, and X. Wang, “Berry curvature as a lower bound for multiparameter estimation”, [Physical Review A \*\*93\*\*, 1–7 \(2016\)](#).
- [164] F. Chapeau-Blondeau, “Optimized probing states for qubit phase estimation with general quantum noise”, [Physical Review A \*\*91\*\*, 052310 \(2015\)](#).
- [165] R. Demkowicz-Dobrzański, J. Czajkowski, and P. Sekatski, “Adaptive Quantum Metrology under General Markovian Noise”, [Physical Review X \*\*7\*\*, 041009 \(2017\)](#).
- [166] M. A. Rodríguez-García, I. P. Castillo, and P. Barberis-Blostein, “Efficient qubit phase estimation using adaptive measurements”, [Quantum \*\*5\*\*, 467 \(2021\)](#).
- [167] K. Wan and R. Lasenby, “Bounds on adaptive quantum metrology under Markovian noise”, [Physical Review Research \*\*4\*\*, 033092 \(2022\)](#).
- [168] K. Modi, H. Cable, M. Williamson, and V. Vedral, “Quantum Correlations in Mixed-State Metrology”, [Physical Review X \*\*1\*\*, 021022 \(2011\)](#).
- [169] T. R. Bromley, I. A. Silva, C. O. Oncebay-Segura, D. O. Soares-Pinto, E. R. deAzevedo, T. Tufarelli, and G. Adesso, “There is more to quantum interferometry than entanglement”, [Physical Review A \*\*95\*\*, 052313 \(2017\)](#).
- [170] U. Khalid, Y. Jeong, and H. Shin, “Measurement-based quantum correlation in mixed-state quantum metrology”, [Quantum Information Processing \*\*17\*\*, 343 \(2018\)](#).
- [171] F. Chapeau-Blondeau, “Optimizing qubit phase estimation”, [Physical Review A \*\*94\*\*, 022334 \(2016\)](#).
- [172] O. E. Barndorff-Nielsen and R. D. Gill, “Fisher information in quantum statistics”, [Journal of Physics A: Mathematical and General \*\*33\*\*, 4481–4490 \(2000\)](#).
- [173] C. Li, M. Chen, and P. Cappellaro, “A geometric perspective: experimental evaluation of the quantum Cramer-Rao bound”, [\(2022\)](#).
- [174] M. Yu, X. Li, Y. Chu, B. Mera, F. N. Ünal, P. Yang, Y. Liu, N. Goldman, and J. Cai, “Experimental demonstration of topological bounds in quantum metrology”, [\(2022\)](#).
- [175] M. Hayashi, *Asymptotic Theory of Quantum Statistical Inference* (WORLD SCIENTIFIC, 2005).



- [176] B. Teklu, S. Olivares, and M. G. A. Paris, “Bayesian estimation of one-parameter qubit gates”, [Journal of Physics B: Atomic, Molecular and Optical Physics](#) **42**, 035502 (2009).
- [177] D. Brivio, S. Cialdi, S. Vezzoli, B. T. Gebrehiwot, M. G. Genoni, S. Olivares, and M. G. A. Paris, “Experimental estimation of one-parameter qubit gates in the presence of phase diffusion”, [Physical Review A](#) **81**, 012305 (2010).
- [178] T. Wasak, A. Smerzi, L. Pezzé, and J. Chwedeńczuk, “Optimal measurements in phase estimation: simple examples”, [Quantum Information Processing](#) **15**, 2231–2252 (2016).
- [179] R. Schmied, “Quantum state tomography of a single qubit: comparison of methods”, [Journal of Modern Optics](#) **63**, 1744–1758 (2016).
- [180] D. Šafránek, “Discontinuities of the quantum Fisher information and the Bures metric”, [Physical Review A](#) **95**, 052320 (2017).
- [181] K Micadei, D. A. Rowlands, F. A. Pollock, L. C. Céleri, R. M. Serra, and K Modi, “Coherent measurements in quantum metrology”, [New Journal of Physics](#) **17**, 023057 (2015).
- [182] F. Belliardo and V. Giovannetti, “Achieving Heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error”, [Physical Review A](#) **102**, 042613 (2020).
- [183] L. Pezzé, A. Smerzi, G. Khoury, J. F. Hodelin, and D. Bouwmeester, “Phase Detection at the Quantum Limit with Multiphoton Mach-Zehnder Interferometry”, [Physical Review Letters](#) **99**, 223602 (2007).
- [184] F. Albarelli, M. A. C. Rossi, D. Tamascelli, and M. G. Genoni, “Restoring Heisenberg scaling in noisy quantum metrology by monitoring the environment”, [Quantum](#) **2**, 110 (2018).
- [185] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, “General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology”, [Nature Physics](#) **7**, 406–411 (2011).
- [186] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, “General Optimality of the Heisenberg Limit for Quantum Metrology”, [Physical Review Letters](#) **105**, 180402 (2010).
- [187] M Bruderer and D Jaksch, “Probing BEC phase fluctuations with atomic quantum dots”, [New Journal of Physics](#) **8**, 87–87 (2006).
- [188] D. C. McKay and B DeMarco, “Cooling in strongly correlated optical lattices: prospects and challenges”, [Reports on Progress in Physics](#) **74**, 054401 (2011).
- [189] C. Sabín, A. White, L. Hackermuller, and I. Fuentes, “Impurities as a quantum thermometer for a Bose-Einstein condensate”, [Scientific Reports](#) **4**, 6436 (2015).
- [190] M. Fujiwara and Y. Shikano, “Diamond quantum thermometry: from foundations to applications”, [Nanotechnology](#) **32**, 482002 (2021).
- [191] U. Marzolino and D. Braun, “Precision measurements of temperature and chemical potential of quantum gases”, [Physical Review A](#) **88**, 063609 (2013).

- [192] M. Mehboudi, A. Lampo, C. Charalambous, L. A. Correa, M. Á. García-March, and M. Lewenstein, “Using Polarons for sub-nK Quantum Nondemolition Thermometry in a Bose-Einstein Condensate”, *Physical Review Letters* **122**, 030403 (2019).
- [193] Q. Bouton, J. Nettersheim, D. Adam, F. Schmidt, D. Mayer, T. Lausch, E. Tiemann, and A. Widera, “Single-Atom Quantum Probes for Ultracold Gases Boosted by Nonequilibrium Spin Dynamics”, *Physical Review X* **10**, 011018 (2020).
- [194] M. M. Khan, M. Mehboudi, H. Terças, M. Lewenstein, and M. A. Garcia-March, “Subnanokelvin thermometry of an interacting d -dimensional homogeneous Bose gas”, *Physical Review Research* **4**, 023191 (2022).
- [195] G. Kucsko, P. C. Maurer, N. Y. Yao, M. Kubo, H. J. Noh, P. K. Lo, H. Park, and M. D. Lukin, “Nanometre-scale thermometry in a living cell”, *Nature* **500**, 54–58 (2013).
- [196] H. Rangani Jahromi, S. E. A. Mamaghani, and R. Lo Franco, “Relativistic quantum thermometry through a moving sensor”, *Annals of Physics* **448**, 169172 (2023).
- [197] R. B. Mann and E. Martín-Martínez, “Quantum Thermometry”, *Foundations of Physics* **44**, 492–511 (2014).
- [198] O. Pereira de Sa Neto, “Temperature Estimation in an Gravitational-Wave”, *SSRN Electronic Journal*, 10.2139/ssrn.4294880 (2022).
- [199] A. De Pasquale, K. Yuasa, and V. Giovannetti, “Estimating temperature via sequential measurements”, *Physical Review A* **96**, 012316 (2017).
- [200] K. V. Hovhannisyán and L. A. Correa, “Measuring the temperature of cold many-body quantum systems”, *Physical Review B* **98**, 045101 (2018).
- [201] V. Mukherjee, A. Zwick, A. Ghosh, X. Chen, and G. Kurizki, “Enhanced precision bound of low-temperature quantum thermometry via dynamical control”, *Communications Physics* **2**, 162 (2019).
- [202] M. T. Mitchison, T. Fogarty, G. Guarnieri, S. Campbell, T. Busch, and J. Goold, “*IN Situ* Thermometry of a Cold Fermi Gas via Dephasing Impurities”, *Physical Review Letters* **125**, 080402 (2020).
- [203] M Salado-Mejía, R Román-Ancheyta, F Soto-Eguibar, and H. M. Moya-Cessa, “Spectroscopy and critical quantum thermometry in the ultrastrong coupling regime”, *Quantum Science and Technology* **6**, 025010 (2021).
- [204] G. Planella, M. F. B. Cenni, A. Acín, and M. Mehboudi, “Bath-Induced Correlations Enhance Thermometry Precision at Low Temperatures”, *Physical Review Letters* **128**, 040502 (2022).
- [205] F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso, eds., *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions*, Vol. 195, Fundamental Theories of Physics (Springer International Publishing, Cham, 2018).
- [206] M. Mehboudi, A. Sanpera, and L. A. Correa, “Thermometry in the quantum regime: recent theoretical progress”, *Journal of Physics A: Mathematical and Theoretical* **52**, 303001 (2019).

- [207] S. Deffner and S. Campbell, *Quantum Thermodynamics: An introduction to the thermodynamics of quantum information*, July 2019.
- [208] S. Jevtic, D. Newman, T. Rudolph, and T. M. Stace, “Single-qubit thermometry”, [Physical Review A \*\*91\*\*, 012331 \(2015\)](#).
- [209] L. A. Correa, M. Perarnau-Llobet, K. V. Hovhannisyanyan, S. Hernández-Santana, M. Mehboudi, and A. Sanpera, “Enhancement of low-temperature thermometry by strong coupling”, [Physical Review A \*\*96\*\*, 062103 \(2017\)](#).
- [210] L. A. Correa, M. Perarnau-Llobet, K. V. Hovhannisyanyan, S. Hernández-Santana, M. Mehboudi, and A. Sanpera, “Enhancement of low-temperature thermometry by strong coupling”, [Physical Review A \*\*96\*\*, 062103 \(2017\)](#).
- [211] M. R. Jørgensen, P. P. Potts, M. G. A. Paris, and J. B. Brask, “Tight bound on finite-resolution quantum thermometry at low temperatures”, [Physical Review Research \*\*2\*\*, 033394 \(2020\)](#).
- [212] P. P. Potts, J. B. Brask, and N. Brunner, “Fundamental limits on low-temperature quantum thermometry with finite resolution”, [Quantum \*\*3\*\*, 161 \(2019\)](#).
- [213] M. Brunelli, S. Olivares, M. Paternostro, and M. G. A. Paris, “Qubit-assisted thermometry of a quantum harmonic oscillator”, [Physical Review A \*\*86\*\*, 012125 \(2012\)](#).
- [214] L. A. Correa, M. Mehboudi, G. Adesso, and A. Sanpera, “Individual Quantum Probes for Optimal Thermometry”, [Physical Review Letters \*\*114\*\*, 220405 \(2015\)](#).
- [215] K. V. Hovhannisyanyan, M. R. Jørgensen, G. T. Landi, Á. M. Alhambra, J. B. Brask, and M. Perarnau-Llobet, “Optimal Quantum Thermometry with Coarse-Grained Measurements”, [PRX Quantum \*\*2\*\*, 020322 \(2021\)](#).
- [216] V. Giovannetti and G. M. Palma, “Master Equations for Correlated Quantum Channels”, [Physical Review Letters \*\*108\*\*, 040401 \(2012\)](#).
- [217] S. Seah, S. Nimmrichter, and V. Scarani, “Nonequilibrium dynamics with finite-time repeated interactions”, [Physical Review E \*\*99\*\*, 042103 \(2019\)](#).
- [218] S. Campbell and B. Vacchini, “Collision models in open system dynamics: A versatile tool for deeper insights?”, [Europhysics Letters \*\*133\*\*, 60001 \(2021\)](#).
- [219] Y. Li, X. Li, and J. Jin, “Dissipation-Induced Information Scrambling in a Collision Model”, [Entropy \*\*24\*\*, 345 \(2022\)](#).
- [220] G. De Chiara, G. Landi, A. Hewgill, B. Reid, A. Ferraro, A. J. Roncaglia, and M. Antezza, “Reconciliation of quantum local master equations with thermodynamics”, [New Journal of Physics \*\*20\*\*, 113024 \(2018\)](#).
- [221] F. L. S. Rodrigues, G. De Chiara, M. Paternostro, and G. T. Landi, “Thermodynamics of Weakly Coherent Collisional Models”, [Physical Review Letters \*\*123\*\*, 140601 \(2019\)](#).
- [222] G. T. Landi, “Battery Charging in Collision Models with Bayesian Risk Strategies”, [Entropy \*\*23\*\*, 1627 \(2021\)](#).
- [223] S. Seah, M. Perarnau-Llobet, G. Haack, N. Brunner, and S. Nimmrichter, “Quantum Speed-Up in Collisional Battery Charging”, [Physical Review Letters \*\*127\*\*, 100601 \(2021\)](#).

- [224] R. Salvia, M. Perarnau-Llobet, G. Haack, N. Brunner, and S. Nimmrichter, *Quantum advantage in charging cavity and spin batteries by repeated interactions*, Apr. 2022.
- [225] S. Cusumano, “Quantum Collision Models: A Beginner Guide”, [Entropy](#) **24**, 1258 (2022).
- [226] F. Ciccarello, S. Lorenzo, V. Giovannetti, and G. M. Palma, “Quantum collision models: open system dynamics from repeated interactions”, [Physics Reports](#) **954**, 1–70 (2022).
- [227] J. A. Gross, C. M. Caves, G. J. Milburn, and J. Combes, “Qubit models of weak continuous measurements: markovian conditional and open-system dynamics”, [Quantum Science and Technology](#) **3**, 024005 (2018).
- [228] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*, First (Oxford University Press, Oxford, Jan. 2007).
- [229] V. Scarani, M. Ziman, P. Štelmachovič, N. Gisin, and V. Bužek, “Thermalizing Quantum Machines: Dissipation and Entanglement”, [Physical Review Letters](#) **88**, 097905 (2002).
- [230] A. Shu, S. Seah, and V. Scarani, “Surpassing the thermal Cramér-Rao bound with collisional thermometry”, [Physical Review A](#) **102**, 042417 (2020).
- [231] E. O’Connor, B. Vacchini, and S. Campbell, “Stochastic Collisional Quantum Thermometry”, [Entropy](#) **23**, 1634 (2021).
- [232] J. Fiurášek, “Maximum-likelihood estimation of quantum measurement”, [Physical Review A](#) **64**, 024102 (2001).
- [233] A. Ly, M. Marsman, J. Verhagen, R. P. Grasman, and E.-J. Wagenmakers, “A Tutorial on Fisher information”, [Journal of Mathematical Psychology](#) **80**, 40–55 (2017).
- [234] Y. Suzuki, S. Uno, R. Raymond, T. Tanaka, T. Onodera, and N. Yamamoto, “Amplitude estimation without phase estimation”, [Quantum Information Processing](#) **19**, 75 (2020).
- [235] U. von Toussaint, “Bayesian inference in physics”, [Reviews of Modern Physics](#) **83**, 943–999 (2011).
- [236] A. H. Kiilerich and K. Mølmer, “Bayesian parameter estimation by continuous homodyne detection”, [Physical Review A](#) **94**, 032103 (2016).
- [237] F. Hanamura, W. Asavanant, K. Fukui, S. Konno, and A. Furusawa, “Estimation of Gaussian random displacement using non-Gaussian states”, [Physical Review A](#) **104**, 062601 (2021).
- [238] S. Morelli, A. Usui, E. Agudelo, and N. Friis, “Bayesian parameter estimation using Gaussian states and measurements”, [Quantum Science and Technology](#) **6**, 025018 (2021).
- [239] Y. Li, L. Pezzè, M. Gessner, Z. Ren, W. Li, and A. Smerzi, “Frequentist and Bayesian Quantum Phase Estimation”, [Entropy](#) **20**, 628 (2018).

- [240] J. Rubio, J. Anders, and L. A. Correa, “Global Quantum Thermometry”, [Physical Review Letters](#) **127**, 190402 (2021).
- [241] H. B. Prosper, “Temperature fluctuations in a heat bath”, [American Journal of Physics](#) **61**, 54–58 (1993).
- [242] M. R. Jørgensen, J. Kołodyński, M. Mehboudi, M. Perarnau-Llobet, and J. B. Brask, “Bayesian quantum thermometry based on thermodynamic length”, [Physical Review A](#) **105**, 042601 (2022).
- [243] J. Glatthard, J. Rubio, R. Sawant, T. Hewitt, G. Barontini, and L. A. Correa, *Optimal cold atom thermometry using adaptive Bayesian strategies*, Oct. 2022.
- [244] J. Rubio, “Quantum scale estimation”, [Quantum Science and Technology](#) **8**, 015009 (2023).
- [245] G. O. Alves and G. T. Landi, “Bayesian estimation for collisional thermometry”, [Physical Review A](#) **105**, 012212 (2022).
- [246] R. Olf, F. Fang, G. E. Marti, A. MacRae, and D. M. Stamper-Kurn, “Thermometry and cooling of a Bose gas to 0.02 times the condensation temperature”, [Nature Physics](#) **11**, 720–723 (2015).
- [247] S. Gammelmark and K. Mølmer, “Bayesian parameter inference from continuously monitored quantum systems”, [Physical Review A](#) **87**, 032115 (2013).
- [248] F. Ciccarello, “Stochastic versus Periodic Quantum Collision Models”, [Open Systems & Information Dynamics](#) **29**, 2250006 (2022).
- [249] R. R. Camasca and G. T. Landi, “Memory kernel and divisibility of Gaussian collisional models”, [Physical Review A](#) **103**, 022202 (2021).
- [250] N. E. Comar and G. T. Landi, “Correlations breaking homogenization”, [Physical Review A](#) **104**, 032217 (2021).
- [251] M. Cohen, “The Fisher information and convexity (Corresp.)”, [IEEE Transactions on Information Theory](#) **14**, 591–592 (1968).
- [252] W. Ehrenberg and R. E. Siday, “The Refractive Index in Electron Optics and the Principles of Dynamics”, [Proceedings of the Physical Society. Section B](#) **62**, 8–21 (1949).
- [253] T. R. G. Peter A. Sturrock, “Concerning the Quantum Interference Effects Associated with Electric and Magnetic Potentials [Letter Submitted to the Editor, Physics Today] by Peter A Sturrock and Timothy R. Groves”.
- [254] Y Aharonov and D. Bohm, “Significance of Electromagnetic Potentials in the Quantum Theory”, [Physical Review](#) **115**, 485–491 (1959).
- [255] Y. Aharonov and D. Bohm, “Further Considerations on Electromagnetic Potentials in the Quantum Theory”, [Physical Review](#) **123**, 1511–1524 (1961).
- [256] E. Cohen, H. Larocque, F. Bouchard, F. Nejdassattari, Y. Gefen, and E. Karimi, “Geometric phase from Aharonov–Bohm to Pancharatnam–Berry and beyond”, [Nature Reviews Physics](#) **1**, 437–449 (2019).
- [257] M. V. Berry, “Quantal phase factors accompanying adiabatic changes”, [Proceedings of the Royal Society of London. Series A](#) **392**, 45–57 (1984).

- [258] S. Pancharatnam, “Generalized theory of interference, and its applications - Part I. Coherent pencils”, [Proceedings of the Indian Academy of Sciences - Section A](#) **44**, 247–262 (1956).
- [259] O. Gamel and D. F. V. James, “Measures of quantum state purity and classical degree of polarization”, [Physical Review A](#) **86**, 033830 (2012).
- [260] T. Kato, “On the Adiabatic Theorem of Quantum Mechanics”, [Journal of the Physical Society of Japan](#) **5**, 435–439 (1950).
- [261] B. Simon, “Tosio Kato’s work on non-relativistic quantum mechanics, Part 2”, [Bulletin of Mathematical Sciences](#) **09**, 1950005 (2019).
- [262] Y. Aharonov and J. Anandan, “Phase change during a cyclic quantum evolution”, [Physical Review Letters](#) **58**, 1593–1596 (1987).
- [263] J. Anandan, “Non-adiabatic non-abelian geometric phase”, [Physics Letters A](#) **133**, 171–175 (1988).
- [264] F. Wilczek and A. Zee, “Appearance of Gauge Structure in Simple Dynamical Systems”, [Physical Review Letters](#) **52**, 2111–2114 (1984).
- [265] E. Sjöqvist, “Geometric phases in quantum information”, [International Journal of Quantum Chemistry](#) **115**, 1311–1326 (2015).
- [266] E. Sjöqvist, V. Azimi Mousolou, and C. M. Canali, “Conceptual aspects of geometric quantum computation”, [Quantum Information Processing](#) **15**, 3995–4011 (2016).
- [267] D. Chruściński and A. Jamiolkowski, *Geometric Phases in Classical and Quantum Mechanics* (Birkhäuser Boston, Boston, MA, 2004).
- [268] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics* (Cambridge University Press, 2020).
- [269] S. M. Girvin and K. Yang, *Modern Condensed Matter Physics* (Cambridge University Press, 2019).
- [270] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanziger, *The Geometric Phase in Quantum Systems* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003).
- [271] J. H. Hannay, “Angle variable holonomy in adiabatic excursion of an integrable Hamiltonian”, [Journal of Physics A: Mathematical and General](#) **18**, 221–230 (1985).
- [272] P. Solinas, P. Zanardi, and N. Zanghì, “Robustness of non-Abelian holonomic quantum gates against parametric noise”, [Physical Review A](#) **70**, 042316 (2004).
- [273] P. Solinas, M. Sassetti, P. Truini, and N. Zanghì, “On the stability of quantum holonomic gates”, [New Journal of Physics](#) **14**, 093006 (2012).
- [274] L. Viotti, F. C. Lombardo, and P. I. Villar, “Geometric phase in a dissipative Jaynes-Cummings model: Theoretical explanation for resonance robustness”, [Physical Review A](#) **105**, 022218 (2022).
- [275] J. Zhang, T. H. Kyaw, D. M. Tong, E. Sjöqvist, and L.-C. Kwek, “Fast non-Abelian geometric gates via transitionless quantum driving”, [Scientific Reports](#) **5**, 18414 (2016).



- [276] M. Johansson, E. Sjöqvist, L. M. Andersson, M. Ericsson, B. Hessmo, K. Singh, and D. M. Tong, “Robustness of nonadiabatic holonomic gates”, [Physical Review A - Atomic, Molecular, and Optical Physics](#) **86**, 062322 (2012).
- [277] C. Zu, W.-B. Wang, L. He, W.-G. Zhang, C.-Y. Dai, F. Wang, and L.-M. Duan, “Experimental realization of universal geometric quantum gates with solid-state spins”, [Nature](#) **514**, 72–75 (2014).
- [278] S. Arroyo-Camejo, A. Lazariev, S. W. Hell, and G. Balasubramanian, “Room temperature high-fidelity holonomic single-qubit gate on a solid-state spin”, [Nature Communications](#) **5**, 4870 (2014).
- [279] A. A. Abdumalikov Jr, J. M. Fink, K. Juliusson, M. Pechal, S. Berger, A. Wallraff, and S. Filipp, “Experimental realization of non-Abelian non-adiabatic geometric gates”, [Nature](#) **496**, 482–485 (2013).
- [280] E. Sjöqvist, “Nonadiabatic holonomic single-qubit gates in off-resonant  $\Lambda$  systems”, [Physics Letters A](#) **380**, 65–67 (2016).
- [281] E. Herterich and E. Sjöqvist, “Single-loop multiple-pulse nonadiabatic holonomic quantum gates”, [Physical Review A](#) **94**, 052310 (2016).
- [282] B. B. Zhou, P. C. Jerger, V. O. Shkolnikov, F. J. Heremans, G. Burkard, and D. D. Awschalom, “Holonomic Quantum Control by Coherent Optical Excitation in Diamond”, [Physical Review Letters](#) **119**, 140503 (2017).
- [283] Y. Xu, W. Cai, Y. Ma, X. Mu, L. Hu, T. Chen, H. Wang, Y. P. Song, Z.-Y. Xue, Z.-q. Yin, and L. Sun, “Single-Loop Realization of Arbitrary Nonadiabatic Holonomic Single-Qubit Quantum Gates in a Superconducting Circuit”, [Physical Review Letters](#) **121**, 110501 (2018).
- [284] C. J. G. Mommers and E. Sjöqvist, “Universal quantum computation and quantum error correction using discrete holonomies”, [Physical Review A](#) **105**, 022402 (2022).
- [285] S. Blanes, F. Casas, J. Oteo, and J. Ros, “The Magnus expansion and some of its applications”, [Physics Reports](#) **470**, 151–238 (2009).
- [286] D. Zeuch, F. Hassler, J. J. Slim, and D. P. DiVincenzo, “Exact rotating wave approximation”, [Annals of Physics](#) **423**, 168327 (2020).
- [287] N. Rosen and C. Zener, “Double Stern-Gerlach Experiment and Related Collision Phenomena”, [Physical Review](#) **40**, 502–507 (1932).
- [288] I. Roos and K. Mølmer, “Quantum computing with an inhomogeneously broadened ensemble of ions: Suppression of errors from detuning variations by specially adapted pulses and coherent population trapping”, [Physical Review A](#) **69**, 022321 (2004).
- [289] S. E. Economou, L. J. Sham, Y. Wu, and D. G. Steel, “Proposal for optical U(1) rotations of electron spin trapped in a quantum dot”, [Physical Review B](#) **74**, 205415 (2006).
- [290] S. E. Economou and T. L. Reinecke, “Theory of Fast Optical Spin Rotation in a Quantum Dot Based on Geometric Phases and Trapped States”, [Physical Review Letters](#) **99**, 217401 (2007).

- [291] J. Spiegelberg and E. Sjöqvist, “Validity of the rotating-wave approximation in nonadiabatic holonomic quantum computation”, [Physical Review A \*\*88\*\*, 054301 \(2013\)](#).
- [292] A. Sørensen and K. Mølmer, “Quantum Computation with Ions in Thermal Motion”, [Physical Review Letters \*\*82\*\*, 1971–1974 \(1999\)](#).
- [293] P. Z. Zhao, G. F. Xu, and D. M. Tong, “Nonadiabatic holonomic multiqubit controlled gates”, [Physical Review A \*\*99\*\*, 052309 \(2019\)](#).
- [294] G. Kirchmair, J. Benhelm, F. Zähringer, R. Gerritsma, C. F. Roos, and R. Blatt, “Deterministic entanglement of ions in thermal states of motion”, [New Journal of Physics \*\*11\*\*, 023002 \(2009\)](#).
- [295] D. Wineland, C. Monroe, W. Itano, D. Leibfried, B. King, and D. Meekhof, “Experimental issues in coherent quantum-state manipulation of trapped atomic ions”, [Journal of Research of the National Institute of Standards and Technology \*\*103\*\*, 259 \(1998\)](#).
- [296] X.-Q. Li, L.-X. Cen, G. Huang, L. Ma, and Y. Yan, “Nonadiabatic geometric quantum computation with trapped ions”, [Physical Review A \*\*66\*\*, 042320 \(2002\)](#).
- [297] K. Kim, C. F. Roos, L. Aolita, H. Häffner, V. Nebendahl, and R. Blatt, “Geometric phase gate on an optical transition for ion trap quantum computation”, [Physical Review A \*\*77\*\*, 050303 \(2008\)](#).
- [298] M. J. Bremner, C. M. Dawson, J. L. Dodd, A. Gilchrist, A. W. Harrow, D. Mortimer, M. A. Nielsen, and T. J. Osborne, “Practical Scheme for Quantum Computation with Any Two-Qubit Entangling Gate”, [Physical Review Letters \*\*89\*\*, 247902 \(2002\)](#).
- [299] A. Uhlmann, “The “transition probability” in the state space of a  $*$ -algebra”, [Reports on Mathematical Physics \*\*9\*\*, 273–279 \(1976\)](#).
- [300] R. Jozsa, “Fidelity for Mixed Quantum States”, [Journal of Modern Optics \*\*41\*\*, 2315–2323 \(1994\)](#).
- [301] G. O. Alves and E. Sjöqvist, “Time-optimal holonomic quantum computation”, [Physical Review A \*\*106\*\*, 032406 \(2022\)](#).
- [302] G. Florio, “Decoherence in Holonomic Quantum Computation”, [Open Systems & Information Dynamics \*\*13\*\*, 263–272 \(2006\)](#).
- [303] C. Lupo, P. Aniello, M. Napolitano, and G. Florio, “Robustness against parametric noise of nonideal holonomic gates”, [Physical Review A \*\*76\*\*, 012309 \(2007\)](#).
- [304] PennyLane dev Team, “Understanding the Haar Measure”, (2021).
- [305] M. F. Cenni, L. Lami, A. Acín, and M. Mehboudi, “Thermometry of Gaussian quantum systems using Gaussian measurements”, [Quantum \*\*6\*\*, 743 \(2022\)](#).
- [306] F. S. Luiz, A. d. O. Junior, F. F. Fanchini, and G. T. Landi, “Machine classification for probe-based quantum thermometry”, [Physical Review A \*\*105\*\*, 022413 \(2022\)](#).
- [307] E. Martínez-Vargas, C. Pineda, F. Leyvraz, and P. Barberis-Blostein, “Quantum estimation of unknown parameters”, [Physical Review A \*\*95\*\*, 012136 \(2017\)](#).



- [308] Y.-R. Zhang and H. Fan, “Quantum metrological bounds for vector parameters”, [Physical Review A \*\*90\*\*, 043818 \(2014\)](#).
- [309] X.-M. Lu and M. Tsang, “Quantum Weiss-Weinstein bounds for quantum metrology”, [Quantum Science and Technology \*\*1\*\*, 015002 \(2016\)](#).
- [310] J. Ziv and M. Zakai, “Some lower bounds on signal parameter estimation”, [IEEE Transactions on Information Theory \*\*15\*\*, 386–391 \(1969\)](#).
- [311] A. Weiss and E. Weinstein, “A lower bound on the mean-square error in random parameter estimation (Corresp.)”, [IEEE Transactions on Information Theory \*\*31\*\*, 680–682 \(1985\)](#).
- [312] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă, “Multi-parameter estimation beyond quantum Fisher information”, [Journal of Physics A: Mathematical and Theoretical \*\*53\*\*, 363001 \(2020\)](#).
- [313] J. Rubio, P. Knott, and J. Dunningham, “Non-asymptotic analysis of quantum metrology protocols beyond the Cramér–Rao bound”, [Journal of Physics Communications \*\*2\*\*, 015027 \(2018\)](#).
- [314] J. Rubio and J. Dunningham, “Quantum metrology in the presence of limited data”, [New Journal of Physics \*\*21\*\*, 043037 \(2019\)](#).
- [315] F. Zhang, J. Zhang, P. Gao, and G. Long, “Searching nonadiabatic holonomic quantum gates via an optimization algorithm”, [Physical Review A \*\*100\*\*, 012329 \(2019\)](#).
- [316] L.-N. Ji, Y. Liang, P. Shen, and Z.-Y. Xue, “Nonadiabatic Holonomic Quantum Computation via Path Optimization”, [\(2022\)](#).
- [317] G. F. Xu, C. L. Liu, P. Z. Zhao, and D. M. Tong, “Nonadiabatic holonomic gates realized by a single-shot implementation”, [Physical Review A \*\*92\*\*, 052302 \(2015\)](#).
- [318] N. Ramberg and E. Sjöqvist, “Environment-Assisted Holonomic Quantum Maps”, [Physical Review Letters \*\*122\*\*, 140501 \(2019\)](#).
- [319] Z. Han, Y. Dong, B. Liu, X. Yang, S. Song, L. Qiu, D. Li, J. Chu, W. Zheng, J. Xu, T. Huang, Z. Wang, X. Yu, X. Tan, D. Lan, M.-H. Yung, and Y. Yu, “Experimental Realization of Universal Time-optimal non-Abelian Geometric Gates”, [\(2020\)](#).
- [320] T. Chen, P. Shen, and Z.-Y. Xue, “Robust and Fast Holonomic Quantum Gates with Encoding on Superconducting Circuits”, [Physical Review Applied \*\*14\*\*, 034038 \(2020\)](#).
- [321] P. Z. Zhao, G. F. Xu, and D. M. Tong, “Nonadiabatic holonomic multiqubit controlled gates”, [Physical Review A \*\*99\*\*, 1–6 \(2019\)](#).
- [322] G. F. Xu, P. Z. Zhao, E. Sjöqvist, and D. M. Tong, “Realizing nonadiabatic holonomic quantum computation beyond the three-level setting”, [Physical Review A \*\*103\*\*, 052605 \(2021\)](#).
- [323] M. M. Gontijo and J. C. A. Barata, “Phase factors of periodically driven two-level systems”, [Physica Scripta \*\*95\*\*, 055101 \(2020\)](#).
- [324] P. J. Dhrymes, “Matrix Vectorization”, in *Mathematics for econometrics* (Springer New York, New York, NY, 2000), pp. 117–145.

- [325] A. Jamiołkowski, “Linear transformations which preserve trace and positive semidefiniteness of operators”, [Reports on Mathematical Physics](#) **3**, 275–278 (1972).
- [326] M.-D. Choi, “Positive Linear Maps on Complex”, [Linear Algebra and its Applications](#) **10**, 285–290 (1975).
- [327] K. Życzkowski and I. Bengtsson, “On duality between quantum maps and quantum states”, [Open Systems and Information Dynamics](#) **11**, 3–42 (2004).
- [328] M. Jiang, S. Luo, and S. Fu, “Channel-state duality”, [Physical Review A - Atomic, Molecular, and Optical Physics](#) **87**, 1–8 (2013).
- [329] D. P. Hardin, T. Michaels, and E. B. Saff, “A Comparison of Popular Point Configurations on  $\mathbb{S}^2$ ”, [Dolomites Research Notes on Approximation](#) **9**, 16–49 (2016).
- [330] M. E. Muller, “A note on a method for generating points uniformly on n -dimensional spheres”, [Communications of the ACM](#) **2**, 19–20 (1959).