UNIVERSITY OF WARSAW

DOCTORAL THESIS

Precision bounds in
noisy quantum metrology

Author:
Jan KOŁODYŃSKI

Supervisor:
Dr hab. Rafał
DEMKOWICZ-DOBRZAŃSKI

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

in the

Chair of Quantum Optics and Atomic Physics

Institute of Theoretical Physics

FACULTY OF PHYSICS

September 2014

"They say the first sentence in any speech is always the hardest. Well, that one's behind me, anyway."

Wisława Szymborska
(Nobel Lecture, 1996)

UNIVERSITY OF WARSAW

Faculty of Physics

Abstract

Doctor of Philosophy

Precision bounds in noisy quantum metrology

by Jan Kołodyński

Quantum metrology is a vividly developing topic of current research in both theoretical and experimental physics. Its main goal is to explore the capabilities of quantum systems that, when employed as probes sensing physical parameters, allow to attain resolutions that are beyond the reach of classical protocols. Spectacularly, one may show that in an idealistic scenario, by utilising the phenomena of *quantum entanglement* and super-classical correlations, a parameter of interest may be in principle determined with mean squared error that scales as $1/N^2$ with the number of particles the system consists of—surpassing the $1/N$-scaling characteristic to classical statistics. However, a natural question arises, whether such an impressive quantum enhancement persists when one takes into account the *decoherence* effects, i.e. the noise that distorts the system and is unavoidably present in any real-life implementation.

In this thesis, we resolve a major part of this issue by describing general techniques that allow to quantify the attainable precision in metrological schemes, while accounting for the impact of uncorrelated noise-types—ones that independently disturb each constituent particle (atom, photon) in the system. In particular, we show that the abstract geometrical structure of a *quantum channel* describing the noisy evolution of a single particle dictates critical bounds on the achievable quantum enhancement. Importantly, our results prove that an infinitesimal amount of noise is enough to restrict the precision to scale classically in the asymptotic N limit, what then constrains the maximal improvement to a constant factor. Although for relatively low numbers of particles the decoherence may be ignored, for large N the presence of noise heavily alters the form of both states and measurements that should be employed to achieve the ultimate resolution. Crucially, however, the established bounds are then typically attainable with use of states and detection techniques that are natural to current experiments.

As we thoroughly introduce the necessary concepts and mathematical tools lying behind the quantum metrological tasks, including the estimation theory techniques in both *frequentist* and *Bayesian* frameworks, we hope that this work may be found attractive by researchers coming from the quantum information theory background and willing to become more familiar with the current approaches to quantum metrology problems. Throughout the work, we provide examples of applications of the methods presented to typical *qubit noise models*, yet we also discuss in detail the phase estimation task in *Mach-Zehnder interferometry* both in the classical and quantum setting—with particular emphasis given to the photonic-losses model while analysing the impact of decoherence.

Acknowledgements

List of Publications

Jan Kołodyński and Rafał Demkowicz-Dobrzański, "*Efficient tools for quantum metrology with uncorrelated noise*", New Journal of Physics **15**, 073043 (2013).

Rafał Demkowicz-Dobrzański, Jan Kołodyński, and Mădălin Guţă, "*The elusive Heisenberg limit in quantum-enhanced metrology*", Nature Communications **3**, 1063 (2012).

Jan Kołodyński and Rafał Demkowicz-Dobrzański, "*Phase estimation without a priori phase knowledge in the presence of loss*", Physical Review A **82**, 053804 (2010).

Beyond the scope of the thesis

Rafael Chaves, Jonatan Bohr Brask, Marcin Markiewicz, Jan Kołodyński, and Antonio Acín, "*Noisy metrology beyond the Standard Quantum Limit*", Physical Review Letters **111**, 120401 (2013).

Jan Kołodyński, Jan Chwedeńczuk, and Wojciech Wasilewski, "*Eigenmode description of Raman scattering in atomic vapors in the presence of decoherence*", Physical Review A **86**, 013818, (2012).

In print

Rafał Demkowicz-Dobrzański, Marcin Jarzyna, and Jan Kołodyński, "*Quantum limits in optical interferometry*", in *Progress in Optics*, edited by Emil Wolf (Elsevier), Vol. 60, (2015), arXiv:1405.7703.

Contents

List of Figures

List of Tables

List of Notes

Abbreviations

SQL	**S**tandard **Q**uantum **L**imit: $\Delta^2 \tilde{\varphi} \sim 1/N$
HL	**H**eisenberg **L**imit: $\Delta^2 \tilde{\varphi} \sim 1/N^2$
POVM	**P**ositive **O**perator **V**alued **M**easure
CPTP	**C**ompletely **P**ositive **T**race **P**reserving map
LGKS	**L**indblad-**G**orini-**K**ossakowski-**S**udarshan form
CJ	**C**hoi-**J**amiołkowski isomorphism/matrix
PDF	**P**robability **D**istribution **F**unction
CLT	**C**entral **L**imit **T**heorem
MSE	**M**ean **S**quared **E**rror
CRB	**C**ramér-**R**ao **B**ound
FI	Classical **F**isher **I**nformation
ML	**M**aximum **L**ikelihood estimator
$\overline{\textbf{MSE}}$	Average **M**ean **S**quared **E**rror
MMSE	**M**inimum **M**ean **S**quared **E**rror estimator
QCRB	**Q**uantum **C**ramér-**R**ao **B**ound
QFI	**Q**uantum **F**isher **I**nformation
SLD	**S**ymmetric **L**ogarithmic **D**erivative
RLD	**R**ight **L**ogarithmic **D**erivative
NOON	$\frac{1}{\sqrt{2}}(|N,0\rangle + |0,N\rangle)$ state
GHZ	**G**reenberger-**H**orne-**Z**eilinger state
BW	**B**erry-**W**iseman state
SDP	**S**emi-**D**efinite **P**rogram
CS	**C**lassical **S**imulation method
QS	**Q**uantum **S**imulation method
CE	**C**hannel **E**xtension method
JS	**J**ordan-**S**chwinger map

Chapter 1

Introduction

1.1 Metrology – the science of measurement

The *International Bureau of Weights and Measures* (BIPM), located in Sèvres (France) and serving since 1875 as one of the primary guards ensuring uniformity of weights and measures around the world, defines the term *metrology* as[1]:

> "(...) the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science and technology. (...)"

On the other hand, the "*Springer Handbook of Metrology and Testing*" [Czichos *et al.*, 2011] divides metrology into three sub-fields: *scientific* (fundamental), *technical* (industrial) and *legal* (imposed by the national and international law). Although the latter two stress the daily-life importance of precise measurements—which from the technological point of view play one of the main roles in the rapidly developing industry, but also must be standardised to ensure legal requirements vital to existence of modern society—their improvement can only be achieved owing to the first sub-field being constantly pursued by researchers around the globe. It is the development of metrology at the fundamental level that leads to the desired refinement of the up-to-date standards of quantities such as *mass*, *length* and *time*, often requiring the research to be lead at the borderline of the current scientific state of the art.

One of the extensively contributing fields is *quantum metrology* that is a relatively young area of physics, currently intensively studied both at the theoretical and experimental levels. As the ultra-precise measurement schemes require the finest possible resolution of sensing, they are eventually condemned to be limited by the fundamental building blocks describing the nature at the microscopic level, i.e. by the laws of *quantum mechanics* which deals with physical phenomena at the nanoscopic scales. Yet, the quantum theory has also to offer effects that importantly allow to surpass the notions that may be naively inferred from classical statistics. In particular, the "spooky"—[Einstein *et al.*, 1935]—feature of quantum theory known as *entanglement* has been shown to significantly enhance capabilities of precise-measurement techniques by exploring super-classical correlations between the system building blocks (e.g. individual atoms or photons) that sense the quantity of interest. Furthermore, with the advent of new technologies allowing to control quantum systems by means of light-matter interactions, such measurement precisions 'beyond

---

[1] http://www.bipm.org/en/convention/wmd/2004/

classical scaling-laws' have been experimentally demonstrated. In fact, these achievements constituted one of the milestones accomplished in recent years within the field of quantum physics and motivated the Royal Swedish Academy of Sciences to present the *Nobel Prize* in physics for 2012 to Serge Haroche and David J. Wineland "*for ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems*"[2] [Haroche, 2013; Wineland, 2013]. An important accomplishment of these experiments that has been crucial for their success, was the ability to defy the so-called quantum *decoherence* that typically disallows the quantum effects to be observed. Such a destructive phenomenon is a consequence of the inevitable interactions with the environment surrounding the quantum system examined, and leads to the presence of *noise* affecting any measurements performed. However, in the above experiments regimes have been spectacularly attained in which the impact of noise is negligible. On the other hand, this has imposed a novel open problem—which since then has been analysed by many researchers also at the theoretical level—as to what extent the quantum enhancement of metrological protocols can be actually observed, but when dealing with real-life quantum systems in which the noise effects cannot be any more assumed to be small.

In this work, we would like to theoretically address the above issue and discuss the consequences of the presence of *decoherence*, which strongly affects the quantum system and thus delimits the accuracy of the measurements performed. In particular, we employ the techniques developed in *quantum information theory*, in order to establish general precision bounds that account for any sources of noise that *independently* affect each of the system building blocks, i.e. the particles (atoms, photons) that a given quantum system consists of.

1.2 Classical metrology

Yet, before diving into the quantum mechanical framework designed to describe the quantum metrological tasks, one should acknowledge the immense field of *estimation theory* [Kay, 1993; Lehmann and Casella, 1998] that constitutes a major branch of statistics, and has been developed to establish most efficient techniques that allow to most accurately infer *parameters* (representing the quantities of interest) encoded in any randomly distributed data. Thus, in any—potentially quantum-based—metrological problem one is bound to use such techniques, as all that is always at hands of an experimentalist are the statistics of the particular outcomes collected. Typically, two philosophically differing approaches to such type of problems are pursued, depending whether one assumes the parameter being estimated to be a fixed (deterministic) variable just parametrising the physical model that predicts the outcome statistics, or accepts also the stochastic nature of the parameter and thus its intrinsic random fluctuations.

Within the *frequentist* approach to estimation, one supposes the estimated parameter to be a deterministic variable which, if known, could in principle be stated up to any precision. Hence, the inference process is limited only by the probabilistic nature of the measurement procedure, and one may thus utilise the well-established statistical techniques that allow to bound the effective *Mean Squared Error* (MSE) of the estimation protocol. In particular, when restricting to *unbiased* estimators that on average output the correct parameter value, the so-called *Fisher Information* (FI) can be evaluated and the *Cramér-Rao Bound* (CRB) may then be directly applied [Kay, 1993; Lehmann and Casella, 1998].

The *Bayesian* approach is somehow complementary, yet by many advocated to be the "practically valid" method of inference [see e.g. Jarzyna and Demkowicz-Dobrzański, 2014]. By assuming the parameter of interest to also be randomly distributed, not only it explicitly accounts for the knowledge one possesses about the quantity of interest before conducting the measurements, but also for the fact that the inference

---

[2]http://www.nobelprize.org/nobel_prizes/physics/laureates/2012/press.html

process should be interpreted as a data-updating routine. From the pragmatic point of view, the Bayesian techniques turn out to be more approachable when dealing with parameters exhibiting *symmetries*, what simplifies the form of the adequate Bayesian *estimation costs* that must be minimised in order to establish the optimal inference strategy. Furthermore, the structure of the Bayesian approach allows to naturally analyse the so-called *adaptive measurement schemes* in which the knowledge about the parameter is consecutively updated after each data-collection step, so that the measurement settings may be gradually adjusted for the procedure to be more efficient.

Nevertheless, the distinction between the above two approaches blurs and eventually vanishes in the limit of infinitely many experimental trials, so that it is often the case that both methods may be straightforwardly interrelated [van der Vaart, 1998]. However, let us already remark that in the quantum mechanical setting one of the key purposes of employing a quantum system consisting of many particles is the ability to profit from its 'stronger than classical' inter-particle correlations. As a result, even when restricting to single-particle measurements, their outcomes are *not* independently distributed, what opens doors to many novel interesting questions. In particular, as by raising then the number of particles present in the setup one does not naturally increase the number of independent trials, no longer the classical intuitions relating the frequentist and Bayesian approaches apply [Gill and Guţă, 2013; Jarzyna and Demkowicz-Dobrzański, 2014].

1.3 Quantum metrology

The classical *problem of parameter estimation* has been explicitly restructured by Helstrom [1976] and Holevo [1982], in order to incorporate the laws of quantum theory that dictate the measurement-outcome statistics. In their seminal works, both frequentist and Bayesian tools known from classical estimation theory have been adapted to apply in the quantum setting. Within the frequentist framework, the notions of *Quantum Fisher Information* (QFI) and the *Quantum Cramér-Rao Bound* (QCRB) have been established [Barndorff-Nielsen and Gill, 2000; Braunstein and Caves, 1994; Hayashi, 2005; Nagaoka, 1989], whereas within the Bayesian paradigm the structure of the most general quantum measurements has been determined that naturally incorporates the estimated parameter symmetries. In particular, for parameters exhibiting *group symmetries*, the notion of the so-called *covariant measurements* has been introduced, which are provably always the optimal ones being also parametrised with the group elements induced by the estimation problem [Chiribella *et al.*, 2005; Hayashi, 2005].

On the other hand, without resorting to general *quantum estimation* techniques, but rather focusing on particular experimental scenarios and concrete measurement strategies, it has been explicitly demonstrated that quantum enhancement of precision is theoretically possible in *optical inteferometry* [Bondurant and Shapiro, 1984; Caves, 1981; Dowling, 1998; Holland and Burnett, 1993; Sanders and Milburn, 1995; Yurke *et al.*, 1986] and *atomic spectroscopy* [Bollinger *et al.*, 1996; Wineland *et al.*, 1994, 1992]. What is more, these pioneering results have manifested the ability of reaching super-classical accuracy already by utilising measurement schemes such as photon-counting, and by employing the well-known *quantum-optical* techniques of *squeezing* applicable to both quantum states of atoms and light.

Ten years later, the topic of quantum metrology has been extensively revisited owing to the vividly developing field of *quantum information theory*. This time, the community has been stimulated to establish general frameworks allowing to predict ultimate, strategy-independent limits on the achievable parameter estimation precision, with particular interest in the role of *quantum entanglement* enhancing the accuracy of parameter inference [Giovannetti *et al.*, 2001, 2004, 2006]. As a result, the notions of the *Standard Quantum Limit* (SQL) and the *Heisenberg Limit* (HL) have been grounded, which generally

quantify the scaling of the estimated parameter *Mean Squared Error* (MSE) with the number of particles employed, N. In the case of uncorrelated particles, the $1/N$ *SQL-like scaling* of the MSE is a natural consequence of the classical notion of independent probing, or in other words the Central Limit Theorem (CLT). On the other hand, when quantum correlations in between the constituent particles are allowed, the $1/N^2$ *HL-like scaling* may be attained, which defines the ultimate quadratic enhancement determined abstractly by the structure of the quantum theory.

At the beginning of XXIst century, with the advent of complex experimental techniques allowing to control light and matter, a vast number of experiments have been conducted that demonstrated that the SQL can be indeed surpassed. These included the ones achieving super-classical *phase* resolution in *optical interferometry* [Kacprowicz *et al.*, 2010; Mitchell *et al.*, 2004; Nagata *et al.*, 2007; Okamoto *et al.*, 2008; Resch *et al.*, 2007; Xiang *et al.*, 2010] with a spectacular implementation in the *gravitational-wave detection* schemes [LIGO Collaboration, 2011, 2013]. Furthermore, the precision of estimation beyond SQL has also been attained in atomic-ensemble experiments in various configurations, in cases when one tries to most accurately: establish a time-duration reference – *atomic clocks* [Appel *et al.*, 2009; Leroux *et al.*, 2010; Louchet-Chauvet *et al.*, 2010; Ospelkaus *et al.*, 2011], register an atomic transition-frequency – *atomic spectroscopy* [Leibfried *et al.*, 2004; Roos *et al.*, 2006; Schmidt *et al.*, 2005], or sense an external magnetic field – *atomic magnetometry* [Koschorreck *et al.*, 2010; Napolitano *et al.*, 2011; Sewell *et al.*, 2012; Wasilewski *et al.*, 2010; Wolfgramm *et al.*, 2010].

1.4   Noisy quantum metrology

In all the above mentioned experiments, precision beyond the SQL could have been attained owing to the spectacular control of the quantum systems involved, what allowed to diminish the amount of *noise* inevitably present in the apparatus and nevertheless achieve the super-classical accuracy. Yet, the issue of robustness of the above setups against various sources of *decoherence* has soon been raised, in particular, questioning whether such experimental schemes could be utilised not only to surpass the SQL for a given number of particles (atoms, photons) employed, N, but also to observe a precision scaling that exceeds the classical $1/N$ dependence.

Such an issue, however, has long been left unanswered, as the original theoretical tools lacked the ability to efficiently incorporate the noise effects into the metrological models considered. In fact, the above problem initiated a new topic of research aiming to establish general techniques that would allow to quantify the precision attained in quantum estimation protocols, but would also explicitly account for the impact of noise either: by accurately describing the destructive processes of *decoherence* [Zurek, 2003] and *quantum fluctuations* [Gardiner and Zoller, 2000]; by adequately treating the overall ensemble of particles as an *open quantum system* [Breuer and Petruccione, 2002] constantly interacting with the environment; or by utilising the abstract formalism of *quantum channels* [Bengtsson and Życzkowski, 2006; Nielsen and Chuang, 2000] to most generally model the noisy evolution of a given quantum system.

In a first step, the impact of decoherence on quantum metrological protocols has been studied while considering particular experimentally motivated measurement schemes and models of noise [Banaszek *et al.*, 2009; Giovannetti *et al.*, 2011; Maccone and Giovannetti, 2011]. In *optical interferometry*, the primary attention has been paid to the effect of *photonic losses* that distorts the information about the phase being estimated [Demkowicz-Dobrzański *et al.*, 2009; Dorner *et al.*, 2009; Knysh *et al.*, 2011; Kołodyński and Demkowicz-Dobrzański, 2010], but also to the impact of *phase diffusion* collectively affecting all the photons present in the setup [Escher *et al.*, 2012; Genoni *et al.*, 2012, 2011]. In the *atomic experiments*, on the other hand, particular focus has been given to the *dephasing*-like noises

affecting the atoms in either correlated or uncorrelated manner [Andrè and Lukin, 2002; Andrè *et al.*, 2004; Borregaard and Sørensen, 2013; Chaves *et al.*, 2013; Dorner, 2012; Huelga *et al.*, 1997; Macieszczak, 2014; Macieszczak *et al.*, 2014; Shaji and Caves, 2007; Ulam-Orgikh and Kitagawa, 2001].

Nevertheless, the desired general frameworks have been soon proposed that are capable of assessing the performance of quantum metrological protocols in the presence of generic types of noise. These have been possible due to the abstract mathematical description employed, which utilises the language of quantum channels originally adapted to the metrological setting by Fujiwara and Imai [2008] and Matsumoto [2010]. The resulting general approaches quantifying the precision bounds in the presence of decoherence include: the *purification-based* methods – allowing to consider a particular purified version of the system without affecting its metrological properties [Escher *et al.*, 2011]; the *simulation-based* methods – allowing to simulate a given quantum channel representing the system evolution by means of either independently distributed classical random variables [Demkowicz-Dobrzański *et al.*, 2012] or uncorrelated quantum states [Kołodyński and Demkowicz-Dobrzański, 2013]; as well as the *channel-extension–based* methods that despite being numerical are always efficiently computable by means of *semi-definite programming* [Demkowicz-Dobrzański *et al.*, 2012; Kołodyński and Demkowicz-Dobrzański, 2013].

Importantly, the above techniques have demonstrated that an *infinitesimal* amount of generic noise, which independently affects each of the constituent particles of the system, is enough to force the asymptotic precision scaling with N to be SQL-like. Thus, the maximal quantum enhancement is then always limited to reach at most a *constant factor* improvement over the classical strategies. Moreover, the above results suggested that, due to the decoherence, the ultimate precision may always be attained in the asymptotic N regime without need of employing complex measurement schemes or preparing the system in exotic quantum states, but rather resorting to simple experimental techniques, e.g. the ones well-established in quantum optics such as the photon-counting or light-squeezing. Furthermore, the corresponding precision bounds, which obey the SQL-like scaling imposed by the noise, may be typically saturated in a single experimental trial (shot), what indicates the less prominent role of the inter-particle correlations [Jarzyna and Demkowicz-Dobrzański, 2013, 2014], and contrasts the noiseless scenario in which maximally correlated states of particles must be employed to attain the HL [Giovannetti *et al.*, 2006; Pezzé and Smerzi, 2009]. Abstractly, the effects of uncorrelated noise may be seen as providing the necessary reason for the *Quantum Local Asymptotic Normality* to hold [Guţă and Jenčová, 2007; Guţă and Kahn, 2006; Kahn and Guţă, 2009], so that the notion of the *Central Limit Theorem* (CLT) of statistics can then be naturally generalised to the quantum setting. In particular, due to the impact of uncorrelated noise, one may thus effectively represent in the asymptotic N limit any overall quantum state of the system by an "metrologically-equivalent" Gaussian state. On the other hand, as some of the precision-bounding techniques allow to assess the accuracy of estimation not only in the asymptotic limit of many particles [Kołodyński and Demkowicz-Dobrzański, 2013], they may be also utilised to verify the performance of quantum estimation protocols in the regime of finite N, in which it is generally not clear what type of measurements should be performed on a given system, and in what kind of states should it be prepared to achieve the optimal accuracy. Although they prove that for sufficiently low N one may disregard the impact of decoherence, and thus assume the optimal noiseless solution that generally employs highly non-classical states and measurements, it is an interesting open question how the structure of such optimal states and measurements varies with an increase in the number of particles—so that in the asymptotic N limit, dominated by the noise, they eventually take a much simpler form.

1.5 Outline of the thesis

The main purpose of this thesis is to explicitly describe such general techniques that allow to establish precision bounds in noisy quantum metrological schemes, both in the asymptotic limit of infinitely many particles [Demkowicz-Dobrzański *et al.*, 2012] and for the quantum systems of finite size [Kołodyński and Demkowicz-Dobrzański, 2013]. In particular, we demonstrate in this work how the *uncorrelated noise* modifies the effective evolution of each *single* constituent particle of the system, so that it may be effectively described by means of a *quantum channel* which structure is sufficient to determine the ultimate-precision bounds of interest.

In order to do so, we introduce the mathematical language necessary to describe the problems of noisy quantum metrology in Chap. 2, where we also present the typical quantum estimation schemes of metrological relevance—the *noisy-phase–estimation models* which we explicitly analyse throughout the work. In Chap. 3, we discuss in detail the tools of *classical* and *quantum estimation theory*, i.e. the *frequentist* and *Bayesian* approaches to parameter inference, in order to explicitly apply them to the example of *phase estimation* in *Mach-Zehnder interferometry* both in the classical and quantum setting.

In Chap. 4, we move onto the main results of this work and demonstrate methods which allow us to limit the ultimate precision in a general *N-parallel–channel* estimation protocol, stemming just from the structure of a *single* quantum channel describing the evolution of each constituent particle of the system. We revisit the Mach-Zehnder interferometer and the phase estimation scenario in Chap. 5 in order to analyse it in further detail, taking into account the impact of *photonic losses* both when following the frequentist and Bayesian approaches.

Finally, we summarise and conclude the discussion in Chap. 6, where we present some of interesting problems that remain unresolved within the topic of noisy quantum metrology, as well as indicate the relations to other methods being complementary to our results. We also discuss the potential generalisations of the established techniques to further metrologically relevant scenarios that are beyond the scope of this work.

Chapter 2

Metrology with realistic quantum systems

In the following chapter, we discuss the mathematical framework that is required in general to study the quantum metrological protocols, which we analyse in further parts of this work. In particular, we present the abstract language of quantum mechanics allowing to describe a given *quantum system* and importantly the *measurements* performed on the system to investigate its features. We further discuss the *evolution* of such a system, which may be equivalently described with help of *quantum channels* or the time-differential *master equation*, and give examples of models (i.e. the *noisy-phase–estimation* models analysed in various configurations throughout this work) that are of particular metrological relevance. We argue that quantum metrology should be viewed as a problem of determining the *latent parameters* of the system evolution, which—in contrast to the *observables* describing physical properties—are not directly measurable and correspond to the variables that parametrise the *model* assumed in the description. Thus, from maybe more philosophical perspective, they should *not* be associated with the properties of the system per se. In the final part of this chapter, we discuss in more detail the *geometrical picture of quantum channels*, which in the case of metrological problems not only describe the evolution, but also are importantly parametrised by the latent variable, φ, being *estimated*. In particular, we explain the notion of φ-*extremality* of a given channel that will be later shown to be one of the key ingredients that forbid the asymptotic HL-like precision scaling, $1/N^2$, with the number, N, of the constituent particles.

2.1   Mathematical description of a quantum system

As the description of the subtleties of the *quantum theory* is far beyond the scope of this work, we can only point the reader to common textbooks introducing quantum mechanics for explicit details [see e.g. Griffiths, 2013; Phillips, 2003], and instead summarise—maybe brutally and iconoclastically—the notion of quantum mechanics in a single sentence stating that:

> The *state* of a quantum system is described by its *wave function*, whereas its measurable physical properties correspond to *Hermitian operators* (*observables*), which respectively are represented by *vectors* and their adequate *linear transformations*, so that the natural language of quantum mechanics is just the *linear algebra*.

However, we would like to explicitly deal with quantum systems subjected to *noise*, which is manifested by the imperfect knowledge we possess about the system at a given time-instance. Hence, we must further generalise the above framework of state vectors to *density matrices*, or equivalently *density operators*, in order to naturally incorporate in the description the classical notion of *statistical ensembles*. In what follows, we introduce the necessary physical concepts and mathematical tools at a sufficient level from the quantum metrology point of view, yet their more detailed analysis may be found in [Bengtsson and Życzkowski, 2006; Nielsen and Chuang, 2000].

2.1.1 Quantum states

If a state of a given quantum system is perfectly known, it is described by a vector, i.e. a *pure state*: $|\psi\rangle \in \mathcal{H}$, living in the Hilbert space $\mathcal{H}$. On the other hand, if the system evolves according to some stochastic process, its state at a given moment must be represented by an *ensemble of pure states*: $\{p_i, |\psi_i\rangle\}_i$ indexed by i with p_i representing the probability of the system being in a particular state $|\psi_i\rangle$. As a result, one then effectively denotes the state of the system with help of a *density operator/matrix*:

$$\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \tag{2.1}$$

which mathematically corresponds to a linear operator on the Hilbert space $\mathcal{H}$ that is *non-negative* ($\forall_i : p_i \geq 0 \Rightarrow \forall_{|\phi\rangle \in \mathcal{H}} : \langle\phi|\varrho|\phi\rangle \geq 0$) and of *unit trace* ($\sum_i p_i = 1 \Rightarrow \mathrm{Tr}\{\varrho\} = 1$), what we formally denote as[1] $\varrho \in \mathcal{T}(\mathcal{H})$. In other words, Eq. (2.1) describes a *probabilistic mixture* of state vectors, and thus is normally termed to represent a *mixed state* of the system.

> **Note 2.1: Convexity of the space of density operators.**
> Notice that consistently a 'mixture of mixed states', $\sum_i p_i \varrho_i$, is also a mixed state of the form (2.1), as it similarly corresponds to an ensemble $\{p_i p_{i,j}, |\psi_{i,j}\rangle\}_{i,j}$, which is built after explicitly defining the ensembles for each ϱ_i as $\varrho_i = \sum_j p_{i,j} |\psi_{i,j}\rangle\langle\psi_{i,j}|$. More formally, this means that for any two mixed states $\varrho_{1/2}$ that belong to the space $\mathcal{T}(\mathcal{H})$, i.e. the space of density matrices defined on the Hilbert space $\mathcal{H}$, their convex sum $\varrho = \lambda\varrho_1 + (1-\lambda)\varrho_2$ with $0 \leq \lambda \leq 1$ also lies with within the space $\mathcal{T}(\mathcal{H})$. Geometrically, such a fact proves that the space of density operators is *convex*.

In order to study the quantum metrological scenarios, we must be able to correctly describe *composite systems*; in particular, quantum systems consisting of many particles. Therefore, we very briefly review their representation within the language of density matrices. Consider two distinct systems, labelled by A and B, which are isolated from one another (e.g. not in physical contact or space-like separated), so that they may be generally treated as separate ensembles $\{p_i^{A/B}, |\psi_i\rangle_{A/B}\}_i$ leading to $\varrho^{A/B}$ respectively. Hence, we may intuitively construct the ensemble representing the composite (bipartite) system AB as $\{p_i^A p_j^B, |\psi_i\rangle_A |\psi_j\rangle_B\}_{i,j}$, and thus the overall mixed state of the form (2.1):

$$\varrho^{AB} = \sum_{i,j} p_i^A p_j^B |\psi_i\rangle_A |\psi_j\rangle_B {}_A\langle\psi_i|{}_B\langle\psi_j| = \sum_i p_i^A |\psi_i\rangle_A\langle\psi_i| \otimes \sum_j p_j^B |\psi_j\rangle_B\langle\psi_j| = \varrho^A \otimes \varrho^B, \tag{2.2}$$

what proves that combining isolated quantum systems results in their *tensor-product* structure. As a consequence, any state describing a quantum system consisting of N particles that are *uncorrelated* between one another—treated as isolated subsystems—most generally reads: $\rho^N = \bigotimes_{n=1}^N \rho^{(n)}$, and in the case when each particle is prepared in an identical state, say ρ, further simplifies to $\rho^N = \rho^{\otimes N}$.

On the other hand, we may ask the inverse question; how to describe the subsystem A given the most general combined state: $\varrho^{AB} \in \mathcal{T}(\mathcal{H}_A \otimes \mathcal{H}_B)$, when we do not have access to the subsystem B (or equivalently

---

[1] A density matrix is thus an element of the Banach space of non-negative *trace-class operators* of trace 1, which we denote as $\mathcal{T}(\mathcal{H})$ [Reed and Simon, 1981].

vice versa). The state of the subsystem A corresponds then to the *reduced density matrix*, $\varrho^{\mathrm{A}} = \mathrm{Tr}_{\mathrm{B}}\{\varrho^{\mathrm{AB}}\}$, obtained after performing the *partial trace* operation over the system B, which is most conveniently defined after choosing any orthonormal basis $\{|i\rangle_{\mathrm{B}}\}_i$ in $\mathcal{H}_{\mathrm{B}}$ for which: $\varrho^{\mathrm{A}} = \sum_{i}{}_{\mathrm{B}}\langle i|\varrho^{\mathrm{AB}}|i\rangle_{\mathrm{B}}$. One may easily prove that such an interpretation is the *only one* consistent with the quantum mechanical description (introduced in the following section) of the measurements, which in such a situation are allowed to be performed *only* on the part of the system that is at our disposal [Nielsen and Chuang, 2000]. Let us emphasise that the important consequence of the lack of access to the whole system is the randomness introduced while discarding (partial-tracing) some of its constituents. Such a fact becomes most evident when considering any pure, and hence *deterministic*, state $|\psi\rangle_{\mathrm{AB}}$ of the composite system that is *not separable*, i.e. cannot be decomposed into a tensor product, $|\psi\rangle_{\mathrm{AB}} \neq |\psi_1\rangle_{\mathrm{A}} \otimes |\psi_2\rangle_{\mathrm{B}}$. It is so, as after tracing any of its subsystems one necessarily obtains a mixture (e.g. $\varrho^{\mathrm{A}} = \mathrm{Tr}_{\mathrm{B}}\{|\psi\rangle_{\mathrm{AB}}\langle\psi|\} \neq |\phi\rangle_{\mathrm{A}}\langle\phi|$ for A) and thus introduces *stochasticity* into the description. Yet, in case of composite systems possessing a (tensor-) product structure, by throwing away any of the subsystems we correctly do not affect the other ones, as most generally $\forall_{\varrho_{1/2}\in\mathcal{H}_{\mathrm{A/B}}} : \mathrm{Tr}_{\mathrm{B}}\{\varrho_1^{\mathrm{A}} \otimes \varrho_2^{\mathrm{B}}\} = \varrho_1^{\mathrm{A}}$, and similarly when tracing out over A. For example, when considering quantum systems that comprise of particles that are uncorrelated with one another, after losing some of them the overall state of the rest will be unaffected, despite the total number of the constituent particles being diminished. In contrast, when particles are correlated—in fact *entangled* (see Sec. 2.1.3 below)—by discarding any of them, we effectively introduce stochastic *noise* that "disturbs" the surviving ones. Yet, let us emphasise that such a "disturbance" should be understood as *blurring of the information* we possess about the system, and *not* as a process in which the particles are actually physically affected [Englert, 2013].

2.1.2 Quantum measurements

As mentioned above, the *observables* of a quantum system determine its physical properties and thus describe the quantities that may be directly measured. Formally, they correspond to linear *Hermitian*[2] operators acting on the Hilbert space, i.e.[3] $\hat{O} \in \mathcal{B}(\mathcal{H})$ such that $\hat{O} = \hat{O}^\dagger$, what means that they can always be expressed in the complete basis of projectors[4]: $\{P_i\}_i$ satisfying $P_i P_j = P_i \delta_{ij}$ and $\sum_i P_i = \mathbb{I}$, such that $\hat{O} = \sum_i o_i P_i$. Importantly, the real eigenvalues of $\hat{O}$, o_i, determine the possible values taken by the random variable O that describes the outcomes obtained in a measurement of the observable. In particular, for a given state ϱ of the system, in "each shot" one of the o_i-s is measured with probability $p_i = \mathrm{Tr}\{\varrho\, P_i\}$, so that on average $\langle O \rangle = \sum_i p_i o_i = \sum_i \mathrm{Tr}\{\varrho\, P_i\} o_i = \mathrm{Tr}\{\varrho\, \hat{O}\}$ is obtained. Hence, such an observable-based *projective measurement* is fully described by the operators $\{P_i\}_i$, each leading to an outcome labelled by i occurring with probability p_i.

Let us note, however, that so-defined projective measurements are highly over-idealised, as by assuming all P_i to be orthogonal with one another, we really demand "by hand" that we are able to accurately measure the observable of interest without heavily disturbing the system. In such a peculiar setting, quantum mechanics predicts the system to be left in the relevant eigenstate of $\hat{O}$ after the measurement, what has maybe confusingly been termed in the literature as a phenomenon of the "wave-function collapse" suggesting some dramatic dynamical process [Englert, 2013]. One should always bear in mind,

---

[2]Strictly speaking, observables correspond to *self-adjoint* operators, which apart from Hermiticity also demand the domains of $\hat{O}$ and $\hat{O}^\dagger$ to coincide [Reed and Simon, 1981]. However, such an extra condition is naturally fulfilled by any *bounded* operator—hence, the notation $\mathcal{B}(\mathcal{H})$ above—so that such mathematical subtleties may only play a role when dealing with infinite dimensional Hilbert spaces (e.g. for position or momentum observables) [Hall, 2013].

[3]In this work, we primarily reserve the notation of a 'hat', i.e. $\hat{O}$, only to operators that correspond to *observables*. Yet, without loss of clarity, we also employ it in the final chapters to denote bosonic operators, while working within the second-quantisation formalism utilised in the description of optical interferometry setups.

[4]If the eigenvalues of $\hat{O} = \sum_i o_i |i\rangle\langle i|$ are non-degenerate, P_i just correspond to projections onto the eigenvectors: $P_i = |i\rangle\langle i|$.

that the quantum-state description should be just treated as a bookkeeping tool for the knowledge we possess about the system. In reality, in order to precisely measure the observable, we must *strongly* interact with the system. This, however, should not be seen as a problem, as we may still perfectly register one of the eigenvalues o_i while disrupting the system or even destroying it (e.g. consider optical experiments in which all the detection configurations yielding various outcomes always lead to the same final state of the system—the vacuum—after absorbing all the photons) [Steinberg, 2014].

Furthermore, as in metrological problems we just seek measurement schemes that lead to probability distributions from which the encoded evolution parameter may be most accurately inferred, the post-measurement state of the system is essentially of no interest. Hence, it is most appropriate to consider the most general formalism of the so-called *Postive Operator Valued Measure* elements (POVMs) [Bengtsson and Życzkowski, 2006; Nielsen and Chuang, 2000], which is designed to model all potential measurement-outcome statistics at the price of being ambiguous in determining post-measurement state of the system[5]. A POVM is generally represented by any set of *quantum measurement operators* acting on the system Hilbert space, $\{M_i\}_i$, that are *non-negative* $M_i \geq 0$ and satisfy the *completeness constraint* $\sum_i M_i = \mathbb{I}$. Each outcome i of the measurement is then obtained with probability $p_i = \text{Tr}\{\varrho M_i\}$, yet (due to the ambiguity of the measurement-action on the state[5]) there exists an infinite number of experimental apparatuses that yield the measurement statistics of a particular POVM. Thus, although the language of POVMs is widely used and spectacularly successful in quantum information theory—in particular, in the optimisation of protocols over measurement strategies—it does not in principle provide a recipe how to construct its given implementation, which must be thus independently determined bearing in mind the experimental context considered. Let us note that nothing prevents us to also consider POVMs that are continuously parametrised by a random variable X, and thus consist of an infinite number of non-negative elements. Such $\{M_x\}_{X=x}$ yield then the outcome *Probability Distribution Function* (PDF): $p(x) = \text{Tr}\{\varrho M_x\}$, with the completeness constraint now reading: $\int \mathrm{d}x\, M_x = \mathbb{I}$, so that consistently $\int \mathrm{d}x\, p(x) = 1$.

2.1.3   Entanglement of subsystems

One of the most surprising phenomena that quantum theory has to offer is its "spooky" feature of *entanglement*, recognised already at the beginning of XX century in the seminal papers of Einstein *et al.* [1935]; von Neumann [1932] and Schrödinger [1935]. Although it has been long considered rather as a peculiarity of quantum mechanics, after the advent of rapidly developing field of quantum information theory and modern experimental techniques, it has been explicitly shown to be a genuine, tangible *resource* being at the heart of real-life quantum-based protocols in quantum: cryptography, communication, dense coding, teleportation and more [see Horodecki *et al.*, 2009, for a review].

Importantly, from the quantum metrology perspective, it is exactly the presence of entanglement—or equivalently the *non-classical correlations* in between the constituent particles of the system—that allows to surpass the standard limits imposed by classical statistics on the precision of parameter inference [Giovannetti *et al.*, 2001, 2004, 2006]. That is why, we review the basic concepts of this phenomenon (with focus on systems consisting of many, potentially indistinguishable particles), so that our discussions of the quantum-enhanced—or really 'entanglement-enhanced'—metrological protocols in the further parts of this work may be clearer to the reader. Nevertheless, let us remark that a comprehensive explanation of the role of entanglement in quantum metrology has not yet been fully established, being a vivid problem of current research [Hyllus *et al.*, 2012; Pezzé and Smerzi, 2009; Tóth, 2012]. Although the presence

---

[5]For each POVM element M_i, one may construct an infinite number of *generalised measurement operators* E_i satisfying $M_i = E_i^\dagger E_i$, each of which leads to a distinct post-measurement state of the system: $\varrho \rightarrow \frac{E_i \varrho E_i^\dagger}{\text{Tr}\{E_i \varrho E_i^\dagger\}}$.

of entanglement is *necessary* for a super-classical precision to be observed in a metrological scenario, there also exist highly particle-entangled states yielding no enhancement [Hyllus *et al.*, 2010a]. Yet, in the absence of any noise, one may establish a link between the so-called *N-parallel–channel* schemes (discussed later in Chap. 4) that employ *N* 'maximally entangled' particles with each individually sensing the parameter of interest, to the *sequential* protocols in which a *single* particle just senses the parameter *N*-times in a row [Berry *et al.*, 2009; Maccone, 2013]. Such an observation suggests that the *N*-particle entangled states are beneficial, as they effectively simulate an *N*-fold amplification of the parameter sensitivity (see also Sec. 3.2.4). Such an interpretation, however, generally fails to be valid in the presence of noise [Maccone, 2013], in which parallel strategies employing entangled particles seem to be more beneficial [Demkowicz-Dobrzański and Maccone, 2014]. On the other hand, it has been very recently shown that (very noisy) states exhibiting *bound-entanglement*—considered to be the weakest form entanglement possessing limited use in quantum protocols (not being distillable) [Horodecki *et al.*, 2009]—can also lead to the ultimate HL-like, $1/N^2$, precision scaling [Czekaj *et al.*, 2014], what implies that even very cumbersome types of inter-particle correlations may be *witnessed* [Horodecki *et al.*, 2009] by inspecting the enhancement of precision in an adequate metrological setting.

In general, a bipartite state consisting of subsystems A and B is called *separable*, if and only if it may be written as a mixture of (tensor-)product states of the subsystems, i.e.

$$\varrho^{\mathrm{AB}} = \sum_i p_i \, \varrho_i^{\mathrm{A}} \otimes \varrho_i^{\mathrm{B}}. \tag{2.3}$$

On the other hand, ϱ^{AB} is said to be *entangled* if it is *not* separable, i.e. cannot be written in the above form. Notice that, due to the overall probabilistic distribution $\{p_i\}_i$, a separable state (2.3) generally exhibits *classical*[6] *correlations* in between A and B, so that the subsystems may be assumed to be uncorrelated only in the case of Eq. (2.2). Crucially, the above definition of entanglement requires the notion of a *division into distinct parties* w.r.t. which the convex combination of product states (2.3) may be defined. Thus, for more than two constituent subsystems a given state may be considered entangled depending on the division chosen.

Note 2.2: Entanglement of a three-qubit state.
For instance, consider a joint state of three qubits: $\varrho^{\mathrm{ABC}} = \frac{1}{2} \sum_{i,j=0}^{1} |i\rangle_{\mathrm{A}} \langle j| \otimes |i\rangle_{\mathrm{B}} \langle j| \otimes |0\rangle_{\mathrm{C}} \langle 0|$ which is separable w.r.t. the AB|C cut, but entangled w.r.t. the A|BC cut. Interestingly, making things even more complicated, one may find three-qubit states, ϱ^{ABC}, that are separable w.r.t. *all* three bipartite cuts (AB|C, A|BC, AC|B), but *not* w.r.t. the tripartite one (A|B|C) [Acín *et al.*, 2001]. Such states are in fact *bound-entangled* [Horodecki *et al.*, 2009] and naturally constructable with use of the formalism of the so-called *unextendible product bases* [Bennett *et al.*, 1999a,b].

Nevertheless, we may generalise the notion of separability (2.3) to composite systems consisting of many particles, and define an *N*-particle quantum state to be *fully separable*[6] if and only if it may be written in the form:

$$\rho^N = \sum_i p_i \bigotimes_{n=1}^{N} \rho_i^{(n)} = \sum_i p_i \, \rho_i^{(1)} \otimes \rho_i^{(2)} \otimes \cdots \otimes \rho_i^{(N)}, \tag{2.4}$$

so that it may be confidently termed *not* to contain *any inter-particle entanglement*, being separable w.r.t. to *all* the possible cuts—particle groupings. In case we restrict only to pure states, the above full-separability condition simplifies to:

$$|\psi^N\rangle = \bigotimes_{n=1}^{N} |\psi^{(n)}\rangle = |\psi^{(1)}\rangle \otimes |\psi^{(2)}\rangle \otimes \cdots \otimes |\psi^{(N)}\rangle, \tag{2.5}$$

---

[6]Let us note that states (2.3)/(2.4), despite being separable, may still possess *non-classical correlations* typically quantified with use of the so-called *quantum discord* [Streltsov, 2015], which analysis is beyond the scope of this work.

showing that the particles must then be in a product state. Let us emphasise that in order to apply the definitions (2.4)/(2.5) and verify the inter-particle entanglement of a given state ρ^N, we must be eligible to make the statement "the n-th particle", as the notion of entanglement requires the subsystems to be unambiguously defined. Hence, the decompositions (2.4)/(2.5) are natural when dealing with systems consisting of *distinguishable* particles (e.g. photons prepared in distinct time-bins, or atoms resident in different optical-lattice sites), which in principle may thus be individually targeted. However, we can also utilise Eqs. (2.4) and (2.5) when analysing systems consisting of a *definite*[7] number of *indistinguishable bosonic*[8] particles. Yet, in order then to verify the inter-particle entanglement, we cannot use their natural description within the *second-quantisation* formalism [Altland and Simons, 2010; Schwabl, 2008; Scully and Zubairy, 1997], but rather must return to the *first-quantisation* picture [Killoran *et al.*, 2014], in which the overall Hilbert space is unambiguously divided into a product of the individual-particle Hilbert spaces, so that conditions (2.4)/(2.5) apply. We discuss this issue in more detail below, but let us also remark that, although such a procedure is appropriate from the point of view of the metrological protocols [Demkowicz-Dobrzański *et al.*, 2015]—that (as later discussed) rely on a well-defined notion of distinct particles which number quantifies the resources—other approaches to quantify entanglement of indentical-particle systems are also possible [Benatti *et al.*, 2014; Shi, 2003; Stockton *et al.*, 2003; Wiseman and Vaccaro, 2003].

2.1.4 Entanglement of indistinguishable particles

Consider a general pure state consisting of N particles in two, *bosonic* modes labelled by a and b [Scully and Zubairy, 1997]:

$$\left|\psi_{\mathrm{bos}}^{N}\right\rangle = \sum_{n=0}^{N} \alpha_n \left|n\right\rangle_a \left|N-n\right\rangle_b = \sum_{n=0}^{N} \alpha_n \left|n, N-n\right\rangle, \tag{2.6}$$

written in the basis of vectors: $\{|n, N-n\rangle\}_{n=0}^{N}$, each representing $(n, N-n)$ particles that occupy modes (a,b) respectively. Hence, $|\psi_{\mathrm{bos}}^{N}\rangle \in \mathcal{H}_{\mathrm{bos}}$, where $\mathcal{H}_{\mathrm{bos}}$ is the bosonic subspace of the Hilbert space with $\dim\{\mathcal{H}_{\mathrm{bos}}\} = N+1$. Now, if we want to return to the first-quantisation picture and treat particles as separate subsystems, we must associate a 2-dimensional Hilbert space with each of them, so that each corresponds to a *qubit* with basis vectors $\{|0(\equiv a)\rangle, |1(\equiv b)\rangle\}$ representing the particle being in either of the modes. As a result, we are able to write every $|n, N-n\rangle$ vector in the form: $|0\rangle^{\otimes n} \otimes |1\rangle^{\otimes N-n}$, which, however, must be *symmetrised* over all particle permutations. Thus, adopting a binary notation in which an N-bit sequence $\boldsymbol{n}$ is utilised to represent a product state of N *distinguishable* qubits: $|\boldsymbol{n}\rangle = |n_1\rangle \otimes \cdots \otimes |n_N\rangle$ with $|n_i\rangle \in \{|0\rangle, |1\rangle\}$, we can rewrite Eq. (2.6) as

$$
\begin{aligned}
\left|\psi_{\mathrm{bos}}^{N}\right\rangle &= \sum_{n=0}^{N} \alpha_n \frac{1}{\sqrt{|\Pi|}} \sum_{\Pi} |\Pi[\underbrace{0,\ldots,0}_{n}, \underbrace{1,\ldots,1}_{N-n}]\rangle = \sum_{n=0}^{N} \alpha_n \frac{1}{\sqrt{\binom{N}{n}}} \sum_{\substack{\boldsymbol{n}=\boldsymbol{0}^N \\ |\boldsymbol{n}|=n}}^{\boldsymbol{1}^N} |\boldsymbol{n}\rangle \\
&= \sum_{\boldsymbol{n}=\boldsymbol{0}^N}^{\boldsymbol{1}^N} \left(\sum_{n=0}^{N} \frac{\alpha_n \, \delta_{|\boldsymbol{n}|,n}}{\sqrt{\binom{N}{n}}} \right) |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_N\rangle .
\end{aligned} \tag{2.7}
$$

We denote by $\sum_{\Pi}$ the sum over all permutations, Π, which is also equivalently represented above in the binary notation by fixing the number of 1s ($n = |\boldsymbol{n}|$) appearing in a bit sequence. We have explicitly written out the product state in Eq. (2.7) to emphasise that indeed the notion of the "n-th particle"

---

[7]In fact (see Chap. 5), we may also return to the first-quantisation picture in case of *indefinite* number of particles when we do *not* possess a *global phase reference* [Bartlett *et al.*, 2007; Jarzyna and Demkowicz-Dobrzański, 2012; Mølmer, 1997], as we may then independently consider each N-particle sector [Vaccaro *et al.*, 2003].

[8]Similarly for *fermionic* systems which, however, we do not consider within this work.

is now clearly defined, so that the full-separability condition (2.5) directly applies. In fact, Eq. (2.7) indicates that Eq. (2.5) can only be satisfied, if Eq. (2.7) may be rewritten as a tensor product $|\xi\rangle^{\otimes N}$ with $|\xi\rangle = \alpha|a\rangle + \beta|b\rangle$ being an arbitrary pure qubit state.

For comparison, notice that the most general pure state of N *distinguishable* particles in two modes, or equivalently N qubits, is supported by a vastly larger—2^N-dimensional—Hilbert space and reads:

$$|\psi^N\rangle = \sum_{\mathbf{n}=\mathbf{0}^N}^{\mathbf{1}^N} \alpha_{\boldsymbol{n}} \, |\boldsymbol{n}\rangle = \sum_{j=\{0,\frac{1}{2}\}}^{\frac{N}{2}} \sum_{m=-j}^{j} \tilde{\alpha}_{j,m} \, |j,m\rangle, \qquad (2.8)$$

so that the bosonic state (2.7) may be interpreted as a special instance of $|\psi^N\rangle$ after matching the coefficients $\alpha_{\boldsymbol{n}}$ with the ones appearing in the curly brackets in Eq. (2.7). In order to make such a statement even clearer, we have written $|\psi^N\rangle$ in Eq. (2.8) also in its *angular-momentum* representation [Biedenharn and Louck, 1981; Devanathan, 1999], in which the quantum numbers (j,m) respectively represent the eigenvalues of the total angular-momentum, $\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$, and the $\hat{J}_z = \frac{1}{2} \sum_{n=1}^{N} \hat{\sigma}_z^{(n)}$ operators[9]. As a consequence, the bosonic (fully symmetric) subspace, $\mathcal{H}_{\text{bos}}$, may then be associated with the one corresponding to the maximal total angular-momentum, $j = N/2$, for which Eqs. (2.6) and (2.8) become equivalent after identifying all $|n, N-n\rangle$ vectors with the so-called Dickes states [Dicke, 1954]: $|j = \frac{N}{2}, m = n - \frac{N}{2}\rangle$, and thus α_n with $\tilde{\alpha}_{\frac{N}{2}, n - \frac{N}{2}}$.

Note 2.3: Inter-particle entanglement of two-mode bosonic states.

Let us consider two simple examples of N-particle two-mode bosonic states: a *Fock state* resident in one of the modes $- |N\rangle_a |0\rangle_b$, and a *twin-Fock state*[a] $- |\frac{N}{2}\rangle_a |\frac{N}{2}\rangle_b$; and investigate whether they contain any inter-particle entanglement. In case of a Fock state, the situation is completely straightforward, as it just corresponds to N photons contained in mode a, i.e. $|N\rangle_a |0\rangle_b = |a\rangle^{\otimes N}$, that by definition are uncorrelated with one another. Yet, such an obvious observation indicates that we need the indistinguishable particles to be somehow distributed between the two modes for any inter-particle entanglement to be present. On the other hand, in case of the twin-Fock state, as only $\alpha_{n=\frac{N}{2}} = 1$ is non-zero in Eq. (2.6), we obtain according to Eq. (2.7):

$$\left|\frac{N}{2}\right\rangle_a \left|\frac{N}{2}\right\rangle_b = \frac{(N/2)!}{\sqrt{N!}} \sum_{\Pi} \Pi\left[|a\rangle^{\otimes \frac{N}{2}} |b\rangle^{\otimes \frac{N}{2}}\right], \qquad (2.9)$$

where again $\sum_{\Pi}$ stands for a sum over permutations. Importantly, as Eq. (2.9) cannot be rewritten into a product state of particles (2.5), it proves that a twin-Fock state contains inter-particle entanglement[b]. As a consequence, in contrast to the Fock state (representing the *classical stategy* in a phase-estimation protocol discussed in Sec. 3.1.4), $|\frac{N}{2}\rangle_a |\frac{N}{2}\rangle_b$ may be utilised to achieve quantum enhancement in metrological protocols [Holland and Burnett, 1993].

[a]Assuming without loss of generality N to be even.

[b]Actually, the states (2.6) that do *not* possess any inter-particle entanglement are *only* the ones that can be obtained by imping Fock and vacuum states respectively on the input ports of a general beam-splitter (see also Sec. 3.2.4).

On the other hand, the first-quantisation–representation (2.7) of $|\psi_{\text{bos}}^N\rangle$ allows to extract from a general transformation of a bosonic N-particle system, $|\psi_{\text{bos}}^N\rangle \to \rho^N$, the mathematical form of the evolution (see next section) of each constituent particle. This may be achieved by re-expressing also ρ^N with help of Eq. (2.7) and identifying the effective map according to which each *single* particle evolves. Importantly, the properties of such a map will be shown later within this work to play a crucial role in quantum metrology, in principle determining the maximal capabilities of a given metrological protocol.

Moreover, let us also remark that, when considering *noisy* quantum systems, the noise may often introduce extra degrees of freedom to the particles, opening thus doors to their potential distinguishability.

---

[9]Throughout this work we adopt the standard notation $\hat{\sigma}_i$ with $i = \{x, y, z\}$ for Pauli spin-1/2 operators.

Mathematically, this means that the system is *not* supported by the bosonic subspace during the evolution, as it is "taken out" from $\mathcal{H}_{\mathrm{bos}}$ by the action of the noise. For instance, consider N-photons prepared in a state (2.6), $|\psi_{\mathrm{bos}}^N\rangle$, representing two modes of light impinged on the input ports of a beam-splitter. However, due to the imperfect (spatial/temporal) mode-matching of the beams, some of the photons do not contribute to the splitting process and lead to a non-coherent admixture additionally present in the outputted light-modes (e.g. yielding imperfect visibility of an interferometer) [Demkowicz-Dobrzański *et al.*, 2015]. Within the second-quantisation formalism, the non-contributing photons should be assumed to occupy extra distinct modes, which, however, are mistaken for the output ones at future detection stages. Thus, the effective state describing all the outputted photons restricted to the two original modes is *not* pure any more, despite still being permutation-invariant[10]. Furthermore, it is no longer supported only by the $j = N/2$ subspace in the angular-momentum representation, and hence extends beyond the bosonic subspace. As a result, the effective noise model does not preserve the *bosonicity* of particles (cannot be described by means of just creation and annihilation operators of the output modes [Altland and Simons, 2010; Schwabl, 2008; Scully and Zubairy, 1997]), and in fact requires the first-quantisation picture of Eq. (2.7) to be employed.

2.2 Quantum system dynamics

In quantum metrological problems, the parameter of interest to be estimated from the measurements performed on a given system is *encoded* during the system *evolution*, so that—apart from the formalism describing the state of the system at some time-instance—we must discuss more formally how the system evolves in time. Let us remind the reader that according to quantum mechanics the evolution of any state vector $|\psi(t)\rangle$ is governed by the Schrödinger equation [Griffiths, 2013; Phillips, 2003], which for a *closed* isolated system reads[11]:

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle \quad \Longrightarrow \quad \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\varrho(t) = \left[\hat{H}, \varrho(t)\right], \tag{2.10}$$

and, as shown above, naturally generalises to the so-called von Neumann equation [Breuer and Petruccione, 2002] when considering the density matrix representation (2.1) with $\varrho(t) = \sum_i \lambda_i(t)|\psi_i(t)\rangle\langle\psi_i(t)|$. As a result, a closed isolated system evolves in between times t_0 and t under a *unitary* transformation $U_\tau = \exp\left[-\mathrm{i}\hat{H}\tau\right]$, so that the final state of the system reads: $\varrho(t) = U_{t-t_0}\varrho(t_0)U_{t-t_0}^\dagger$.

Now, as we would like to describe the evolution of noisy *open* systems, we must account for the degrees of freedom that are not under our control and lead to the effect of *decoherence* [Zurek, 2003], making the overall evolution non-unitary. In such a case, Eq. (2.10) still applies describing the evolution of the state $\varrho^{\mathrm{SE}}(t) \in \mathcal{H}_{\mathrm{S}} \otimes \mathcal{H}_{\mathrm{E}}$ containing the system (S) of interest, but *also* the environment (E) which is beyond our reach. However, we must ensure that at a given time instance $t_0 = 0$, from which we would like to describe the evolution and at which we importantly have full control of the system knowing its state $\varrho^{\mathrm{S}}(0) \in \mathcal{H}_{\mathrm{S}}$, the environment is isolated, i.e. $\varrho^{\mathrm{SE}}(0) = \varrho^{\mathrm{S}}(0) \otimes \varrho^{\mathrm{E}}$, making the overall process physical. Then, after the environment comes into contact with the system it introduces noise, as the final state $\varrho^{\mathrm{SE}}(t)$ must be traced-out (see Sec. 2.1.1) over the subspace $\mathcal{H}_{\mathrm{E}}$ representing degrees of freedom we do not have access

---

[10]The fact that all permutation-invariant states are bosonic is true for *pure*, but *not* for *mixed* states. All permutation-invariant mixed states may be written in a block-diagonal form: $\rho_{\mathrm{p.i.}}^N = \bigoplus_{j=0}^{N/2} \rho_j \otimes (\mathbb{I}_{d_j}/d_j)$, where each subspace is parametrised by the total angular-momentum number j and d_j represents the dimension of each "multiplicity" space being proportional to $\mathbb{I}$ [Bartlett *et al.*, 2007]. The *bosonic mixed states* may lie only in the block corresponding to $j = N/2$.

[11]Without loss of generality, throughout this work we assume the Planck's constant $\hbar$ to be equal to 1.

to. As a result, after defining $U_\tau^{\text{SE}} = \exp[-i\hat{H}_{\text{SE}}\tau]$, we may most generally write:

$$\varrho^{\text{S}}(t) = \text{Tr}_{\text{E}}\left\{U_t^{\text{SE}}\left(\varrho^{\text{S}}(0) \otimes \varrho^{\text{E}}\right)U_t^{\text{SE}\dagger}\right\} = \Lambda_t\left[\varrho^{\text{S}}(0)\right] \quad \Longleftrightarrow \quad i\frac{\text{d}}{\text{d}t}\varrho^{\text{S}}(t) = \text{Tr}_{\text{E}}\left\{\left[\hat{H}_{\text{SE}}, \varrho^{\text{SE}}(t)\right]\right\}, \quad (2.11)$$

where Λ_t is thus the effective *quantum channel* describing the evolution of $\varrho^{\text{S}}(0)$ to $\varrho^{\text{S}}(t)$. Notice already that $\Lambda_t : \mathcal{T}(\mathcal{H}_{\text{S}}) \to \mathcal{T}(\mathcal{H}_{\text{S}})$ must be a linear map, as it is constructed by combining linear operations: a unitary rotation followed by the partial trace of the environmental subspace $\mathcal{H}_{\text{E}}$.

As from the perspective of quantum metrological protocols, we will be primarily interested (see Sec. 3.2 later) in the overall transformation of the system from its 'input' state at $t_0 = 0$ onto its 'output' state at t for a *fixed* evolution duration, we drop in the next section the time-dependence of Λ_t, in order to discuss the general structure of *quantum channels—Completely Positive Trace-Preserving* (CPTP) maps—describing all possible physical 'input-output transformations' of the system. However, such an approach is not appropriate when considering metrological schemes (e.g. the frequency estimation scenarios studied later in Sec. 4.4) in which the time duration t is an extra degree of freedom that may be adjusted to purposefully vary the form of the effective quantum channel Λ_t. Thus, we also review the general description of an open system via the *master equation*, i.e. the differential equation on the r.h.s. of Eq. (2.11) (subject to the initial condition $\varrho^{\text{SE}}(0) = \varrho^{\text{S}}(0) \otimes \varrho^{\text{E}}$), which often is the one that phenomenologically specifies the dynamics of a given quantum system.

2.2.1 *Quantum channel* picture of system evolution

Let us consider a general *quantum channel* defined via Eq. (2.11), i.e. $\Lambda : \mathcal{T}(\mathcal{H}_{\text{in}}) \to \mathcal{T}(\mathcal{H}_{\text{out}})$ [Bruß and Leuchs, 2007], which describes the system evolution from a given *input* state $\varrho_{\text{in}} \in \mathcal{T}(\mathcal{H}_{\text{in}})$ (previously $\varrho^{\text{S}}(0)$) to the adequate *output* state $\varrho_{\text{out}} = \Lambda[\varrho_{\text{in}}] \in \mathcal{T}(\mathcal{H}_{\text{out}})$ (previously $\varrho^{\text{S}}(t)$). For generality, let us also assume that the system dimensions may vary during the evolution, i.e. $d_{\text{in}} \neq d_{\text{out}}$ with $d_{\text{in/out}} = \dim\{\mathcal{H}_{\text{in/out}}\}$, what is still in agreement with Eq. (2.11), in which the dimension of the traced-out environment E changes between t_0 and t. Physically, this corresponds to the situation in which the noise either destroys or introduces new degrees of freedom to the system. For instance, one may consider atoms which are perturbed by the interaction with an environment in a way, so that respectively either some of their energetic transitions become forbidden by the presence of decoherence, or they are excited into energetic levels that originally would not have been allowed while considering only the atomic free-evolution.

Complete Positivity (CP) and other properties of quantum channels

Firstly, one should impose that the probability has to be conserved, so that $\text{Tr}\{\varrho_{\text{in}}\} = \text{Tr}\{\varrho_{\text{out}}\} = 1$ and thus any quantum map Λ must be *Trace-Preserving* (TP). On the other hand, as *none* of the eigenvalues (i.e. mixing probabilities in Eq. (2.1)) of the output state can be negative regardless of ϱ_{in} chosen, any quantum channel must also be *positive*, i.e. $\forall_{\varrho_{\text{in}} \in \mathcal{T}(\mathcal{H}_{\text{in}})} : \Lambda[\varrho_{\text{in}}] \geq 0$.

However, this is not the end of the story, as for a given quantum channel Λ to be physical it must also describe a valid evolution of a system (S) being in contact with some ancilla (A), and thus constituting a part of an overall system+ancilla (SA) composite system. This is true only if for any initial bipartite state ($\varrho^{\text{SA}} \in \mathcal{T}(\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{A}})$ with arbitrary $d_{\text{A}} = \dim\{\mathcal{H}_{\text{A}}\}$) the overall density matrix of the system SA remains positive after the action of Λ on S. Hence, this means that any quantum channel must also be *Completely Positive* (CP), i.e.

$$\forall_{\varrho^{\text{SA}} \in \mathcal{T}(\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{A}})} : \quad \Lambda \otimes \mathcal{I}\left[\varrho^{\text{SA}}\right] \geq 0, \quad (2.12)$$

where $\mathcal{I}$ denotes the identity map that trivially acts on the subspace $\mathcal{H}_\mathrm{A}$ without affecting it. Notice that, if we restricted the bipartite inputs in the CP-definition (2.12) to be separable (see Eq. (2.3)), Eq. (2.12) would be trivially satisfied due to the positivity property of Λ. Thus, it is really the ability of considering entangled states between the system and the ancilla, which strengthens the notion of positivity of a quantum channel to its complete positivity. Let us also note that, due to the linearity of Λ, we could have equivalently restricted ourselves in the CP-definition (2.12) only to pure states $|\Psi\rangle_\mathrm{SA} \in \mathcal{H}_\mathrm{in} \otimes \mathcal{H}_\mathrm{A}$. As a result, one may directly see that it is sufficient to consider only $d_\mathrm{A} = d_\mathrm{in}$ in order to ensure the CP property of Λ, as any bipartite pure state $|\Psi\rangle_\mathrm{SA}$ may always be expressed by utilising its *Schmidt decomposition* as $|\Psi\rangle_\mathrm{SA} = \sum_{i=1}^{\min\{d_\mathrm{in},d_\mathrm{A}\}} \lambda_i |\psi_i\rangle_\mathrm{S} |\phi_i\rangle_\mathrm{A}$ with some $\lambda_i \geq 0$ and orthonormal $|\psi_i\rangle \in \mathcal{H}_\mathrm{in}$, $|\phi_i\rangle \in \mathcal{H}_\mathrm{A}$ [Bengtsson and Życzkowski, 2006; Nielsen and Chuang, 2000].

Summarising, any physical quantum channel corresponds to a *Completely Positive Trace-Preserving* (CPTP) map that is defined to satisfy the above conditions. We return once more to the analysis of CPTP maps in Sec. 2.4, where we further parametrise them w.r.t. the estimated parameter of interest in quantum metrological protocols and discuss their *geometrical* properties.

Kraus representation of a quantum channel and the *Stinespring dilation theorem*

Furthermore, the construction of Eq. (2.11) allows to naturally introduce the so-called *Kraus representation* [Kraus, 1983] of a given CPTP map $\Lambda : \mathcal{T}(\mathcal{H}_\mathrm{in}) \to \mathcal{T}(\mathcal{H}_\mathrm{out})$. Notice that without loss of generality we may enlarge the subspace of $\mathcal{H}_\mathrm{E} \to \mathcal{H}_\mathrm{E} \otimes \mathcal{H}_{\mathrm{E}'}$, and *purify* ϱ^E to $|\xi\rangle_{\mathrm{EE}'}$ such that $\varrho^\mathrm{E} = \mathrm{Tr}_{\mathrm{E}'}\{|\xi\rangle_{\mathrm{EE}'}\langle\xi|\}$. Then, we may always rewrite Eq. (2.11) after relabelling the extended environment as the "new effective" environment, so that $\mathrm{EE}' \to \mathrm{E}$ and[12] $\mathcal{U}^\mathrm{SE} \otimes \mathcal{I}^{\mathrm{E}'} \to \mathcal{U}^\mathrm{SE}$:

$$\varrho_\mathrm{out} = \Lambda[\varrho_\mathrm{in}] = \mathrm{Tr}_\mathrm{E}\left\{U^\mathrm{SE}\left(\varrho_\mathrm{in} \otimes |\xi\rangle_\mathrm{E}\langle\xi|\right)U^{\mathrm{SE}\dagger}\right\} = \sum_{i=1}^d K_i \varrho_\mathrm{in} K_i^\dagger, \qquad (2.13)$$

where $K_i = {}_\mathrm{E}\langle i|U^\mathrm{SE}|\xi\rangle_\mathrm{E}$ $(K_i : \mathcal{H}_\mathrm{in} \to \mathcal{H}_\mathrm{out})$ are the so-called *Kraus operators* representing the action of Λ. Note that $\{K_i\}_i$ are ambiguously defined after choosing any set of vectors, $\{|i\rangle_\mathrm{E}\}_{i=1}^d$, spanning the enlarged environmental subspace $\mathcal{H}_\mathrm{E}$, such that $\sum_{i=1}^d |i\rangle_\mathrm{E}\langle i| = \mathbb{I}^\mathrm{E}$. Thus, necessarily $d \geq \dim\{\mathcal{H}_\mathrm{E}\}$, but the *minimal* possible dimension of $\mathcal{H}_\mathrm{E}$ which can be chosen to mimic the action of Λ for any ϱ_in defines the so-called *rank*, r, of the quantum channel. In general (see also Sec. 2.4.1), $1 \leq r \leq d_\mathrm{in}^2$ and $r = 1$ corresponds to the case of a *unitary* map ($\Lambda = \mathcal{U}$ with a trivial single Kraus operator $K = U$), whereas a quantum channel with $r = d_\mathrm{in}^2$ is termed to be *full-rank*—i.e. it possesses the maximal possible number of linearly independent Kraus operators for a given $\mathcal{H}_\mathrm{in}$.

On the other hand, one may also show (see e.g. [Nielsen and Chuang, 2000]) that, for a given Kraus representation of a quantum channel, one can always find all U^SE, $|\xi\rangle_\mathrm{E}$ and $|i\rangle_\mathrm{E}$ in Eq. (2.13), which are defined in the enlarged space $\mathcal{H}_\mathrm{S} \otimes \mathcal{H}_\mathrm{E}$ and yield particular Kraus operators. Hence, the general system+environment interpretation (2.11) of the channel action is indeed always valid. Moreover, as any mapping Λ expressed in the *Kraus form* (2.13) must be CP (as Λ admits then a positive Choi-Jamiołkowski matrix representation [Bengtsson and Życzkowski, 2006]—see Sec. 2.4.1), we arrive at the *Stinespring dilation theorem* (see Fig. 2.1), which states that:

Theorem 2.2.1 (Stinespring dilation theorem). Any CPTP map may always be written according to Eq. (2.13) as a unitary transformation acting on an enlarged space with the environmental degrees of freedom eventually traced-out.

---

[12]Throughout this work, we represent by the *calligraphic* fount the *super-operators*—maps acting on density matrices: $\mathcal{T}(\mathcal{H}_\mathrm{in}) \to \mathcal{T}(\mathcal{H}_\mathrm{out})$—e.g. $\mathcal{U}$ and $\mathcal{I}$ for a unitary and identity channels repectively, in contrast to U and $\mathbb{I}$ representing operators acting on Hilbert spaces.
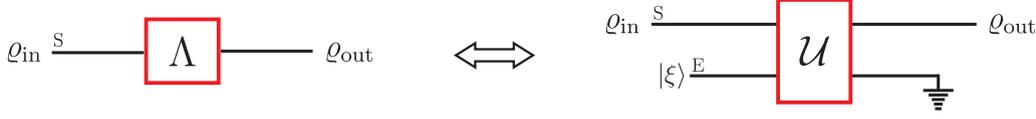
FIGURE 2.1: **Stinespring dilation theorem** stating that *any* quantum channel $\Lambda : \mathcal{T}(\mathcal{H}_{\text{in}}) \rightarrow \mathcal{T}(\mathcal{H}_{\text{out}})$, i.e. CPTP map that transforms all $\varrho_{\text{in}} \in \mathcal{T}(\mathcal{H}_{\text{in}})$ onto some $\varrho_{\text{out}} \in \mathcal{T}(\mathcal{H}_{\text{out}})$, may always be rewritten as a unitary transformation acting on an enlarged input space, $\varrho_{\text{in}} \otimes |\xi\rangle_{\text{E}}\langle\xi|$, with the environmental degrees of freedom eventually traced-out.

Hence, the only constraint that must be generally satisfied by any Kraus operators reads: $\sum_{i=1}^{d} K_i^\dagger K_i = \mathbb{I}^{\text{S}}$, and it ensures that the TP-property is fulfilled by a given quantum channel. As a consequence, given a set $\{K_i\}_{i=1}^{d}$ of Kraus operators of Λ, we may always construct another valid Kraus representation of the map with help of any rectangular matrix[13] u of size $d' \times d$ that for any $1 \le i \le d$ (and similarly for j) satisfies $\sum_{k=1}^{d'} \mathsf{u}_{ik}^\dagger \mathsf{u}_{kj} = \delta_{ij}$. It is so, because we can then always write a new set of d' Kraus operators:

$$\tilde{K}_i = \sum_{j=1}^{d} \mathsf{u}_{ij} K_j, \tag{2.14}$$

that also satisfy $\sum_{i=1}^{d'} \tilde{K}_i^\dagger \tilde{K}_i = \mathbb{I}^{\text{S}}$ and correctly represent the action of the channel, i.e. $\Lambda[\bullet] = \sum_{i=1}^{d'} \tilde{K}_i \bullet \tilde{K}_i^\dagger$. However, typically in problems which require optimisation over Kraus representations of a given quantum channel, it is enough to consider only the sets of linearly independent Kraus operators. Thus, when searching for the optimal Kraus representation, it is then sufficient to consider any $\{K_i\}_{i=1}^{r}$ with r being the rank of the channel Λ (e.g. the unique *canonical* Kraus representation defined with help of the Choi-Jamiołkowski matrix—see Sec. 2.4.1) and generate all other Kraus representations of interest via Eq. (2.14) after restricting to square—and thus *unitary*—matrices u.

Note 2.4: Partial trace as a CPTP map.

Lastly, let us show that the *partial trace* operation (previously introduced in Sec. 2.1.1 to represent the inability of accessing some parts of a system) consistently corresponds to a valid CPTP map, as it may be straightforwardly rewritten in the Kraus form (2.13). Defining a general reduced density matrix as $\varrho^{\text{A}} = \text{Tr}\{\varrho^{\text{AB}}\} = \sum_{i=1}^{d_{\text{B}}} {}_{\text{B}}\langle i|\varrho^{\text{AB}}|i\rangle_{\text{B}}$, we may trivially construct the necessary Kraus operators $K_i : \mathcal{H}_{\text{A}} \otimes \mathcal{H}_{\text{B}} \rightarrow \mathcal{H}_{\text{A}}$, such that each $K_i = \mathbb{I}^{\text{A}} \otimes {}_{\text{B}}\langle i|$ and thus $\varrho^{\text{A}} = \sum_{i=1}^{d_{\text{B}}} K_i \varrho^{\text{AB}} K_i^\dagger$ as required.

2.2.2 *Master equation* picture of system evolution

From an alternative perspective, rather than analysing the properties of a given quantum channel describing the overall evolution between times t_0 and t, one may study the form of the differential equation on the r.h.s. of Eq. (2.11). Although, such a *master equation* picture is equivalent to the quantum map representation given the initial conditions, i.e. the input state $\varrho_{\text{in}} \equiv \varrho^{\text{S}}(t_0)$ at t_0, it allows to determine the dynamical equations of motion at any time instance for the reduced density matrix, and thus generalise the notion of classical stochastic processes into the quantum setting[14] [Breuer and Petruccione, 2002].

Separating from the overall Hamiltonian $\hat{H}_{\text{SE}}$ in Eq. (2.11) the Hamiltonian $\hat{H}_{\text{S}}$, which specifies the free—*unitary*—evolution of the system in the absence of interactions with the environment, we may write the

---

[13] We denote the matrices u with a different fount to indicate that these do *not* represent operators on Hilbert spaces, but should be treated as matrices of complex entries just specifying linear transformations of Kraus operators.

[14] In particular, the terminology of classical Markov master equations, e.g. the Chapman-Kolmogorov differential equation, [Breuer and Petruccione, 2002].

master equation in the form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho(t) \;=\; -\mathrm{i}\big[\hat{H}, \varrho(t)\big] \;+\; \mathcal{L}[\varrho(t)] \tag{2.15}$$

where we have dropped the system-labelling S without loss of generality and introduced the *Louvillian* $\mathcal{L}$ (linear super-operator acting on density matrices), which is responsible for the *dissipative* part of the evolution—caused by the destructive impact of the environment. Notice that for the purpose of this work we have already assumed the Louvillian $\mathcal{L}$ to be *time-independent*, so that Eq. (2.15) constitutes really the *Markovian quantum master equation* [Breuer and Petruccione, 2002]. In general, in order to allow for non-Markovian effects [Addis *et al.*, 2014; Rivas *et al.*, 2014], i.e. the emergence of memory in the evolution via the "information back-flow" from the environment to the system, one must allow $\mathcal{L}$ to depend on time, yet other equivalent generalisations of Eq. (2.15) are also possible [Chruściński and Kossakowski, 2010].

The form of $\mathcal{L}$ may be determined by considering a particular physical model of the evolution under the Markovian assumptions of system-environment weak-coupling and infinitely small correlation time of the environment [Breuer and Petruccione, 2002] (see e.g. the *quantum optical master equations* for the model of atom–electromagnetic-field interactions [Gardiner and Zoller, 2000]). However, one may generally show that for the Louvillian to be physical, it must possess the so-called *Lindblad-Gorini-Kossakowski-Sudarshan* (LGKS) form, which is *necessary and sufficient* for the corresponding effective quantum map of Eq. (2.11), i.e. $\Lambda_t : \varrho(0) \rightarrow \varrho(t)$, to be CPTP (as defined in the previous section) for any t.

Lindblad-Gorini-Kossakowski-Sudarshan form
In their seminal works Gorini *et al.* [1976]; Lindblad [1976] have demonstrated that for the Markovian master equation (2.15) to always yield a valid CPTP map in the quantum channel picture of Sec. 2.2.1, the Louvillian $\mathcal{L}$ must generally read:

$$\mathcal{L}[\varrho(t)] \;=\; \sum_{k=1}^{d^2-1} \gamma_k \left[L_k\,\varrho(t)\,L_k^\dagger \;-\; \frac{1}{2}\left\{L_k^\dagger L_k, \varrho(t)\right\}\right], \tag{2.16}$$

where $\gamma_k \geq 0$ are the non-negative 'decay rates' of the dissipation (decoherence) process, $\{A, B\} = AB + BA$ denotes the anti-commutator, d is the dimension of the system Hilbert space, and L_k are the so-called *Lindblad operators*—that formally are *not* constrained to possess any particular structure[15].

Nevertheless, let us already remark that from the perspective of metrology and, in particular, the results presented in this work, it is the quantum channel picture (2.13) that turns out to be more appropriate, as it provides the geometric properties of the evolution (discussed in Sec. 2.4) that have a crucial impact on the performance of metrological protocols. Yet, when the time duration of the evolution, t, serves as an extra degree of freedom that one can control, or when the dynamics is simply specified at the differential equation level (e.g. for Non-Markovian or non-commuting with $\hat{H}$ noise models [Chaves *et al.*, 2013; Chin *et al.*, 2012; Matsuzaki *et al.*, 2011]), the master equation picture (2.15) must still be utilised in order to determine the form of the effective map Λ_t in Eq. (2.11), which properties will then importantly vary depending on the t considered. In what follows, we describe the qubit evolution models of metrological relevance and determine the relation between their quantum channel and master equation pictures—what happens to be straightforward due to the commutativity of the unitary, $[\hat{H}, \bullet]$, and the dissipative, $\mathcal{L}[\bullet]$, terms in Eq. (2.15) ($[\hat{H}, \mathcal{L}[\bullet]] \equiv \mathcal{L}[[\hat{H}, \bullet]]$) for the evolutions considered.

---

[15]Notice that, due to the master equation formulation (2.15) and the LGKS form (2.16), the TP-property $\mathrm{Tr}\{\varrho(t)\} = 1$ is trivially preserved at all times, as $\mathrm{Tr}\{\mathcal{L}[\varrho(t)]\} = 0$ by construction.
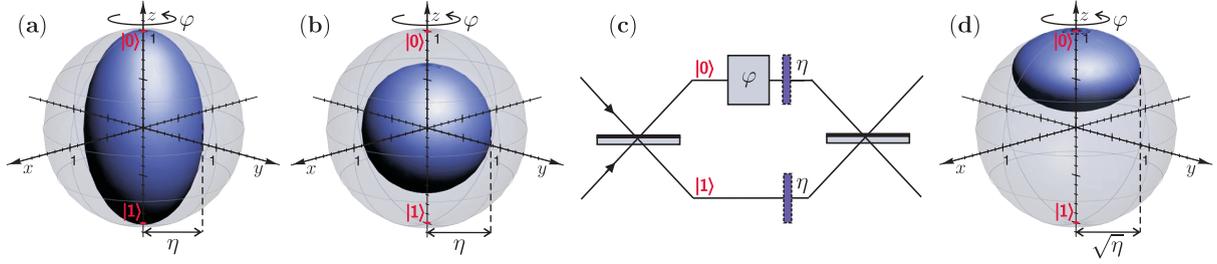
FIGURE 2.2: **Noisy-phase–estimation channels** of relevance to quantum metrological schemes. In each case, a particle is represented by a *qubit* in the basis $\{|0\rangle, |1\rangle\}$, which in the absence of noise solely rotates around the z axis by an angle specified by the parameter φ. The noise is modelled by the quantum maps $\mathcal{D}_\eta$ of strength η (see Tab. 2.1) that commute with such a rotation: (**a**) *dephasing*, (**b**) *depolarisation*, (**c**) *loss*, (**d**) *spontaneous emission* (*amplitude damping*). In case of the *loss* noise model, the overall transformation does not possess a Bloch ball representation as it corresponds to a qubit→qutrit map. Yet, as shown in (**c**), the input qubit has then a natural interpretation of a photon propagating through an interferometer with equal power-transmittances of its arms.

2.2.3   EXAMPLE: **Noisy-phase–estimation channels of relevance to metrology**

Throughout this work, while analysing various tools introduced to study the noise effects in quantum metrology, we consider the natural examples of *qubit* evolutions that are of particular relevance to metrological protocols. As depicted in Fig. 2.2, these correspond to: *dephasing*, *depolarisation*, *loss*, and *spontaneous emission* (*amplitude damping*) noise models; that are typically utilised when accounting for decoherence in optical interferometry [Demkowicz-Dobrzański *et al.*, 2015] and atomic experiments [Hammerer *et al.*, 2010; Leibfried *et al.*, 2003].

Quantum channel picture
In order to describe these exemplary evolution models in the quantum channel picture (2.13), we consider a qubit representing a particle (two-level atom, photon in two modes) prepared in a state ϱ_{in} which undergoes an overall CPTP map Λ_φ according to

$$\varrho_{\mathrm{out}} = \Lambda_\varphi[\varrho_{\mathrm{in}}] = \mathcal{D}_\eta[\mathcal{U}_\varphi[\varrho_{\mathrm{in}}]] = \sum_{i=1}^{r} K_i(\varphi)\, \varrho_{\mathrm{in}}\, K_i^\dagger(\varphi) \tag{2.17}$$

that is composed of consecutively a unitary $\mathcal{U}_\varphi$ and a pure-noise $\mathcal{D}_\eta$ channel. $\mathcal{U}_\varphi[\varrho] = U_\varphi \varrho U_\varphi^\dagger$ with $U_\varphi = \exp[-i\hat{\sigma}_z \varphi/2]$, so that it generates a rotation by an angle φ around the z axis in the Bloch ball representation (see Fig. 2.2), whereas $\mathcal{D}_\eta$ models one of the noise-types depicted in Fig. 2.2, and specified in Tab. 2.1 with the effective noise strength set to η. For convenience, we have chosen in Eq. (2.17) $\mathcal{U}_\varphi$ to act before the decoherence, yet the ordering of the two in principle plays no role, as the unitary rotation commutes with the noise models considered. However, such a choice allows us to similarly define the Kraus representation (2.13) of Λ_φ for all four cases depicted in Fig. 2.2, i.e. as $K_i(\varphi) = K_i U_\varphi$ with K_i representing any valid (see Eq. (2.14)) set of Kraus operators for corresponding pure-noise maps $\mathcal{D}_\eta$. Yet, for each above noise-type, we explicitly specify in Tab. 2.1 the canonical Kraus operators of the adequate channel $\mathcal{D}_\eta$, which number then also determines (see Sec. 2.4.1) the rank[16] r of both $\mathcal{D}_\eta$ and[17] Λ_φ. Notice that the above-introduced formulation correctly applies also to the *loss* noise model depicted

---

[16]Note that $1 \le r \le 4$ for qubit-input channels, as $d_{\mathrm{in}}^2 = 4$.
[17]Also for Λ_φ, as the concatenation with a unitary map in Eq. (2.17) does not vary the rank of a channel.

Noise model	Kraus representation of $\mathcal{D}_\eta$	Liouvillian $\mathcal{L}[\varrho]$	$\eta(t)$
Dephasing: ($r=2$)	$K_1 = \sqrt{\frac{1+\eta}{2}}\,\mathbb{I}\,,\;\; K_2 = \sqrt{\frac{1-\eta}{2}}\,\hat{\sigma}_z$	$\frac{\gamma}{2}\left(\hat{\sigma}_z\varrho\,\hat{\sigma}_z - \varrho\right)$	$\mathrm{e}^{-\gamma t}$
Depolarization: ($r=4$)	$K_1 = \sqrt{\frac{1+3\eta}{4}}\,\mathbb{I}\,,\;\; \left\{K_i = \sqrt{\frac{1-\eta}{4}}\,\hat{\sigma}_{i-1}\right\}_{i=2}^4$	$\frac{\gamma}{2}\left(\frac{1}{3}\sum_{i=1}^{3}\hat{\sigma}_i\varrho\,\hat{\sigma}_i - \varrho\right)$	$\mathrm{e}^{-\frac{2\gamma}{3}t}$
Loss: ($r=3$)	$K_1 = \begin{pmatrix} \sqrt{\eta} & 0 \\ 0 & \sqrt{\eta} \\ 0 & 0 \end{pmatrix},\;\; K_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \sqrt{1-\eta} & 0 \end{pmatrix},$ $K_3 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \sqrt{1-\eta} \end{pmatrix}$	$\gamma\sum_{m=0}^{1}\left(\sigma_{m,+}\varrho\,\sigma_{m,-} - \frac{1}{2}\left\{\sigma_{m,-}\sigma_{m,+},\varrho\right\}\right)$	$\mathrm{e}^{-\gamma t}$
Spontaneous emission: ($r=2$)	$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{\eta} \end{pmatrix},\;\; K_2 = \begin{pmatrix} 0 & \sqrt{1-\eta} \\ 0 & 0 \end{pmatrix}$	$\gamma\left(\sigma_+\varrho\,\sigma_- - \frac{1}{2}\left\{\sigma_-\sigma_+,\varrho\right\}\right)$ $\scriptstyle{\sigma_\pm = \frac{1}{2}(\hat{\sigma}_x\pm\mathrm{i}\hat{\sigma}_y)}$	$\mathrm{e}^{-\gamma t}$

TABLE 2.1: **Characteristics of the noisy-phase–estimation channels of Fig. 2.2.** For each qubit-evolution–model considered, the canonical Kraus representation of the pure-noise map $\mathcal{D}_\eta$ is specified, which also determines the rank, r, of the effective map Λ_φ in the quantum channel picture (2.17). Moreover, the corresponding Liouvillians, $\mathcal{L}$, of the LGKS form (2.16) are presented that apply at the level of the master equation (2.18). In case of the *loss* model, $\sigma_{m,+} = |\mathrm{vac}\rangle\langle m|$ are the generators of transitions into the vacuum state from the qubit basis vectors $|m=\{0,1\}\rangle$, such that $\sigma_{m,-} = \sigma_{m,+}^\dagger$. As all pure-noise maps commute with the phase encoding, in order to construct the effective quantum channel, $\Lambda_{\omega,t}$, from the master equation (2.18), it is sufficient to set $\varphi\to\omega t$ in Λ_φ and account for the time-dependence of the decoherence-strength parameters, $\eta\to\eta(t)$, in accordance with the last column.

in Fig. 2.2(**c**), in which case the effective pure-noise map, $\mathcal{D}_\eta$, transforms a qubit into a qutrit with the third basis state (vacuum) representing the particle being lost, i.e. $\{|0\rangle,|1\rangle\}\to\{|0\rangle,|1\rangle,|\mathrm{vac}\rangle\}$.

Master equation picture

On the other hand, we may also determine the equivalent description of the qubit evolution models of Fig. 2.2 in the master equation picture (2.15) by defining $\varrho(t_0)=\varrho_{\mathrm{in}}$ at t_0 and specifying:

$$\frac{\mathrm{d}}{\mathrm{d}t}\varrho(t) = -\mathrm{i}\frac{\omega}{2}\left[\hat{\sigma}_z,\varrho(t)\right] + \mathcal{L}[\varrho(t)], \tag{2.18}$$

with the Liouvillians $\mathcal{L}$ (of the LGKS form (2.16)) listed in Tab. 2.1 for each of the noise-types considered. As a result, by integrating Eq. (2.18) over the interval $[t_0,t]$ we can construct the effective map[18] such that $\varrho(t)=\Lambda_{\omega,t}[\varrho(t_0)]$, i.e. the equivalent of Λ_t in Eq. (2.11). However, for the above qubit models, $\Lambda_{\omega,t}$ may also be determined with help of the quantum channel picture by simply setting $\varphi\to\omega t$ in Λ_φ of Eq. (2.17) and letting the decoherence strength be also a time-dependent function $\eta\to\eta(t)$—that we specify for each of the noise-types in the last column of Tab. 2.1. Importantly, such a straightforward relation between Eq. (2.17) – in which we have *artificially* treated the noise and unitary maps to be separate, and Eq. (2.18) – that *appropriately* models the noise to occur simultaneously to phase acquisition at all times, is only valid due to the commutativity of the unitary $\mathcal{U}_\varphi$ and pure-noise $\mathcal{D}_\eta$ channels in Eq. (2.17). In case of more general noise models, e.g. non-Markovian [Chin *et al.*, 2012; Matsuzaki *et al.*, 2011] or non-commuting [Chaves *et al.*, 2013], we cannot naively separate the decoherence from the free evolution,

---

[18]We have on purpose chosen to denote the overall map between t_0 and t as $\Lambda_{\omega,t}$, in order to match the convention later used in Sec. 4.4, where ω of Eq. (2.18) represents then the estimated detuning *frequency* in atomic spectroscopy setups.

but must rather start with the phenomenologically determined equivalent of the master equation (2.18) (e.g. with $\mathcal{L}_t$ being now t-dependent), and by integrating it over $[t_0, t]$ construct the effective quantum map Λ_t of Eq. (2.11) such that $\varrho(t) = \Lambda_t[\varrho(t_0)]$ for a given t. On one hand, such a construction is always possible once the expressions for both $\varrho(t_0)$ and $\varrho(t)$ are known[19] [Bengtsson and Życzkowski, 2006], yet one may also follow a general recipe and often construct the Kraus representation of the channel Λ_t directly from the master equation, i.e. without need of explicit integration [Andersson *et al.*, 2007].

2.3 Quantum metrology – estimation of *latent* parameters of a quantum system

In statistics [Everitt and Skrondal, 2010], while analysing classical experiments that yield random data, one divides the variables used in the description of the investigated process into two[20] natural categories of: the *observable* (manifest) variables – ones representing the *properties* of the system undergoing the process that may be directly measured; and the *latent* variables – ones corresponding to the *parameters* of the mathematical model assumed to describe the stochastic process itself. Importantly, the latent parameters are by definition intrinsic to the description and cannot be measured, so that for their determination one must resort to the techniques of *statistical inference* [Wasserman, 2004] and, in particular, the *estimation theory* [Kay, 1993].

Such a formalism naturally carries over into the quantum setting, in which the observable variables correspond exactly to the *quantum observables*—Hermitian operators $\hat{O}$ introduced in Sec. 2.1.2—that, as explained before, indeed represent the physical properties (e.g. position, momentum, spin, energy etc.) of a given quantum system. The *quantum latent parameters*, on the other hand, represent then the quantities that characterise the system evolution (e.g. the time t in Eq. (2.10), decay rates γ_k in Eq. (2.16), φ and η parameters in Eq. (2.17) or ω in Eq. (2.18) etc.), and should be interpreted as classical (possibly random) variables inbuilt in the description, extrinsic to the quantum system.

Importantly, given a fixed state of the system, ϱ, the latent parameters cannot be even defined unless an explicit dynamical model is provided that explains how the parameters of interest have been *encoded*. Formally, when we label the state as ϱ_φ with φ representing a latent parameter, we also implicitly have in mind the φ-encoding process and thus the knowledge of how the form of ϱ_φ varies with φ. On the other hand, as the most general quantum measurements that may be performed on such a state, i.e. the POVMs $\{M_i\}_i$ introduced in Sec. 2.1.2, just yield classical outcomes distributed with (φ-dependent) probabilities $p_i(\varphi) = \mathrm{Tr}\{\varrho_\varphi M_i\}$, in order to most accurately determine the value of φ we must still utilise the classical estimation techniques [Kay, 1993], so that the parameter can be most efficiently inferred from the sampled data. Hence, the extra freedom that the quantum framework really gives us is the choice of the *optimal input state* on which the parameter φ may be encoded, but also the choice of the most effective measurement strategy—the *optimal POVM*—that leads to the distribution of outcomes from which φ can then be most accurately deduced. As a consequence, one may neatly conclude that the *purpose of quantum metrology* simply boils down to

"the most precise determination of the latent parameters of a quantum system",

---

[19]As one may then construct the 'dynamical' matrix transforming $\varrho(t_0) \to \varrho(t)$ and 'reshuffle' it to obtain the Choi-Jamiołkowski matrix (2.24) [Bengtsson and Życzkowski, 2006], from which the canonical Kraus representation may be directly determined (see Sec. 2.4.1).

[20]In fact, one could also define a third category that is beyond the scope of this work, i.e. the *hidden* variables – ones that could in principle be measured and observed but are not accessible for practical reasons. Notice that these also have a direct application in the quantum setting, being naturally utilised when describing the phenomenon of *quantum non-locality* and, in particular, the *local hidden-variable models* [Brunner *et al.*, 2014].

what, in practice, combines the optimisation of both states and measurements at the quantum mechanical level with classical estimation tools that must be utilised at the statistical-data interpretation stage.

2.3.1 Uncertainty relations for quantum observables

Before focussing on the problem of quantum latent-parameters determination, let us briefly discuss the consequences of the physical properties of a quantum system being described by means of the observable formalism. Crucially, an observable $\hat{O}$ fully determines the statistics of its measurement, as every moment of the random variable O representing the measurement outcomes (being the eigenvalues, o_i, of $\hat{O}$, see Sec. 2.1.2), can be generally written as

$$\forall_{k\in\mathbb{N}}: \quad \langle O^k \rangle = \sum_i p_i o_i^k = \sum_i \langle P_i \rangle o_i^k = \left\langle \sum_i P_i o_i^k \right\rangle = \left\langle \left(\sum_i P_i o_i \right)^k \right\rangle = \langle \hat{O}^k \rangle, \tag{2.19}$$

where we have acknowledged that the projectors P_i satisfy $P_i P_j = P_i \delta_{ij}$, and naturally generalised the averaging operation, $\langle \bullet \rangle = \sum_i p_i \bullet$, to the quantum setting, so that for a given state ϱ: $\langle \bullet \rangle = \text{Tr}\{\varrho \bullet\}$. Importantly, Eq. (2.19) proves that all the statistical measures of the random variable O are just represented by their quantum mechanical ("hatted") equivalents evaluated for the observable $\hat{O}$. For instance, the *variance* $\text{Var}[O] = \langle (O - \langle O \rangle)^2 \rangle$ is directly translated onto[21] $\Delta^2 \hat{O} = \langle (\hat{O} - \langle \hat{O} \rangle)^2 \rangle$, which thus equivalently quantifies the spread of the outcomes distribution, when the observable of interest is measured.

One of the spectacular features of the quantum theory is the phenomenon of *incompatibility of the observables* [Griffiths, 2013; Phillips, 2003], which states that if two observables, say $\hat{A}$ and $\hat{B}$, do not commute with one another, i.e. $[\hat{A}, \hat{B}] \neq 0$, there does not exist a quantum system which possesses both physical properties associated with $\hat{A}$ and $\hat{B}$ precisely defined. Such a fact is a direct consequence of the so-called *Heisenberg uncertainty relation* ensuring that one can only decrease one of the observable variances at the expense of the other, i.e.

$$\Delta^2 \hat{A} \, \Delta^2 \hat{B} \quad \geq \quad \frac{1}{4} \left| \langle [\hat{A}, \hat{B}] \rangle \right|^2 + \left| \text{cov}(\hat{A}, \hat{B}) \right|^2 \quad \geq \quad \frac{1}{4} \left| \langle [\hat{A}, \hat{B}] \rangle \right|^2, \tag{2.20}$$

where the above stronger and weaker bounds have been discovered by (and are named after) Schrödinger [1930] and Robertson [1934] respectively. Similarly to random variables, $\text{cov}(\bullet, \bullet)$ denotes the *covariance* of quantum observables, i.e. $\text{cov}(\hat{A}, \hat{B}) = \frac{1}{2} \langle \{\hat{A}, \hat{B}\} \rangle - \langle \hat{A} \rangle \langle \hat{B} \rangle$. Operationally, Eq. (2.20) means that given *infinitely many* copies of the same system, if we measure $\hat{A}$ on half of them, while $\hat{B}$ on the rest, the variances determined by the two data sets collected must satisfy the above Robertson-Schrödinger inequalities. For completeness, let us note that this contrasts the scenario of the recently vividly researched topic of the *noise-disturbance uncertainty relations* [Branciard, 2013; Busch *et al.*, 2013; Ozawa, 2004] which aim to relate the precision with which one measures one of the observables, say $\hat{A}$, on a *single* copy of the system to the error in measurement of the other observable, $\hat{B}$, performed afterwards.

2.3.2 Inferring a latent parameter from a quantum observable

At this preliminary stage of presentation, let us consider a simple strategy of latent parameter inference, in which we measure some observable $\hat{O}$ of a system in a state ϱ_φ, in order to most accurately determine

---

[21]Following the standard convention, we label by $\Delta^2 \hat{O}$ the *variance* of an observable. Yet, as for random variables $\Delta^2 O$ is typically utilised to express the *Mean Squared Error* (3.3) in estimation theory, we use the $\text{Var}[O]$ notation instead.

a fixed, *deterministic* latent parameter φ. We achieve this by constructing an estimate[22] of φ, call it $\tilde{\varphi}$, which is built on a single outcome and thus corresponds to a function of the random variable O: $\tilde{\varphi}(O)$. Hence, if we assume that O fluctuates in a very narrow region around its mean[23], so that $\mathrm{Var}[O] = \Delta^2 \hat{O} \ll 1$, we may *locally* Taylor-expand $\tilde{\varphi}$ up to first order around $\langle \hat{O} \rangle$ [Barlow, 2013]:

$$\tilde{\varphi}(O) \;=\; \tilde{\varphi}\big(\langle \hat{O} \rangle\big) + \frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}O}\bigg|_{\langle \hat{O} \rangle} \big(O - \langle \hat{O} \rangle\big) + \dots \tag{2.21}$$

and by squaring and averaging Eq. (2.21) obtain the *error-propagation formula* (e.g. [Wineland *et al.*, 1992]):

$$\Delta^2 \tilde{\varphi} \;\approx\; \frac{\Delta^2 \hat{O}}{\left| \frac{\mathrm{d}\langle \hat{O} \rangle}{\mathrm{d}\varphi} \right|^2}, \tag{2.22}$$

where we have defined[24] $\Delta^2 \tilde{\varphi} = \big\langle (\tilde{\varphi}(O) - \varphi)^2 \big\rangle$ and assumed that at the mean value $\langle \hat{O} \rangle$ we precisely estimate the true parameter, i.e. $\tilde{\varphi}\big(\langle \hat{O} \rangle\big) = \varphi \implies \frac{\mathrm{d}\tilde{\varphi}(\langle \hat{O} \rangle)}{\mathrm{d}\varphi} = 1$, so that $\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}O}\big|_{\langle \hat{O} \rangle} = \frac{\mathrm{d}\tilde{\varphi}(\langle \hat{O} \rangle)}{\mathrm{d}\varphi} \left[\frac{\mathrm{d}\langle \hat{O} \rangle}{\mathrm{d}\varphi} \right]^{-1} = \left[\frac{\mathrm{d}\langle \hat{O} \rangle}{\mathrm{d}\varphi} \right]^{-1}$.

Importantly, the error-propagation formula (2.22) quantifies the fluctuations of our latent-parameter–estimate $\tilde{\varphi}$ around the true value φ, which are unavoidable due to the stochasticity of the observable measurement. As a result, Eq. (2.22) also specifies the ultimate sensitivity of $\tilde{\varphi}$ to any variations of φ (still being deterministic!) in such a "small fluctuations of O" (local—see later Sec. 3.1.2) regime, for a particular observable that is assumed to be measured. This makes it a powerful tool that may be utilised to quantify performance of quantum metrological protocols employing a specific observable-based measurement strategy. In fact, it was Eq. (2.22) that has been utilised in the pioneering works of [Bondurant and Shapiro, 1984; Caves, 1981; Dowling, 1998; Holland and Burnett, 1993; Sanders and Milburn, 1995; Yurke *et al.*, 1986] to determine the maximal precision with which *phase* may be resolved in optical interferometry, and similarly for the estimation of atomic transition *frequency* in atomic spectroscopy [Bollinger *et al.*, 1996; Wineland *et al.*, 1994, 1992].

Note 2.5: Mandelstam-Tamm inequality – time-energy uncertainty relation.

A remarkable consequence of both the Robertson inequality (2.20) and the error-propagation formula (2.22) is the *time-energy uncertainty relation* originally discovered by Mandelstam and Tamm [1945][a]. In the special case of *time* being the parameter estimated, i.e. $\varphi \equiv t$, the derivative of the operator mean in Eq. (2.22), may be generally rewritten utilising the dynamical von Neumann equation (2.10), as $\frac{\mathrm{d}\langle \hat{O} \rangle}{\mathrm{d}t} = \mathrm{i}\big\langle [\hat{H}, \hat{O}] \big\rangle$. As a result, after substituting $\hat{A} = \hat{H}$ and $\hat{B} = \hat{O}$ into Eq. (2.20), as well as for $\Delta^2 \hat{O}$ with help the error-propagation formula, we obtain (independently of $\hat{O}$ assumed) the so-called *Mandelstam-Tamm inequality*:

$$\Delta^2 \hat{H}\, \Delta^2 \tilde{t} \;\geq\; \frac{1}{4}, \tag{2.23}$$

which states that the variance of the Hamiltonian $\hat{H}$ sets a lower limit on the magnitude of fluctuations of the time-estimate $\tilde{t}$, and thus the resolution with which one can sense the variations of the elapsed time t. Notice that $\Delta^2 \hat{H}$ is really determined by the energy spectrum of a given system, so that, for instance, assuming the system to be in a pure state $|\psi\rangle$ written in the basis of the energy-eigenstates as $|\psi\rangle = \sum_i \alpha_i |E_i\rangle$, where $\forall_i : \hat{H}|E_i\rangle = E_i|E_i\rangle$, the Hamiltonian variance just reads: $\Delta^2 \hat{H} = \sum_i E_i^2 |\alpha_i|^2 - \big(\sum_j E_j |\alpha_j|^2\big)^2$.

[a]For an alternative version of the time-energy uncertainty relation based on different principles, see [Margolus and Levitin, 1998].

---

[22]In fact, $\tilde{\varphi}$ is formally termed an *estimator*, as defined later in Sec. 3.1.

[23]What (owing to the Central Limit Theorem) can always be assured, if we conduct a large enough number of procedure repetitions. In fact, this is exactly the assumption of *locality* that is essential when pursuing the frequentist approach to parameter estimation (see Sec. 3.1.2.3).

[24]$\Delta^2 \tilde{\varphi}$ actually corresponds to the *Mean Squared Error* (3.3) introduced later in Sec. 3.1.

2.4 Geometry of φ-parametrised quantum channels

Lastly, in order to describe in more detail the evolution of a quantum system employed in a metrological scenario, we follow Sec. 2.2.1 and apply the language of quantum channels that are then crucially responsible for the encoding of the latent parameter to be determined. Hence, after identifying ϱ_φ as the output state (previously labelled as ϱ_{out} in Sec. 2.2.1), we generally write $\varrho_\varphi = \Lambda_\varphi[\varrho_{\text{in}}]$, where the evolution CPTP map, $\Lambda_\varphi : \mathcal{T}(\mathcal{H}_{\text{in}}) \to \mathcal{T}(\mathcal{H}_{\text{out}})$, is now explicitly parametrised by the estimated parameter. Importantly, we can thus effectively treat the variations of φ as changes in the form of Λ_φ, and—by defining a *family* of CPTP maps $\{\Lambda_\varphi\}_\varphi$ parametrised by the latent parameter φ—interpret the φ-estimation task as a problem of determining which of the channels from the family has acted on the input state ϱ_{in}.

As it is thus the family $\{\Lambda_\varphi\}_\varphi$ that contains all the information about the latent parameter, it is necessary to describe in more detail its mathematical form. That is why, in what follows, we introduce yet another tool of quantum-channel formalism, i.e. the *Choi-Jamiołkowski* (CJ) *isomorphism*, which allows us to establish the geometrical structure of the space of quantum channels. As result, we are able to study the *geometry* of φ-parametrised CPTP maps and, in particular, define the notion of φ-*extremality* of a given channel Λ_φ. We later show that the φ-extremality property plays an important role when analysing metrological scenarios, in which Λ_φ is responsible for the parameter encoding. In the last part of this section, in order to make these ideas clear, we discuss in detail the geometrical properties of the exemplary qubit noisy-phase–estimation channels introduced in Sec. 2.2.3.

2.4.1 Choi-Jamiołkowski isomorphism

Given a general quantum channel, $\Lambda : \mathcal{T}(\mathcal{H}_{\text{in}}) \to \mathcal{T}(\mathcal{H}_{\text{out}})$, introduced in Sec. 2.2.1, we define its *Choi-Jamiołkowski* (CJ) *matrix* [Choi, 1975; Jamiołkowski, 1972], Ω_Λ, as a state supported by an enlarged space $\mathcal{T}(\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{\text{A}})$ (as in Eq. (2.12)) with $\dim \mathcal{H}_{\text{A}} = d_{\text{in}}$, such that[25]

$$\Omega_\Lambda = \Lambda \otimes \mathcal{I}[|\mathbb{I}\rangle\langle\mathbb{I}|] , \tag{2.24}$$

where $|\mathbb{I}\rangle = \sum_{i=1}^{d_{\text{in}}} |i\rangle_{\text{S}} |i\rangle_{\text{A}}$ is an (unnormalised) maximally entangled state defined on $\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{A}}$—the system (S) input Hilbert space and the one of the ancilla (A). We have without loss of generality defined $|\mathbb{I}\rangle$ to be unnormalised and thus $\text{Tr}\{\Omega_\Lambda\} = d_{\text{in}}$, in order to benefit from a concise notation for bipartite states, in which $|\phi\rangle = \sum_{i,j=1}^{d_{\text{in}}} \langle i|\phi|j\rangle |i\rangle_{\text{S}} |j\rangle_{\text{A}} = \phi \otimes \mathbb{I}|\mathbb{I}\rangle = \mathbb{I} \otimes \phi^T |\mathbb{I}\rangle$. For instance, for any three operators $A, B, C \in \mathcal{T}(\mathcal{H})$ with $\dim(\mathcal{H}) = d_{\text{in}}$, we may then equivalently write $A \otimes C|B\rangle = |ABC^T\rangle = ABC^T \otimes \mathbb{I}|\mathbb{I}\rangle = \mathbb{I} \otimes CB^T A^T |\mathbb{I}\rangle$.

As a result, given *any* Kraus representation of Λ, e.g. $\{K_i\}_{i=1}^d$ of Eq. (2.13), we can always write the CJ matrix (2.24) as $\Omega_\Lambda = \sum_{i=1}^d |K_i\rangle\langle K_i|$, so that it may be interpreted as a mixture of states $|K_i\rangle$ and thus must be *positive semi-definite*[25]. On the other hand, assuming that $\Omega_\Lambda \geq 0$, we may always perform its eigendecomposition to obtain $\Omega_\Lambda = \sum_{i=1}^r \lambda_i |\psi_i\rangle\langle\psi_i|$ with $\lambda_i > 0$, and by utilising the above bipartite notation *unambiguously* construct the *canonical Kraus operators* of the corresponding quantum channel Λ as $\{K_i = \sqrt{\lambda_i}\psi_i\}_{i=1}^r$ that satisfy $K_i \otimes \mathbb{I}|\mathbb{I}\rangle = \sqrt{\lambda_i}|\psi_i\rangle$. Importantly, the *rank* r (number of $\lambda_i > 0$) of the CJ matrix thus represents exactly to the *rank* r (defined in Sec. 2.2.1) of its corresponding quantum channel. Furthermore, by the above argumentation *the map Λ is CP—admits a Kraus representation— if and only if* $\Omega_\Lambda \geq 0$. The TP property of Λ, on the other hand, is assured by a constraint on the CJ matrix: $\text{Tr}_{\text{S}}\{\Omega_\Lambda\} = \mathbb{I}^{\text{A}} \Leftrightarrow \sum_i K_i^\dagger K_i = \mathbb{I}^{\text{S}}$.

---

[25]Notice the similarity to the CP-property definition (2.12). In general, the CJ-isomorphism (2.24) is valid for *any* linear transformation Λ, but $\Omega_\Lambda \geq 0$ *if and only if* Λ is CP.

Formally, Eq. (2.24) defines a linear mapping from the *space of quantum channels* $\Lambda : \mathcal{T}(\mathcal{H}_{\mathrm{in}}) \to \mathcal{T}(\mathcal{H}_{\mathrm{out}})$ onto the *space of their CJ matrices* $\Omega_\Lambda \in \mathcal{T}(\mathcal{H}_{\mathrm{out}} \otimes \mathcal{H}_{\mathrm{A}})$. On the other hand, an inverse linear mapping may also be constructed, by realising that the action of any channel Λ may be written with use of its CJ matrix (2.24) (see e.g. [Bengtsson and Życzkowski, 2006; Keyl and Werner, 2007]) via:

$$\forall_{\varrho_{\mathrm{in}} \in \mathcal{H}_{\mathrm{in}}}: \quad \Lambda[\varrho_{\mathrm{in}}] = \mathrm{Tr}\{\Omega_\Lambda (\mathbb{I} \otimes \varrho_{\mathrm{in}}^T)\}. \tag{2.25}$$

Thus, the mapping between the two spaces actually corresponds to an *isomorphism* [Jamiołkowski, 1972], so that importantly the space of all quantum channels $\Lambda : \mathcal{T}(\mathcal{H}_{\mathrm{in}}) \to \mathcal{T}(\mathcal{H}_{\mathrm{out}})$ possesses the same *geometric properties* as the space of the quantum states[26] $\Omega \in \mathcal{T}(\mathcal{H}_{\mathrm{out}} \otimes \mathcal{H}_{\mathrm{A}})$.

Note 2.6: Convexity of the space of quantum channels.
In particular, as the space of density operators is *convex* (see Note 2.1), so is the space of quantum channels. In order to prove such a fact explicitly, note that if we construct a convex combination of two CPTP maps $\Lambda_{1/2}$, i.e. $\Lambda = \lambda \Lambda_1 + (1-\lambda)\Lambda_2$ with any $0 \le \lambda \le 1$, owing to the linearity of Eq. (2.24) the corresponding CJ matrix is also just a convex sum of $\Omega_{\Lambda_{1/2}}$, i.e. $\Omega_\Lambda = \lambda \Omega_{\Lambda_1} + (1-\lambda)\Omega_{\Lambda_2}$. As Ω_Λ is thus trivially positive semi-definite and satisfies $\mathrm{Tr}_{\mathrm{S}}\{\Omega_\Lambda\} = \mathbb{I}^{\mathrm{A}}$, its equivalent channel Λ must respectively fulfil the CP and TP properties, and hence belong to the space of valid quantum maps.

2.4.2 Extremal and φ-extremal quantum channels

Extremal channels

Having shown that the space of all quantum channels can be equivalently interpreted as a *convex* space of the corresponding CJ matrices (2.24), we define a given channel to be *extremal*, if it cannot be decomposed into a convex sum of other CPTP maps [Bengtsson and Życzkowski, 2006]:

Definition 2.4.1 (Extremality of a quantum channel). A CPTP map Λ is *extremal* if and only if there do *not* exist any distinct CPTP maps $\Lambda_{1/2}$, such that $\Lambda = \lambda \Lambda_1 + (1-\lambda)\Lambda_2$ for some $0 < \lambda < 1$.

Geometrically, a natural consequence of Def. 2.4.1 is the statement that in the space of all CPTP maps there cannot exist a ball of valid quantum channels surrounding an extremal map, as then the above decomposition could always be constructed. Hence, as schematically depicted in Fig. 2.3(**a**), extremal channels can lie within the space of all quantum channels only at the *boundaries*, which further cannot be "flat", as this would still allow for a decomposition into a mixture of other CPTP maps.

Following the same argumentation as in Note 2.6, Def. 2.4.1 can be directly translated onto the space of CJ matrices and redefined in terms of their convex combinations. Hence, a natural class of extremal CPTP maps may be specified by considering the ones that yield CJ matrices of rank one, i.e. lead to pure $\Omega_\Lambda = |\psi\rangle\langle\psi|$ in Eq. (2.24). As this may only occur when $|\psi\rangle = U \otimes \mathbb{I}|\mathbb{I}\rangle$ (pure CJ matrices (2.24) correspond to unitary channels and vice versa [Bengtsson and Życzkowski, 2006]), all *unitary* quantum channels serve as examples of extremal maps. From the geometrical perspective (see Fig. 2.3(**a**)), the unitary maps $\mathcal{U}$ being continuously parametrisable must not only lie at the boundary of the channel space, but also should intuitively form a "smooth convex surface" containing the (trivial unitary) identity map $\mathcal{I}$.

Notice that if we (incorrectly) assumed the CP condition to be sufficient for a quantum channel to be physical, there would not exist any other extremal channels, as in the space of positive semi-definite density operators, $\Omega_\Lambda \ge 0$, pure states are the only extremal points [Bengtsson and Życzkowski, 2006]. Yet, one must not forget also to impose the TP property corresponding to an extra linear constraint on the CJ matrices: $\mathrm{Tr}_{\mathrm{S}}\{\Omega_\Lambda\} = \mathbb{I}^{\mathrm{A}}$, which geometrically selects a hyperplane of CPTP maps in a larger space

---

[26]Here, normalised to $\mathrm{Tr}\{\Omega\} = d_{\mathrm{in}}$ rather to unity, what, however, is irrelevant.

(a)

(b)
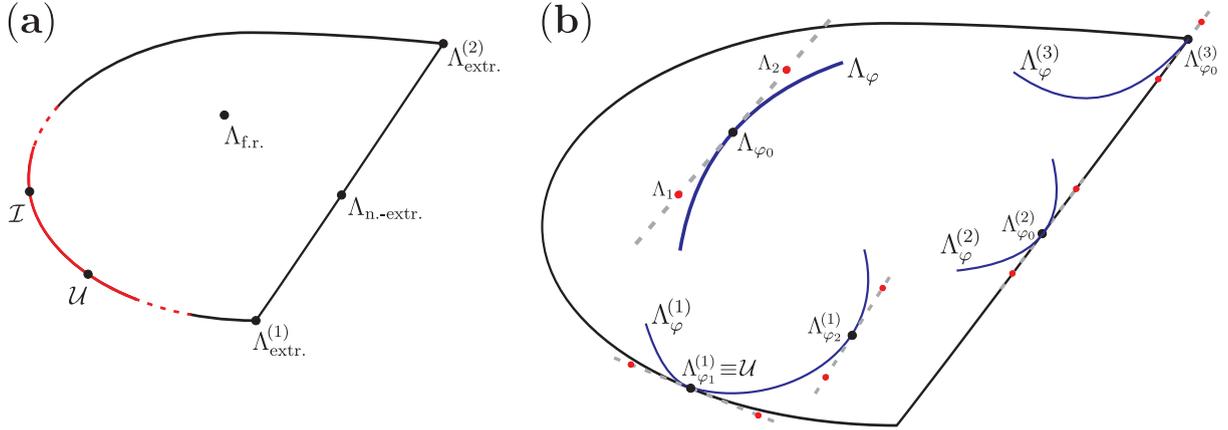
FIGURE 2.3: **Geometry of quantum channels – extremality and φ-extremality**.
The *convex* space of quantum channels is depicted as an oval shape with one of the sides cut to schematically represent the consequences of the TP constraint selecting a hyperplane of CPTP maps in a larger space of CP maps.
(a) **Extremality** property analysed for various channels.
$\mathcal{U}$ – *unitary* channel, being an *extremal* map, lies within a convex surface of unitaries (*red*) that also contain the identity map $\mathcal{I}$; $\Lambda_{\text{f.r.}}$ – exemplary *full-rank* channel is naturally *non-extremal*, as it lies strictly inside the channel space; $\Lambda_{\text{extr.}}^{(1/2)}$ – non-unitary channels that are *extremal*—not being decomposable into a mixture of CPTP maps; $\Lambda_{\text{n.-extr.}}$ – *non-full-rank* channel which is located at a "flat" boundary, what still makes the map *non-extremal* (e.g. may be expressed as a mixture of $\Lambda_{\text{extr.}}^{(1/2)}$).
(b) φ-**extremality** property studied for four φ-parametrised families of CPTP maps: Λ_φ, $\{\Lambda_\varphi^{(i)}\}_{i=1}^3$.
Λ_φ (*thick blue*) – presented to explicitly show the tangential decomposition specified in Def. 2.4.3 that yields the channel to be φ-*non-extremal* at φ_0. As Λ_{φ_0} is *full-rank* such a decomposition is possible irrespectively of the direction at which the curve crosses it; $\Lambda_\varphi^{(1)}$ – at φ_1 the channel is unitary and thus extremal, what implies φ-*extremality*. Yet, for any other parameter value, e.g. φ_2, it is φ-*non-extremal*; $\Lambda_\varphi^{(2)}$ – at φ_0 the channel ceases to be full-rank, but as the boundary is "flat" in the direction of $\partial_\varphi \Lambda_\varphi|_{\varphi_0}$ $\Lambda_{\varphi_0}^{(2)}$ is φ-*non-extremal*; $\Lambda_\varphi^{(3)}$ – at φ_0 the channel becomes extremal and hence φ-*extremal*.

of CP maps. This may lead to non-trivial facets of the space of quantum channels, what we schematically represent in Fig. 2.3 by a flat fragment of the boundary. Importantly, it is thus the TP-property that is responsible for the existence of non-unitary extremal CPTP maps. In general, such non-trivial extremal channels may be identified by utilising the *Choi criterion* that—as proven in App. A following [Choi, 1975]—is equivalent to the channel-extremality definition (Def. 2.4.1), but also provides a recipe of how to verify if a given channel is extremal basing on its Kraus representation:

Criterion 2.4.2 (Choi criterion for channel extremality). Given a quantum channel Λ of rank r and a set of its linearly independent Kraus operators $\{K_i\}_{i=1}^r$, Λ is *extremal* if and only if $\left\{K_i^\dagger K_j\right\}_{ij}$ is a set of r^2 *linearly independent* matrices.

On other hand, we term all quantum channels that do *not* fulfil the above Choi criterion—or equivalently, by Def. 2.4.1, *are* decomposable into a convex sum of CPTP maps—to be *non-extremal*.

φ-extremal channels
A family of CPTP maps $\{\Lambda_\varphi\}_\varphi$, where φ is the latent parameter to be estimated in a metrological scenario, geometrically corresponds to a *curve* in the space of quantum channels (*solid blue* line(s) in Fig. 2.3(**b**)). Crucially, the nature of a quantum metrological problem intrinsically defines a "sense of

direction" along which the parameter of interest varies. Hence, we naturally adapt the concept of channel extremality specified in Def. 2.4.1, so that it encapsulates such notion of parameter-induced direction:

Definition 2.4.3 (φ-extremality of a quantum channel)**.** An element of a family of CPTP maps $\{\Lambda_\varphi\}_\varphi$ is φ-*extremal* at a given φ_0 if and only if there do *not* exist any distinct CPTP maps $\Lambda_{1/2}$ that lie in the space of quantum channels along the line *tangential* at φ_0 to the curve representing the family, and yield $\Lambda = \lambda\,\Lambda_1 + (1-\lambda)\Lambda_2$ for some $0 < \lambda < 1$.

Although the above definition possesses a neat geometrical interpretation, which we explicitly depict in Fig. 2.3(**b**) (*thick blue line*), it can be formalised with help of the CJ-matrix representation (2.24). Notice that the tangential direction in the space of quantum channels (and equivalently CJ matrices) is specified at a given φ_0 by the derivative[27] of the CJ matrix: $\dot\Omega_{\Lambda_{\varphi_0}} \equiv \partial_\varphi \Omega_{\Lambda_\varphi}\big|_{\varphi=\varphi_0}$ (which then by Eq. (2.25) also defines the derivative of the channel $\partial_\varphi \Lambda_\varphi\big|_{\varphi=\varphi_0}$). Thus, all the maps lying along the tangent (i.e. along a given *dashed grey line* in Fig. 2.3(**b**)) may be defined as the ones with CJ matrices reading $\Omega_{\Lambda_{\varphi_0}} + \epsilon\,\dot\Omega_{\Lambda_{\varphi_0}}$ for any $\epsilon \in \mathbb{R}$. On the other hand, the construction of the convex decomposition in Def. 2.4.3 is possible, if one may "follow" the tangential curve by *any* (even infinitesimally small) distances in *both* directions away from Λ_{φ_0}, while remaining within the space of CPTP maps. Hence, for this *not* to be true, so that Λ_φ is φ-extremal at φ_0 (and *red dots* in Fig. 2.3(**b**) lie outside the channel space), there must not exist $\epsilon > 0$ such that both $\Omega_{\Lambda_{\varphi_0}} \pm \epsilon\,\dot\Omega_{\Lambda_{\varphi_0}}$ are positive semi-definite. As explicitly shown in App. B, such a statement may be further reformulated to define the criterion for φ-extremality as:

Criterion 2.4.4 (Criterion for channel φ-extremality)**.** An element of a family of CPTP maps $\{\Lambda_\varphi\}_\varphi$ is φ-*extremal* at a given φ_0 if and only if the derivative of its CJ matrix, $\dot\Omega_{\Lambda_{\varphi_0}}$, is *not* contained within the support of the CJ matrix $\Omega_{\Lambda_{\varphi_0}}$.

Similarly to non-extramal channels, we may also define the class of φ-*non-extremal* CPTP maps, i.e. all that do *not* satisfy Crit. 2.4.4. Then, as φ-non-extremality guarantees existence of a valid decomposition (one along the tangent), any φ-non-extremal map must naturally be non-extremal. Equivalently, an extremal channel is by definition φ-extremal. We prove explicitly these statements in App. B by showing that Def. 2.4.1 and Crit. 2.4.2 indeed imply respectively Def. 2.4.3 and Crit. 2.4.4. However, let us clearly emphasise that, as the notion of φ-extremality strongly depends on the geometry of a particular channel family, *there exist φ-extremal channels which are non-extremal* (see e.g. the loss noise model analysed in the following section).

Note 2.7: Geometry of full-rank quantum channels.

Lastly, let us discuss the *full-rank* quantum channels that, as defined in Sec. 2.2.1, possess a maximal number of linearly independent Kraus operators ($r = d_{\mathrm{in}}^2$). In the CJ-matrix picture (2.24), such condition translates onto the statement that *all* the eigenvalues of Ω_Λ are strictly greater than zero. Geometrically, this corresponds exactly to the situation in which one may construct a ball of valid CPTP maps surrounding Λ of interest, or loosely speaking, one may "move away" from Λ in any direction without crossing any boundary of the channel space (what occurs when one of the eigenvalues of Ω_Λ changes sign). Hence, as indicated in Fig. 2.3(**a**), *the full-rank channels are the ones that lie strictly inside the space of CPTP maps*, and thus are naturally *non-extremal*. Furthermore, considering a family $\{\Lambda_\varphi\}_\varphi$ and a full-rank channel Λ_{φ_0} at some φ_0, it must be also φ-*non-extremal* irrespectively of the parameter-induced geometry. Due to the presence of valid CPTP maps in any direction away from Λ_{φ_0}, no matter what form the derivative $\partial_\varphi \Lambda_\varphi\big|_{\varphi=\varphi_0}$ takes, the convex decomposition along the tangent is always possible. Formally, one may directly see that Crit. 2.4.4 is indeed fulfilled in the CJ-matrix picture, as $\Omega_{\Lambda_{\varphi_0}}$ is full-rank (its eigenvectors span the whole space) so that it definitely supports $\dot\Omega_{\Lambda_{\varphi_0}}$.

Noise model:	*Dephasing*	*Depolarization*	*Loss*	*Spontaneous emission*
Rank:	2	4 (full-rank)	3	2
Extremal:	no	no	no	**yes**
φ-extremal:	no	no	**yes**	**yes**
Geometric interpretation:				
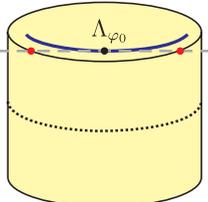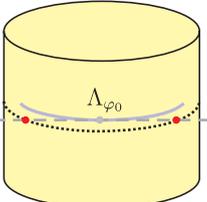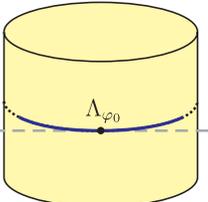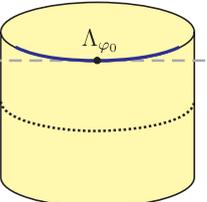
TABLE 2.2: **Geometrical properties of the noisy-phase–estimation models** depicted in Fig. 2.2. The channel-ranks are dictated by the pure-noise maps specified in Tab. 2.1, so that only the *depolarisation* channel corresponds to a full-rank CPTP map. The *loss* and *spontaneous emission* models yield *φ-extremal* channels, where in the latter case this is a consequence of the noise-map being *extremal*. In the last raw, we present an intuitive representation depicting geometry of the curve dictated by each channel family, i.e. its schematic location in the space all CPTP maps.

2.4.3 EXAMPLE: **Noisy-phase–estimation channels**

As an example, let us return to the qubit evolution models introduced in Sec. 2.2.3 (i.e. the noisy-phase-estimation channels depicted in Fig. 2.2) and discuss their geometrical properties, in particular, verifying the notions of extremality and φ-extremality. For each model, we consider the overall channel $\Lambda_\varphi = \mathcal{U}_\varphi \circ \mathcal{D}_\eta$ defined in Eq. (2.17), with an adequate pure-noise map $\mathcal{D}_\eta$ specified in Tab. 2.1. We summarise the results in Tab. 2.2, but let us already note that due to the *unitary* parameter-encoding the channels enjoy a circular symmetry[28], so that their geometrical properties are φ-independent. That is why, we are able to *schematically* represent the space of CPTP maps containing all[29] four channels of interest as a cylinder, where the curve $\{\Lambda_\varphi\}_\varphi$ in all cases corresponds to a circle that lies in a horizontal plane being symmetric around the vertical axis. In the last row of Tab. 2.2, we present such a metaphorical "geometric interpretation" for each noise-type, where in each case we intuitively draw a segment of the curve $\{\Lambda_\varphi\}_\varphi$ in agreement with the geometric features we obtain.

In order to verify extremality of the channels, we apply the Choi criterion (Crit. 2.4.2) to each set of Kraus operators specified in Tab. 2.1, and come to conclusion that only the *spontaneous emission* (*amplitude damping*) model yields an *extremal* CPTP map [Bengtsson and Życzkowski, 2006], which therefore is also *φ-extremal*. In the "cylinder-representation", we thus choose its corresponding curve to coincide with the upper edge of the cylinder, what adequately disallows any convex decomposition of Λ_{φ_0} to be legal. Secondly, we verify the φ-extremality criterion (Crit. 2.4.4) to find out that also the *loss* noise-type

---

[27]Throughout this work, by derivative of a matrix/vector we simply mean the matrix/vector obtained after computing the derivatives of all the entries.

[28]In fact, the *U(1) symmetry of phase* (which we discuss in more detail later in Sec. 3.1.3.3), as the parameter φ appears in the CJ matrices of all four channels only through the $e^{\pm i\varphi}$ factors.

[29]To be precise, the loss channel is a qubit-qutrit map and thus belongs to a different space of qubit-qutrit channels. Yet, this does not stop us to *intuitively* interpret such a space in Tab. 2.2 in the same manner as for the other models.

fulfils it, what thus proves that the loss model yields a φ-*extremal* but *non-extremal* channel. Hence, we interpret such a fact by drawing its curve on the cylinder in a way, so that it is decomposable in the vertical direction, but not along the parametrisation-induced tangent. The *depolarisation* channel is the simplest one to analyse, as it constitutes a full-rank map. Thus, for all φ the corresponding Λ_φ lie strictly inside the cylinder, making the map φ-*non-extremal* (and hence *non-extremal*). On the other hand, the *dephasing* noise model is *not* full-rank and therefore must be located on the boundary of the channel space. Yet, as we verify that it is φ-*non-extremal*—its $\Omega_{\Lambda_{\varphi_0}}$ supports $\dot{\Omega}_{\Lambda_{\varphi_0}}$ in accordance with Crit. 2.4.4—such boundary must be flat in the tangential direction, i.e. we draw its curve on the top facet of the cylinder.

Chapter 3

Limits to precise estimation of latent parameters

3.1 Classical estimation theory

Before going into details of quantum mechanical aspects of metrology, we review the fundamentals of estimation theory that lies at the heart of any, also classical, metrological problem. The essential question that has been addressed by statisticians long before the invention of quantum mechanics is how to most efficiently extract information from a given data set, which is determined by some non-deterministic process. In particular, if there exists a (global) latent parameter that affects the measurement outcomes collected, e.g. temperature at which an experiment is performed or strength of a magnetic field distorting the electromagnetic signal measured, to what extent is one able to determine its value basing on the gathered sample of data. This issue is normally termed as the *problem of parameter estimation* [Kay, 1993; Lehmann and Casella, 1998].

3.1.1 The parameter estimation problem

Mathematically, the parameter estimation problem corresponds to the situation, in which we are given an N-point data set $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ representing a realisation of N *independent* identically-distributed random variables, X^N, each distributed according to a common *Probability Density Function* (PDF), $p_\varphi(X)$, that depends on an unknown parameter φ we wish to determine. Our goal is to construct an *estimator* $\tilde{\varphi}_N(\mathbf{x})$ which should be interpreted as a function that outputs the most accurate estimate of the parameter φ based on a given data set. Importantly, as the estimator $\tilde{\varphi}_N$ is built on random data, it is a random variable itself and its statistical properties, such as the mean or the variance, are dictated by the data statistics, i.e. the collective factorisable PDF: $p_\varphi(\mathbf{x}) = \prod_{i=1}^{N} p_\varphi(x_i)$.

Typically, two approaches to the above problem are undertaken depending on the probabilistic nature of the estimated parameter. In the so called *frequentist* approach, φ is assumed to be a *deterministic* variable with a fixed value that, if known, could in principle be stated to any precision. Moreover, as the sample size is taken to be large enough that the frequencies of the outcomes approximate well their probabilities—hence, the name of the approach—it is presumed that without loss of generality any estimation protocol may be taken to be *local*, i.e. designed for a particular value of φ. In contrast,

when following the *Bayesian* paradigm, the estimated parameter is a *random* variable itself, so that the estimation protocol has to apply *globally*, i.e. it must be optimised for a given range of values the parameter may take. In such a case, the intrinsic fluctuations of φ account for the lack of knowledge about the parameter we possess *prior* to performing the estimation. We describe both approaches in detail below.

3.1.2 Frequentist approach – *local* estimation of a *deterministic* parameter

3.1.2.1 Imposing the unbiasedness and minimising the Mean Squared Error

As within the *frequentist* approach the estimated parameter is assumed to be a deterministic variable of fixed value, we may write the variance of any given estimator $\tilde{\varphi}_N(\mathbf{x})$ built on an N-point data sample as

$$\mathrm{Var}[\tilde{\varphi}_N]\big|_\varphi = \left\langle \left(\tilde{\varphi}_N(\mathbf{x}) - \langle \tilde{\varphi}_N \rangle_\varphi \right)^2 \right\rangle_\varphi = \int \mathrm{d}^N x \, p_\varphi(\mathbf{x}) \left(\tilde{\varphi}_N(\mathbf{x}) - \langle \tilde{\varphi}_N \rangle_\varphi \right)^2, \qquad (3.1)$$

where φ is the true value of the parameter and $\langle \tilde{\varphi}_N \rangle_\varphi = \int \mathrm{d}^N x \, p_\varphi(\mathbf{x}) \, \tilde{\varphi}_N(\mathbf{x})$ is the mean value of the estimator. We term an estimator to be *consistent*, if it outputs with certainty the correct value of the estimated parameter when the sample size is infinitely increased, i.e.

$$\lim_{N\to\infty} \tilde{\varphi}_N(\mathbf{x}) = \varphi, \qquad (3.2)$$

where taking the asymptotic N limit should be understood as convergence in probability, so that the distribution of $\tilde{\varphi}_N$ becomes infinite-narrowly peaked around the nominal value φ. In particular, the consistency of the estimator implies that its mean and variance must converge asymptotically to $\lim_{N\to\infty} \langle \tilde{\varphi}_N \rangle_\varphi = \varphi$ and $\lim_{N\to\infty} \mathrm{Var}[\tilde{\varphi}_N]\big|_\varphi = 0$ respectively. On the other hand, the performance of any estimator is quantified by the *Mean Squared Error* (MSE), i.e. the average *squared distance* of $\tilde{\varphi}_N$ from the true value φ:

$$\Delta^2 \tilde{\varphi}_N \big|_\varphi = \left\langle \left(\tilde{\varphi}_N(\mathbf{x}) - \varphi \right)^2 \right\rangle_\varphi = \int \mathrm{d}^N x \, p_\varphi(\mathbf{x}) \left(\tilde{\varphi}_N(\mathbf{x}) - \varphi \right)^2, \qquad (3.3)$$

which is then minimised by an *optimal* estimator.

Let us emphasise that such an optimal estimator may turn out to be *local*—optimised for a particular value of the parameter—as in general there may not exist a *single* estimator minimising the MSE (3.3) for *all* values of φ. As a matter of fact, such a local estimator seems to be useless from the practical point of view, as it requires the estimated parameter to be exactly known before the estimation procedure! However, as any realistic, or in other words *global*, estimator can perform only worse at a given φ than the estimator specially optimised for that parameter value, we may limit the performance of any potential strategy by establishing the ultimate bounds on precision achieved by the local estimators. Moreover, the issue of locality becomes less and less important with growth of the sample size and in particular may be fully ignored when investigating protocols in the asymptotic N limit, in which the local—frequentist—precision bounds are guaranteed to be saturable [van der Vaart, 1998]. We discuss the consequences of this counter-intuitive paradigm in Sec. 3.1.2.3 below, but one should already bear in mind that such *local estimation regime* is always presumed within the frequentist approach, what has then strong implications on its applicability to any real-life problems. On the other hand, if luckily a *global* estimator may be constructed that minimises the MSE for any parameter value, then trivially it is also *locally* optimal for any φ.

The minimisation of Eq. (3.3) is addressed within the frequentist framework in two steps. When considering *global* estimators, they are firstly restricted to be *unbiased*, so that they always output on average the true parameter value by fulfilling for any φ the condition

$$\langle \tilde{\varphi}_N \rangle_\varphi = \int \mathrm{d}^N x \, p_\varphi(\mathbf{x}) \, \tilde{\varphi}_N(\mathbf{x}) = \varphi. \tag{3.4}$$

As a result, the Eqs. (3.1) and (3.3) become equivalent and the minimisation of the MSE is then tantamount to minimising the variance of a given unbiased estimator. Yet again, as we are allowed to consider and focus on the case of local estimators, we may relax the *unbiasedness constraint* (3.4), so that it holds only up to $O(\delta\varphi^2)$ in the vicinity of particular parameter true value chosen, say φ_0 with $\varphi = \varphi_0 + \delta\varphi$. Eq. (3.4) then fixes only the mean value of the estimator and its differential w.r.t. φ:

$$\langle \tilde{\varphi}_N \rangle_{\varphi_0} = \varphi_0 \,, \qquad \left. \frac{\partial \langle \tilde{\varphi}_N \rangle}{\partial \varphi} \right|_{\varphi_0} = 1. \tag{3.5}$$

However, out of the above constraints only the second one has a non-trivial meaning, because knowing the true value φ_0 while estimating locally we can always adequately shift the estimator and satisfy the first one. Thus effectively, the *local unbiasedness constraints* (3.5) fix only the "speed" of change of the estimator mean with the parameter at its true value. Importantly, any *global* unbiased estimator satisfying Eq. (3.4) trivially satisfies the *local* conditions (3.5) at any φ_0, so that we may utilise Eq. (3.5) in the following section to bound the precision of *any* unbiased estimator.

Yet, before doing so, let us remark that by imposing either of the unbiasedness constraints, (3.4) or (3.5), the frequentist approach naturally excludes all the *biased* estimators that potentially may not only be more accurate but also be the only ones minimising the MSE (3.3). However, as we require any estimator to be *consistent*, by averaging Eq. (3.2) over the outcomes one should realise that such a practical restriction also forces any estimator to be *unbiased* in the asymptotic N limit. Hence, although the biased estimators may lead to improved precision beyond the scope of the frequentist approach for finite sample sizes, they may be ignored in the $N \to \infty$ limit for which the frequentist approach is really designed for.

3.1.2.2 Ultimate precision and the Cramér-Rao Bound

Stemming from the local unbiasedness condition on the "speed" of change of the estimator mean (3.5) and assuming the single-outcome PDF, $p_\varphi(X)$, to be *regular* at a given φ_0, i.e.[1]

$$\left\langle \frac{\partial}{\partial \varphi} \ln p_\varphi \right\rangle_{\varphi_0} = 0 \quad \Longrightarrow \quad \int \mathrm{d}x \left[p_\varphi(x) \frac{\partial \ln p_\varphi(x)}{\partial \varphi} \right]_{\varphi_0} = \int \mathrm{d}x \left. \frac{\partial p_\varphi(x)}{\partial \varphi} \right|_{\varphi_0} = \left[\frac{\partial}{\partial \varphi} \int \mathrm{d}x \, p_\varphi(x) \right]_{\varphi_0} = 0, \quad (3.6)$$

one may construct by means of a Cauchy-Schwarz inequality—see e.g. [Kay, 1993] for the derivation—the so-called *Cramér-Rao Bound* (CRB) that lower-limits the MSE of any unbiased estimator:

$$\left. \Delta^2 \tilde{\varphi}_N \right|_{\varphi_0} \geq \frac{1}{N \, F_{\mathrm{cl}}[p_\varphi]\big|_{\varphi_0}}. \tag{3.7}$$

Although we have specified the CRB as a bound on the MSE (3.3) which is the adequate figure of merit, one should bear in mind that due to unbiasedness $\left. \Delta^2 \tilde{\varphi}_N \right|_{\varphi_0} = \mathrm{Var}[\tilde{\varphi}_N]\big|_{\varphi_0}$ and the CRB equivalently

---

[1]Crucially, the regularity assumption allows to interchange the order of $\int \mathrm{d}x$ and $\partial/\partial\varphi$ in any expression averaged over the outcomes of X at a particular parameter value φ_0.

lower-limits the variance (3.1) of the estimator, which is the more experimentally relevant quantity that may be determined for an unknown φ basing only on the data gathered. We have also kept the notation $\ldots|_{\varphi_0}$ in Eq. (3.7), in order to stress that both its l.h.s. and r.h.s. may in principle depend on φ_0, as the bound is derived fixing a particular value of the estimated parameter. Nevertheless, the CRB applies also to *global* unbiased estimators which form does not vary with φ. The crucial quantity limiting ultimately the MSE in Eq. (3.7) is the so-called (classical) *Fisher Information* (FI), F_{cl}, that can be expressed using one of the equivalent formulae below:

$$F_{\mathrm{cl}}[p_\varphi] = \int \mathrm{d}x \, \frac{1}{p_\varphi(x)} \left[\frac{\partial p_\varphi(x)}{\partial \varphi} \right]^2 = \left\langle \left(\frac{\partial}{\partial \varphi} \ln p_\varphi \right)^2 \right\rangle = -\left\langle \frac{\partial^2}{\partial \varphi^2} \ln p_\varphi \right\rangle. \tag{3.8}$$

Being *non-negative* and *additive* the FI has the interpretation of an *information measure* [Arndt, 2001] which increase indicates a higher precision potentially achievable by estimation protocols. In particular, at a given φ_0, $F_{\mathrm{cl}}[p_\varphi]\big|_{\varphi_0} = 0$ proves that one cannot extract any information about the parameter from a sample, whereas divergent $F_{\mathrm{cl}}[p_\varphi]\big|_{\varphi_0} = \infty$ implies that the true value φ_0 can in principle be perfectly determined, what, however, is guaranteed only in the asymptotic N limit, as discussed in Secs. 3.1.2.4 and 3.1.2.5 below. The non-negativity of the FI follows naturally from the second expression in Eq. (3.8) and its additivity can be easily verified by using the last expression of Eq. (3.8), in order to prove that $F_{\mathrm{cl}}\big[p_\varphi^{(1,2)}\big] = F_{\mathrm{cl}}\big[p_\varphi^{(1)}\big] + F_{\mathrm{cl}}\big[p_\varphi^{(2)}\big]$ for any factorisable $p_\varphi^{(1,2)}(X_1, X_2) = p_\varphi^{(1)}(X_1)\, p_\varphi^{(2)}(X_2)$. Importantly, when dealing with *independently* distributed samples $F_{\mathrm{cl}}\big[p_\varphi^N\big] = N\, F_{\mathrm{cl}}[p_\varphi]$, what indeed leads to the CRB (3.7) being fully determined by the distribution of the single random variable X and, most importantly, to the *SQL-like scaling* of $1/N$ for the MSE.

Note 3.1:   Central Limit Theorem – asymptotic estimation of the PDF mean.
As aside, let us note that the additivity property of FI is consistent with the natural intuition one may infer from the *Central Limit Theorem* (CLT). One may look at the CLT as a special kind of an estimation problem in which the mean, μ, of a distribution (of a random variable X) is treated as the parameter to be determined, $\varphi \equiv \mu$, whereas the sample average, $\tilde\varphi_N(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i$, constitutes an example of a global unbiased estimator that always saturates the CRB (3.7) in the asymptotic N limit. According to the CLT, the PDF of such an estimator converges to a Gaussian distribution with $\Delta^2\tilde\varphi_N \overset{N\to\infty}{=} \mathrm{Var}[X]/N$. Hence, $\Delta^2\tilde\varphi_N$ saturates asymptotically the CRB, as the FI (3.8) calculated with respect to the mean of any Gaussian distribution corresponds to the inverse of its variance:

$$F_{\mathrm{cl}}\left[\sim \exp\left(-\frac{(x-\varphi)^2}{2\,\mathrm{Var}[X]} \right) \right] = -\left\langle \frac{\partial^2}{\partial \varphi^2}\left[-\frac{(x-\varphi)^2}{2\,\mathrm{Var}[X]} \right] \right\rangle = \frac{1}{\mathrm{Var}[X]}. \tag{3.9}$$

Lastly, let us remark that the FI is a *local quantity*, what should be expected acknowledging the fact that it was derived basing on the local unbiasedness conditions (3.5) valid up to $O(\delta\varphi^2)$ around a given φ_0. Consequently, as explicitly stated in the first expression of the definition (3.8), the FI at a given parameter value, $F_{\mathrm{cl}}[p_\varphi]\big|_{\varphi_0}$, is dependent only on $p_{\varphi_0}(X)$ and $(\partial p_\varphi(X)/\partial\varphi)\big|_{\varphi_0}$, so that it is fully specified by just fixing at φ_0: the PDF and its "speed" of change with φ. Crucially, this means that *all* parametrised PDFs of X that coincide up to $O(\delta\varphi^2)$ with one another at φ_0 are *equivalent* from the point of view of their FI. In fact, the locality is also a consequence of the FI possessing a neat geometric interpretation [Amari and Nagaoka, 2007]. Expanding the *angular distance*[2] [Amari *et al.*, 1987; Wootters, 1981] between the neighbouring PDFs $p_{\varphi_0}(X)$ and $p_{\varphi_0+\delta\varphi}(X)$ around φ_0, which is defined as $D_{\mathrm{cl}}(p_1, p_2) = \arccos[\mathsf{Fid}_{\mathrm{cl}}(p_1, p_2)]$ with $\mathsf{Fid}_{\mathrm{cl}}(p_1, p_2) = \int\mathrm{d}x\sqrt{p_1(x)\,p_2(x)}$ being the *fidelity* [Nielsen and Chuang, 2000] of PDFs $p_{1/2}$, we obtain for $\delta\varphi \geq 0$:

$$D_{\mathrm{cl}}(p_{\varphi_0}, p_{\varphi_0+\delta\varphi}) = \frac{1}{2}\sqrt{F_{\mathrm{cl}}[p_\varphi]\big|_{\varphi_0}}\, \delta\varphi + O(\delta\varphi^2). \tag{3.10}$$

---

[2] Also known as the Bhattacharyya distance, for which the fidelity is termed as the Bhattacharyya coefficient [Bengtsson and Życzkowski, 2006].

Hence, the FI (3.8) may be equivalently interpreted as the square of the speed, $F_{\mathrm{cl}}[p_\varphi] = 4\left(\mathrm{d}D_{\mathrm{cl}}/\mathrm{d}\varphi\right)^2$, with which p_φ is "moving" along the path of φ at a particular parameter value.

3.1.2.3 Locality of the frequentist approach and its consequences

When analysing performance of realistic estimation protocols, two issues may arise when utilizing the CRB to quantify the accuracy of any practical, hence *global*, estimator chosen. Firstly, in a more general case than the Gaussian distribution mean estimation (3.9), the FI (3.8) and hence the CRB (3.7) may depend on the true value φ_0. Thus, it is then ambiguous which parameter value to substitute into the CRB to most accurately bound the precision achieved given particular sample of data and a global unbiased estimator. Secondly, even when the evaluated FI turns out to be parameter-independent, one must always verify if the CRB actually corresponds to an inequality which may be saturated by a *single* estimator at any φ. In the following section, we show that for any finite N a global unbiased estimator, which saturates the CRB irrespectively of the parameter value, exists *only* for PDFs belonging to a subclass of the *exponential probability distributions*. Hence, for a general $p_\varphi(X)$, unless we are dealing with very large samples, the CRB should be treated with caution.

On the other hand, we show in Sec. 3.1.2.4 that *locally* one can always[3] construct an estimator that saturates the CRB (3.7). From the physical perspective, this means that the CRB is always meaningful when the goal is really to design an estimator that is most sensitive to small deviations from a known value of the parameter. This corresponds to the situation when one has investigated perfectly all the properties of a given physical system that afterwards is subjected to some external fluctuations, which vary the parameter of interest by a small amount. Intuitively, this must be the most optimistic scenario of estimation that one may consider, so it is consistent that the CRB defines the ultimate bound on the achievable precision. Moreover, this fact also explains why in the situation when the parameter is unknown one may attain the CRB by infinitely increasing the sample size N. Then, one may always use a fraction of the outcomes in order to learn the value of the estimated parameter sufficiently enough to enter the local estimation regime, so that the CRB becomes more and more accurate as $N \to \infty$.

Unfortunately, the frequentist approach does not give a recipe how to quantify N for which the local regime of estimation may already be assured, and thus the accuracy of the CRB as a bound on precision. Yet, as shown in Eq. (3.9) with the example of application of the CLT to estimation, the CRB (3.7) represents a tight inequality when the independently distributed data statistics approach their corresponding asymptotic Gaussian distribution. Such phenomenon is true in general for any (asymptotically) unbiased estimator defined on *independently* distributed data and is known under the name of the Local Asymptotic Normality [van der Vaart, 1998]. Hence, one way to quantify the speed at which the precision of the optimal strategy approaches the CRB with N is to determine the rate at which the overall PDF of the sampled data converges to its asymptotic Gaussian form. Nevertheless, the method which bases on and is constructed to account for the progressive improvement of knowledge about the parameter with growth of N is really the Bayesian inference which we discuss in Sec. 3.1.3.

3.1.2.4 Saturability of the CRB

An unbiased estimator that saturates the CRB (3.7) is said to be *efficient* in that it efficiently uses the sampled data. As the CRB is derived by means of the Cauchy-Schwarz inequality, the sufficient and

---

[3]Yet, one must be careful when dealing with parameters that do not carry a standard *topology* of a real line, e.g. in phase estimation problems in which the parameter is an element of the circle group (see for instance Sec. 3.1.4.1).

necessary condition for its saturability, which is imposed on the PDF and a *global* unbiased estimator, corresponds to the statement [Kay, 1993]:

$$\frac{\partial}{\partial\varphi}\ln p_\varphi(\mathbf{x}) = F_{\mathrm{cl}}\big[p_\varphi^N\big]\,(\tilde{\varphi}_N(\mathbf{x}) - \varphi)\,, \tag{3.11}$$

where the multiplicative factor on the r.h.s. becomes fixed to $F_{\mathrm{cl}}\big[p_\varphi^N\big] = N\,F_{\mathrm{cl}}[p_\varphi]$, so that: by differentiating Eq. (3.11) w.r.t. φ, averaging it over the outcomes, and applying the unbiasedness condition (3.4); we recover the FI definition (3.8) for p_φ^N. One should note that the above requirement ceases to make mathematical sense when $N \to \infty$, so the fact of (3.11) not being satisfiable for any finite N does not stop the CRB from being potentially tight in the asymptotic N limit.

Defining[4] $\ddot{\lambda}(\varphi) = \frac{\partial^2\lambda(\varphi)}{\partial\varphi^2} = F_{\mathrm{cl}}[p_\varphi]$, where $\lambda(\varphi)$ is some general, outcome-independent function, we may write the most general form of the PDF of X^N that satisfies Eq. (3.11) for *any* φ as

$$p_\varphi(\mathbf{x}) = \exp\big\{N\big[\dot{\lambda}(\varphi)\,(\tilde{\varphi}_N(\mathbf{x}) - \varphi) + \lambda(\varphi) + c_N(\mathbf{x})\big]\big\} \tag{3.12}$$

with $c_N(\mathbf{x})$ being an arbitrary, parameter-independent function. Moreover, as the data distribution is described with independently distributed random variables, the l.h.s. of Eq. (3.12) factorizes, so that $p_\varphi(\mathbf{x}) = \prod_{i=1}^{N} p_\varphi(x_i)$. Therefore, if there exists an efficient estimator for a single outcome, $\tilde{\varphi}(x)$ for $N = 1$, then the overall efficient estimator may be simply taken to be the mean of such individual estimators, $\tilde{\varphi}_N(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\tilde{\varphi}(x_i)$, where after choosing $c_N(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} c(x_i)$ the individual PDF must satisfy

$$p_\varphi(x) = \exp\big\{\dot{\lambda}(\varphi)\,(\tilde{\varphi}(x) - \varphi) + \lambda(\varphi) + c(x)\big\} \tag{3.13}$$

that is consistent with the general expression (3.12) after substituting $N = 1$.

The PDFs that possess the form (3.12), and hence allow for a global unbiased estimator that satisfies the criterion (3.11), belong to the so-called *exponential family* which is a well established class of PDFs in probability theory that encapsulates most common distributions such as: Gaussian, Bernoulli, gamma, chi-squared, binomial, Poissonian and many others [Kay, 1993]. In general, an exponential PDF reads:

$$p_\varphi^{\mathrm{exp}}(\mathbf{x}) = h(\mathbf{x})\,\exp\left\{N\left[\eta(\varphi)\,\frac{T(\mathbf{x})}{N} - A(\varphi)\right]\right\}\,, \tag{3.14}$$

where η, A and T, h are some standard parameters that characterise a given distribution and depend on the estimated parameter and the sample outcomes respectively[5]. Then, for the single-outcome, exponential PDF $p_\varphi^{\mathrm{exp}}(X)$ the FI (3.8) may be written as

$$F_{\mathrm{cl}}\big[p_\varphi^{\mathrm{exp}}\big] = \dot{\eta}(\varphi)\,\frac{\partial}{\partial\varphi}\left[\frac{\dot{A}(\varphi)}{\dot{\eta}(\varphi)}\right]\,. \tag{3.15}$$

Comparing Eqs. (3.12) and (3.14) it becomes evident that a legal global estimator for $p_\varphi^{\mathrm{exp}}(\mathbf{x})$ reads $\tilde{\varphi}_N^{\mathrm{exp}}(\mathbf{x}) = T(\mathbf{x})/N$, but for the validity of the CRB (3.7) it must also be assured to satisfy the unbiasedness constraint (3.4), which then naturally guarantees the CRB-saturability requirement (3.11) to be fulfilled. It is easy to show that for an exponential PDF of the form (3.14) the unbiasedness and CRB-saturability

---

[4]In order to shorten the notation, we represent the *derivatives w.r.t. the estimated parameter* throughout this work with '*overdots*', so that e.g. $\dot{\eta}(\varphi) \equiv \partial_\varphi\,\eta(\varphi)$, $\dot{p}_\varphi(x) \equiv \partial_\varphi\,p_\varphi(x)$, $\dot{\varrho}_\varphi \equiv \partial_\varphi\,\varrho_\varphi$, $|\dot{\psi}_\varphi\rangle \equiv \partial_\varphi|\psi_\varphi\rangle$ and $\dot{K}(\varphi) \equiv \partial_\varphi\,K(\varphi)$ etc.

[5]In statistics, the parameters of an exponential PDF are normally termed [Lehmann and Casella, 1998]: $h(\mathbf{x})$–base measure, $T(\mathbf{x})$–sufficient statistic, $\eta(\varphi)$–natural parameter, $A(\varphi)$–log-partition function.

conditions are met by such an estimator *if and only if*

$$\left\langle \tilde{\varphi}_N^{\mathrm{exp}} \right\rangle_\varphi = \int \mathrm{d}^N x \, p_\varphi^{\mathrm{exp}}(\mathbf{x}) \, \frac{T(\mathbf{x})}{N} = \frac{\dot{A}(\varphi)}{\dot{\eta}(\varphi)} = \varphi. \tag{3.16}$$

Furthermore, the expression (3.15) for the FI of $p_\varphi^{\mathrm{exp}}(X)$ simplifies then to $F_{\mathrm{cl}}[p_\varphi^{\mathrm{exp}}] = \dot{\eta}(\varphi)$ what becomes also clear when comparing Eqs. (3.12) and (3.14) and noting that $\eta(\varphi) = \dot{\lambda}(\varphi)$.

On the other hand, by just solving the CRB-saturability condition (3.11) for $\tilde{\varphi}_N(\mathbf{x})$, a *local* efficient estimator may always be constructed for *any*[6] PDF $p_\varphi(X)$ at a given true value φ_0, so that

$$\tilde{\varphi}_{N,\varphi_0}(\mathbf{x}) = \varphi_0 + \frac{1}{N \, F_{\mathrm{cl}}[p_\varphi]\big|_{\varphi_0}} \, \frac{\partial \ln p_\varphi(\mathbf{x})}{\partial \varphi}\bigg|_{\varphi_0}, \tag{3.17}$$

which then trivially satisfies Eq. (3.11) and as necessary fulfils the local unbiasedness conditions (3.5). Although the local estimator $\tilde{\varphi}_{N,\varphi_0}$ is not constructable when the second term in Eq. (3.17) is divergent, this may occur only for pathological parameter values for which the estimation problem is ill-defined, i.e.: when $\partial \ln p_\varphi(\mathbf{x})/\partial \varphi\big|_{\varphi_0}$ is infinite, what makes the PDF regularity assumption (3.6) invalid and the CRB (3.7) not applicable; or when $F_{\mathrm{cl}}[p_\varphi]\big|_{\varphi_0} = 0$ and no information about the parameter is extractable from the PDF. Lastly, while returning to the special case of exponential PDFs and substituting $p_\varphi^{\mathrm{exp}}(\mathbf{x})$ (3.14) into Eq. (3.17), we obtain the corresponding general form of a local efficient estimator:

$$\tilde{\varphi}_{N,\varphi_0}^{\mathrm{exp}}(\mathbf{x}) = \varphi_0 + \left[\frac{T(\mathbf{x})}{N} - \frac{\dot{A}(\varphi_0)}{\dot{\eta}(\varphi_0)} \right] \left[\frac{\partial}{\partial \varphi} \left[\frac{\dot{A}(\varphi)}{\dot{\eta}(\varphi)} \right]\bigg|_{\varphi_0} \right]^{-1}, \tag{3.18}$$

which may be verified to satisfy the local unbiasedness constraints (3.5) as required.

3.1.2.5   Maximum Likelihood estimator

As remarked in the previous section, although a global unbiased estimator may not exist that satisfies the saturability condition (3.11) for a given $p_\varphi(\mathbf{x})$, Eq. (3.11) loses its mathematical validity when $N \to \infty$ opening doors to the CRB (3.7) being potentially saturable. As a matter of fact, there exists an estimator, i.e. the *Maximal Likelihood* (ML) estimator, which always turns out to be efficient in the asymptotic N limit. The ML estimator is formally defined as

$$\tilde{\varphi}_N^{\mathrm{ML}}(\mathbf{x}) = \underset{\varphi}{\mathrm{argmax}} \, p_\varphi(\mathbf{x}) = \underset{\varphi}{\mathrm{argmax}} \, \ln p_\varphi(\mathbf{x}), \tag{3.19}$$

but intuitively should be understood as a function that for a given instance of outcomes, $\mathbf{x}$, outputs the value of parameter for which this data sample is most probable, i.e. the *likelihood function* $l_{\mathbf{x}_0}(\varphi) \equiv p_\varphi(\mathbf{x} = \mathbf{x}_0)$ is maximal. Although $\tilde{\varphi}_N^{\mathrm{ML}}$ is generally biased for finite N, it is unbiased asymptotically for any PDF: $\langle \tilde{\varphi}_N^{\mathrm{ML}} \rangle_\varphi \overset{N \to \infty}{=} \varphi$, so that the CRB then applies and, crucially, is always saturated, as also $\Delta^2 \tilde{\varphi}_N^{\mathrm{ML}}\big|_\varphi \overset{N \to \infty}{=} 1/(N F_{\mathrm{cl}}[p_\varphi])$ [Kay, 1993; Lehmann and Casella, 1998; van der Vaart, 1998]. However, similarly to the problem discussed in Sec. 3.1.2.3 of certifying sufficient sample size that assures locality of estimation, the frequentist approach does not give a recipe how to quantify N for which the ML estimator attains the CRB up to a certain accuracy.

---

[6]Yet, the estimator (3.17) may lead to inconclusive answers when estimating parameters *not defined on a real line*, i.e. *not* of standard topology, e.g. see the case of a circularly symmetric parameter discussed in Sec. 3.1.4.1.

As the logarithm is a monotonic function, in the last expression of Eq. (3.19) we equivalently consider the logarithm of the likelihood (the log-likelihood) to deduce the parameter value that maximizes the probability. Thus, in a typical situation when the log-likelihood possesses a single maximum, the ML estimator may be then interpreted as a solution to the equation

$$\frac{\partial}{\partial \varphi} \ln p_\varphi(\mathbf{x}) = 0, \tag{3.20}$$

which is warranted to indicate the maximum if also the condition $\frac{\partial^2}{\partial \varphi^2} \ln p_\varphi(\mathbf{x}) < 0$ is fulfilled at the φ considered.

Looking back at the CRB-saturability requirement (3.11), one should note that there always exists a parameter value—independently of the particular sampled data $\mathbf{x}$ obtained—$\varphi = \tilde{\varphi}_N(\mathbf{x})$ for which the r.h.s. of Eq. (3.11) vanishes. Importantly, at such φ, condition (3.11) becomes equivalent to the requirement (3.20) determining the ML estimator. Furthermore, the differential of Eq. (3.11) w.r.t. φ there reads

$$\left. \frac{\partial^2}{\partial \varphi^2} \ln p_\varphi(\mathbf{x}) \right|_{\varphi = \tilde{\varphi}_N(\mathbf{x})} = - N \left. F_{\mathrm{cl}}[p_\varphi] \right|_{\varphi = \tilde{\varphi}_N(\mathbf{x})}, \tag{3.21}$$

and, due to the non-negativity of the FI (3.8), guarantees that $\varphi = \tilde{\varphi}_N(\mathbf{x})$ corresponds to a maximum of the likelihood function. Hence, for a PDF that leads to $l_{\mathbf{x}}(\varphi)$ possessing a single maximum in φ, the ML estimator constructed by satisfying (3.20) must always be identical to the global efficient estimator satisfying Eq. (3.11) if such one exists. $\tilde{\varphi}_N^{\mathrm{ML}}$ fulfils then the condition (3.11) after choosing without loss of generality $\varphi = \tilde{\varphi}_N(\mathbf{x})$, which also is the parameter value outputted by the ML estimator (3.19). Therefore, in such a special case, the ML estimator is not only always unbiased, but also saturates the CRB for *any* N and not just in the asymptotic limit.

Note 3.2: ML estimator for the mean of a Gaussian PDF.

For instance, let us consider the mean estimation problem of N independent variables, X^N, distributed according to a Gaussian PDF: $p_\varphi(x) \sim \exp\left[\frac{-(x-\varphi)^2}{2\,\mathrm{Var}[X]}\right]$, for which the condition (3.20) reads

$$\frac{\partial}{\partial \varphi} \ln p_\varphi(\mathbf{x}) = \frac{\partial}{\partial \varphi} \sum_{i=1}^{N} \frac{-(x_i - \varphi)^2}{2\,\mathrm{Var}[X]} = \frac{N}{\mathrm{Var}[X]} \left(\frac{1}{N} \sum_{i=1}^{N} x_i - \varphi \right) = 0, \tag{3.22}$$

so that indeed the ML estimator (3.19) corresponds to the average of a sample, $\tilde{\varphi}_N^{\mathrm{ML}}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} x_i$, which is the global unbiased estimator saturating the CRB (3.7)—as discussed in Note 3.1.

Considering the case of N independent variables, X^N, each being distributed according to a PDF within the *exponential* family of distributions (3.14) satisfying the saturability constraint (3.16), the ML estimator must then correspond to $\tilde{\varphi}_N^{\mathrm{ML}}(\mathbf{x}) = T(\mathbf{x})/N$ that is the corresponding global efficient estimator.

3.1.2.6 Estimating a transformed parameter

As shown in Sec. 3.1.2.4, a given estimation problem may not allow for a global estimator to exist that fulfils the CRB-saturability criterion (3.11) unless the asymptotic N limit is considered, in which the ML estimator (3.19) is guaranteed to be efficient. However, it may happen that the same estimation problem may still be solvable efficiently with a global estimator regardless of N when estimating a *transformed parameter* $g(\varphi)$. One may explicitly prove, see e.g. [Kay, 1993], that for a general $g(\varphi)$ the CRB (3.7) transforms to

$$\left. \Delta^2 \tilde{g}_N \right|_{\varphi_0} \geq \frac{[\dot{g}(\varphi_0)]^2}{N \left. F_{\mathrm{cl}}[p_\varphi] \right|_{\varphi_0}}, \tag{3.23}$$

because the FI (3.8) under such parameter change just rescales, $F_{\mathrm{cl}}\big[p_{g(\varphi)}\big] = F_{\mathrm{cl}}[p_\varphi] / [\dot{g}(\varphi)]^2$, due to the chain rule $\partial/\partial\varphi \equiv \dot{g}(\varphi)\,\partial/\partial g$. As in the case of the CRB, the *transformed CRB* (3.23) lower-bounds the MSE, and hence the variance, of any unbiased estimator $\tilde{g}_N$ and may be saturated by a single estimator if the transformed version of criterion (3.11) is satisfied for any φ:

$$\frac{\partial}{\partial\varphi} \ln p_\varphi(\mathbf{x}) = N\, \frac{F_{\mathrm{cl}}[p_\varphi]}{\dot{g}(\varphi)}\, (\tilde{g}_N(\mathbf{x}) - g(\varphi)). \tag{3.24}$$

Thus, in particular, we may always fulfil the corresponding requirement (3.16) applicable to the exponential family of PDFs (3.14) and the estimator $\tilde{g}_N^{\mathrm{exp}}(\mathbf{x}) = T(\mathbf{x})/N$ by choosing the transformed parameter such that

$$\big\langle \tilde{g}_N^{\mathrm{exp}} \big\rangle_\varphi = \frac{\dot{A}(\varphi)}{\dot{\eta}(\varphi)} = g(\varphi), \tag{3.25}$$

what guarantees[7] $\tilde{g}_N^{\mathrm{exp}}$ to be a global estimator saturating the transformed CRB (3.23) for any N.

3.1.3   Bayesian approach – *global* estimation of a *stochastic* parameter

3.1.3.1   Average Mean Squared Error

Within the *Bayesian* approach, the estimated parameter φ is assumed to be a random variable that is distributed according to a *prior* PDF, $p(\varphi)$, representing the knowledge about φ one possesses before performing the estimation. Therefore, in contrast to the frequentist philosophy of Sec. 3.1.2, where the estimated parameter was assumed to have a fixed, well defined value, it is a particular *realisation* of the parameter that is really estimated in a real-life experiment. As a consequence, an optimal estimator must not only be *global* and minimise the MSE (3.3), but also has to take into account which values of φ are more probable according to $p(\varphi)$. Hence, such an estimator must minimise the *Average Mean Squared Error* ($\overline{\mathrm{MSE}}$):

$$\big\langle \Delta^2\tilde{\varphi}_N \big\rangle = \int\!\mathrm{d}\varphi\, p(\varphi)\, \Delta^2\tilde{\varphi}_N\big|_\varphi = \int\!\mathrm{d}\varphi\, p(\varphi)\!\int\!\mathrm{d}^N x\, p(\mathbf{x}|\varphi)\,(\tilde{\varphi}_N(\mathbf{x}) - \varphi)^2, \tag{3.26}$$

which is the MSE (3.3) averaged over the possible values of the estimated parameter. $p(\mathbf{x}|\varphi)$ is the PDF previously labelled as $p_\varphi(\mathbf{x})$ within the frequentist approach, which due to stochastic character of the parameter now represents a conditional probability predicting for a given realisation of φ the outcome statistics. Consequently, the $\overline{\mathrm{MSE}}$ may be interpreted also as the mean *squared distance* between the estimator and parameter realisations averaged according to the joined PDF $p(\mathbf{x}, \varphi)$, which is defined via the *Bayes' theorem*—hence the name of the approach—in two equivalent ways:

$$p(\mathbf{x}, \varphi) = p(\mathbf{x}|\varphi)\, p(\varphi) = p(\varphi|\mathbf{x})\, p(\mathbf{x}). \tag{3.27}$$

In general, the conditional probabilities satisfy $\int\!\mathrm{d}^N x\, p(\mathbf{x}|\varphi) = \int\!\mathrm{d}\varphi\, p(\varphi|\mathbf{x}) = 1$ and the probability of a particular sample corresponds to the marginal $p(\mathbf{x}) = \int\!\mathrm{d}\varphi\, p(\mathbf{x}, \varphi)$. By utilizing the last expression in Eq. (3.27), one can rewrite Eq. (3.26) as

$$\big\langle \Delta^2\tilde{\varphi}_N \big\rangle = \int\!\mathrm{d}^N x\, p(\mathbf{x}) \left[\int\!\mathrm{d}\varphi\, p(\varphi|\mathbf{x})\,(\tilde{\varphi}_N(\mathbf{x}) - \varphi)^2 \right] \tag{3.28}$$

---

[7]Yet, by transforming the parameter we may introduce more pathological parameter values for which $\dot{g}(\varphi) = \pm\infty$ or $\dot{g}(\varphi) = 0$, as then $F_{\mathrm{cl}}\big[p_{g(\varphi)}\big] = F_{\mathrm{cl}}[p_\varphi]/[\dot{g}(\varphi)]^2 = 0$ or $\frac{\partial}{\partial g}\ln p_{g(\varphi)} = \frac{1}{\dot{g}(\varphi)}\frac{\partial}{\partial\varphi}\ln p_\varphi = \pm\infty$ invalidating the CRB.

to show that the $\overline{\text{MSE}}$ is minimal for an estimator which minimises the term in square brackets for each $\mathbf{x}$. Hence, we may determine the form of the optimal *Minimum Mean Squared Error* (MMSE) estimator:[8]

$$\frac{\partial}{\partial \tilde{\varphi}_N(\mathbf{x})} \int \mathrm{d}\varphi \, p(\varphi|\mathbf{x}) \left(\tilde{\varphi}_N(\mathbf{x}) - \varphi\right)^2 = 0 \quad \Longrightarrow \quad \tilde{\varphi}_N^{\text{MMSE}}(\mathbf{x}) = \int \mathrm{d}\varphi \, p(\varphi|\mathbf{x}) \, \varphi = \langle \varphi \rangle_{p(\varphi|\mathbf{x})} , \qquad (3.29)$$

which simply corresponds to the average parameter value computed with respect to the *posterior* PDF, $p(\varphi|\mathbf{x})$, that in principle may always be computed due to the Bayes' theorem (3.27) via

$$p(\varphi|\mathbf{x}) = \frac{p(\mathbf{x}|\varphi) \, p(\varphi)}{\int \mathrm{d}\varphi \, p(\mathbf{x}|\varphi) \, p(\varphi)}. \qquad (3.30)$$

Within the Bayesian framework, one should view the process of data inference as a procedure in which the effective PDF of the estimated parameter φ becomes updated. Hence, the posterior PDF $p(\varphi|\mathbf{x})$ represents the prior $p(\varphi)$ that has been reshaped and narrowed-down after learning the sample $\mathbf{x}$, whereas the MMSE estimator (3.29) just outputs the mean of such an effective distribution. Moreover, the *minimal* $\overline{\text{MSE}}$ (3.26) then reads

$$\left\langle \Delta^2 \tilde{\varphi}_N^{\text{MMSE}} \right\rangle = \int \mathrm{d}^N x \, p(\mathbf{x}) \left[\int \mathrm{d}\varphi \, p(\varphi|\mathbf{x}) \left(\varphi - \langle \varphi \rangle_{p(\varphi|\mathbf{x})}\right)^2 \right] = \int \mathrm{d}^N x \, p(\mathbf{x}) \, \mathrm{Var}[\varphi]|_{p(\varphi|\mathbf{x})} , \qquad (3.31)$$

so that it represents the variance of the parameter φ computed also with respect to $p(\varphi|\mathbf{x})$ and averaged over all the possible outcomes.

Firstly, let us emphasise that, in contrast to the local approach of Sec. 3.1.2, in order to establish the optimal estimator, i.e. the MMSE estimator (3.29), we did *not* have to force it to be *unbiased*. In fact, assuming a particular parameter realisation φ_0, the "local mean" of the MMSE estimator reads

$$\left\langle \tilde{\varphi}_N^{\text{MMSE}} \right\rangle_{\varphi_0} = \int \mathrm{d}^N x \, p(\mathbf{x}|\varphi_0) \, \tilde{\varphi}_N^{\text{MMSE}}(\mathbf{x}) = \iint \mathrm{d}^N x \, \mathrm{d}\varphi \, p(\varphi|\mathbf{x}) \, p(\mathbf{x}|\varphi_0) \, \varphi = \langle \varphi \rangle_{p^N(\varphi|\varphi_0)} , \qquad (3.32)$$

where $p^N(\varphi|\varphi_0) = \int \mathrm{d}^N x \, p(\varphi|\mathbf{x}) \, p(\mathbf{x}|\varphi_0)$ is the probability of inferring the parameter value φ on average given the true value φ_0. Thus, $\tilde{\varphi}_N^{\text{MMSE}}$ is unbiased from the local perspective only if $\langle \varphi \rangle_{p^N(\varphi|\varphi_0)} = \varphi_0$ what is not true in general.

For instance, when the prior PDF $p(\varphi)$ is a distribution much more sensitive to any parameter changes than the distribution $p(\mathbf{x}|\varphi)$ dictating the outcomes collected, the variations of the posterior PDF (3.30) with φ are predominantly determined by the prior PDF with the sampled data playing a marginal role, so that $p(\varphi|\mathbf{x}) \approx p(\mathbf{x}|\varphi)p(\varphi) \approx p(\varphi)$. As a result, also $p^N(\varphi|\varphi_0) \approx p(\varphi)$ and the prior PDF is then responsible for the bias in Eq. (3.32). Moreover, in such a prior-dominant case, the minimal $\overline{\text{MSE}}$ (3.31) approximately equals the variance of the prior PDF and the distribution of the MMSE estimator becomes narrowly peaked around the prior mean. Therefore, it is really important within the Bayesian approach to choose an appropriate $p(\varphi)$ such that, on one hand, it adequately represents the knowledge about the parameter before the estimation, but, on the other, it does not significantly overshadow the information obtained from the data collected.

Note 3.3: Bayesian approach with a Dirac delta prior distribution.
An extremal but instructive example of a prior PDF is the Dirac delta distribution, $p_\delta(\varphi) = \delta(\varphi - \varphi_0)$, which represents the case when we perfectly know the estimated parameter before performing the estimation. Importantly, such situation is *not* equivalent to the local estimation regime of the frequentist approach, in which (as explained in Sec. 3.1.2.3) one seeks for an estimator most sensitive to parameter fluctuations and the prior knowledge, even being

---

[8]In accordance with the convention introduced in Sec. 3.1.2, where the subscript in $\langle \ldots \rangle_\varphi$ denoted the fixed parameter value, we generalise this notation here, so that it explicitly specifies the PDF w.r.t. which the averaging should be performed, e.g. $\langle \ldots \rangle_{p(\phi|\bullet)} = \int \mathrm{d}\phi \, p(\phi|\bullet)(\ldots)$.

complete, is thus irrelevant. In contrast, when we substitute $p_\delta(\varphi)$ into Eq. (3.30) to compute the MMSE estimator (3.29), we simply obtain $\tilde\varphi_N^{\mathrm{MMSE}}(\mathbf{x}) = \varphi_0$ yielding $\langle\Delta^2\tilde\varphi_N^{\mathrm{MMSE}}\rangle = 0$, so that the Bayesian solution to the estimation problem correctly suggests just to output the perfectly known value of φ and fully ignore any data collected.

Secondly, let us stress that so far we did *not* require at any stage the sampled data to be *independently* distributed. Such property, which previously guaranteed locality in the asymptotic N limit within the frequentist approach (see Sec. 3.1.2.3) is *not necessary* in the derivation of the optimal Bayesian estimator, which relies only on the form of the posterior PDF (3.30). In fact, as independently distributed data may be interpreted as if it was collected carrying out consecutive *repetitions* of the estimation protocol, the Bayesian results in such a case may be understood as a *progressive* updating of the knowledge we possess about the parameter. Notice that the data may then be freely split into parts treated as separate samples, e.g. $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2\}$. Due to the independence property $p(\mathbf{x}|\varphi) = p(\mathbf{x}_1|\varphi)p(\mathbf{x}_2|\varphi)$ and one may rewrite the posterior PDF by dividing the numerator and denominator in (3.30) by $p(\mathbf{x}_1)$ as:

$$p(\varphi|\mathbf{x}) = \frac{p(\mathbf{x}_2|\varphi)\,p(\varphi|\mathbf{x}_1)}{\int\mathrm{d}\varphi\,p(\mathbf{x}_2|\varphi)\,p(\varphi|\mathbf{x}_1)}, \tag{3.33}$$

so that $p(\varphi|\mathbf{x})$ may be reinterpreted as if it was calculated basing only on the outcomes $\mathbf{x}_2$ but for the prior already updated with the results $\mathbf{x}_1$. Hence, for independently distributed data, we could equivalently arrive at the MMSE estimator (3.29) evaluated for $p(\varphi|\mathbf{x})$, if we constructed it progressively by repeating the protocol and varying the prior PDF, while including more and more outcomes in each round, what effectively narrows down the spread of each consecutively obtained estimator that eventually coincides with $\tilde\varphi_N^{\mathrm{MMSE}}(\mathbf{x})$ when all the data is finally employed.

On the other hand, the independent character of the sampled data allows to establish a link between the global and local results and, in particular, give an operational meaning to the $\overline{\mathrm{MSE}}$ (3.26), so that it is not just a figure of merit with respect to which the estimator is derived, but is also related to the estimator variance that, as remarked before, is of experimental significance being determinable basing only on the outcomes gathered. For independent data, one may prove that the role of any well-behaved prior distribution becomes negligible in the asymptotic N limit. This is so, because the Local Asymptotic Normality[9] [van der Vaart, 1998] then assures that for *any* regular[10] $p(\varphi)$, the posterior $p(\varphi|\mathbf{x})$ (3.30)—treated as a distribution of φ for a given $\mathbf{x}$ and the parameter true value φ_0—always becomes equivalent to a Gaussian PDF as $N\to\infty$, with mean that may be viewed as a random variable accounting for the fluctuations of $\mathbf{x}$, which is also normally distributed with mean and variance respectively equal to φ_0 and the inverse of the FI: $1/(N\,F_{\mathrm{cl}}[p(X|\varphi)]|_{\varphi_0})$. As the mean of the posterior PDF corresponds to the MMSE estimator (3.29), such observation proves that for regular priors $\tilde\varphi_N^{\mathrm{MMSE}}$ is always asymptotically unbiased and locally saturates the CRB (3.7). Hence, it converges to the ML estimator (3.19) with $\langle\tilde\varphi_N^{\mathrm{MMSE}}\rangle_{\varphi_0} = \langle\varphi\rangle_{p_N(\varphi|\varphi_0)} \overset{N\to\infty}{=} \varphi_0$ and $\Delta^2\tilde\varphi_N^{\mathrm{MMSE}}\big|_{\varphi_0} \overset{N\to\infty}{=} 1/(N\,F_{\mathrm{cl}}[p(X|\varphi)]|_{\varphi_0})$[11]. As a consequence, one may replace for each realisation of φ the MSE, $\Delta^2\tilde\varphi_N^{\mathrm{MMSE}}\big|_{\varphi}$, in Eq. (3.31) by its corresponding CRB, so that the minimal $\overline{\mathrm{MSE}}$ (3.31) may always be rewritten in the asymptotic N limit as

$$\langle\Delta^2\tilde\varphi_N^{\mathrm{MMSE}}\rangle = \left\langle\Delta^2\tilde\varphi_N^{\mathrm{MMSE}}\big|_{\varphi}\right\rangle_{p(\varphi)} \overset{N\to\infty}{=} \left\langle\frac{1}{N\,F_{\mathrm{cl}}[p(X|\varphi)]}\right\rangle_{p(\varphi)} \geq \frac{1}{N\,\langle F_{\mathrm{cl}}[p(X|\varphi)]\rangle_{p(\varphi)}}, \tag{3.34}$$

---

[9]In particular, the Bernstein–von Mises theorem [van der Vaart, 1998].

[10]By regular prior distribution we mean that it possesses "finite information" about the parameter, so that it is smooth and for all φ: $F_{\mathrm{cl}}[p(\varphi)] < \infty$, what excludes e.g. the Dirac delta distribution $p_\delta(\varphi) = \delta(\varphi-\varphi_0)$ discussed in Note 3.3.

[11]In fact, any other reasonable estimator built on the posterior PDF (3.30) will also converge asymptotically to the local results, e.g. estimators corresponding to the median and the mode of $p(\varphi|\mathbf{x})$ mentioned in the following section [van der Vaart, 1998].

where the last expression follows from the Jensen inequality [Jensen, 1906] stating that for any concave function $f(X)$: $\langle f(X) \rangle \geq f(\langle X \rangle)$. Most importantly, although the lower bound (3.34) becomes trivial when the FI (3.8) is independent of the estimated parameter, i.e. $\forall_\varphi$: $F_{\mathrm{cl}}[p(X|\varphi)] = F_{\mathrm{cl}}$, the minimal $\overline{\mathrm{MSE}}$ (3.31) coincides then asymptotically with the CRB, as $\langle \Delta^2 \tilde{\varphi}_N^{\mathrm{MMSE}} \rangle \stackrel{N \to \infty}{=} 1/F_{\mathrm{cl}}$, and hence by Eq. (3.7) constitutes for $N \to \infty$ a lower bound on the variance of any (locally) unbiased estimator.

Lastly, let us remark that Eq. (3.34) can also be derived by means of the so-called *Bayesian CRB* [Gill and Levit, 1995; van Trees, 1968], which applies *regardless* of N and lower-limits the $\overline{\mathrm{MSE}}$ (3.26) for any Bayesian estimator at a price of requiring the prior PDF not only to be regular, but also to vanish at the end-points, i.e. $p(a) = p(b) = 0$ for $\varphi \in [a, b]$:

$$\langle \Delta^2 \tilde{\varphi}_N \rangle \geq \frac{1}{F_{\mathrm{cl}}[p(\varphi)] + N \langle F_{\mathrm{cl}}[p(X|\varphi)] \rangle_{p(\varphi)}} \stackrel{N \to \infty}{=} \frac{1}{N \langle F_{\mathrm{cl}}[p(X|\varphi)] \rangle_{p(\varphi)}}, \tag{3.35}$$

where $F_{\mathrm{cl}}[p(\varphi)]$ is the FI of the prior distribution. As a result, Eq. (3.35) is unfortunately not valid when considering uniform prior PDFs, i.e. $p(\varphi) \simeq 1$ such that $\int \mathrm{d}\varphi\, p(\varphi) = 1$, which we focus on in this work. Thus, in the following sections when considering Bayesian estimation problems, we always utilise Eq. (3.34) to establish connection with the complementary frequentist results for independently distributed samples.

3.1.3.2 Average *cost*

Within the frequentist approach, one is restricted to use the MSE (3.3), $\Delta^2 \tilde{\varphi}_N$, which figure of merit is the *squared distance* between the estimator and the true parameter value: $(\tilde{\varphi} - \varphi)^2$, as only then by imposing the unbiasedness the CRB (3.7) may be derived and utilized. On the other hand, within the Bayesian framework, nothing prevents us to consider other figures of merit, i.e. *cost functions* $C(\tilde{\varphi}, \varphi)$, in order to generalise the $\overline{\mathrm{MSE}}$, $\langle \Delta^2 \tilde{\varphi}_N \rangle$, and define the *average cost*, $\langle C(\tilde{\varphi}_N) \rangle$ [Kay, 1993]:

$$\langle C(\tilde{\varphi}_N) \rangle = \int \mathrm{d}\varphi\, p(\varphi)\, C(\tilde{\varphi}_N)|_\varphi = \int \mathrm{d}\varphi\, p(\varphi) \int \mathrm{d}^N x\, p(\mathbf{x}|\varphi)\, C(\tilde{\varphi}_N(\mathbf{x}), \varphi), \tag{3.36}$$

which in some situations may turn out to be more appropriate than the $\overline{\mathrm{MSE}}$ (3.26) corresponding to the special case of $C(\tilde{\varphi}, \varphi) = (\tilde{\varphi} - \varphi)^2$. For example, other common cost functions that are often considered include the Absolute Error (AE) and the Hit-or-Miss Error (HME) [Kay, 1993]:

$$C_{\mathrm{AE}}(\tilde{\varphi}, \varphi) = |\tilde{\varphi} - \varphi|, \qquad C_{\mathrm{HME}}(\tilde{\varphi}, \varphi) = \begin{cases} 0, & |\tilde{\varphi} - \varphi| \leq \delta \\ 1, & |\tilde{\varphi} - \varphi| > \delta \end{cases} \text{ with } \delta \ll 1. \tag{3.37}$$

The average cost in the case of AE, $\langle C_{\mathrm{AE}}(\tilde{\varphi}_N) \rangle$, does not differ significantly from the $\langle \Delta^2 \tilde{\varphi}_N \rangle$, but due to the cost function being linear and not quadratic it does not penalise that much for estimates being far from the true parameter value. As a consequence, the optimal Bayesian estimator does not correspond then to the mean of the posterior PDF, $p(\varphi|\mathbf{x})$, as in the case of the $\overline{\mathrm{MSE}}$ and the MMSE estimator (3.29), but rather to its *median*. The HME, on the other hand, is most restrictive rewarding only the estimates being approximately the real value of the parameter, so that the optimal estimator minimising the average cost, $\langle C_{\mathrm{HME}}(\tilde{\varphi}_N) \rangle$, corresponds to choosing the most probable value of φ with respect to the posterior PDF, i.e. its *maximum* (mode)[12]. In particular, such an estimator is equivalent *regardless of* N to the ML estimator (3.19) considered within the frequentist approach when the prior PDF is assumed to be uniform, i.e. $p(\varphi) \simeq const$, [van der Vaart, 1998].

---

[12]For explicit derivations of the optimal estimators for the AE and HME cost functions see e.g. [Kay, 1993].

However, we present the cost functions (3.37) to the reader only as instructive examples, as we do not utilise them explicitly within this work. This is because, we restrict ourselves to average costs (3.36) which converge in the asymptotic N limit to the $\overline{\mathrm{MSE}}$ (3.26), i.e. $\langle \mathcal{C}(\tilde{\varphi}_N) \rangle \overset{N \to \infty}{=} \langle \Delta^2 \tilde{\varphi}_N \rangle$. As any consistent estimator (see Eq. (3.2)) attains the parameter true value with $N \to \infty$, so that $\tilde{\varphi}$ approaches φ with N, such constraint is fulfilled by considering only cost functions that satisfy $C(\tilde{\varphi}, \varphi) = (\tilde{\varphi} - \varphi)^2 + O[(\tilde{\varphi} - \varphi)^3]$. As a result, the *optimal* Bayesian estimator, $\tilde{\varphi}_N^{\mathrm{opt}}$, *always* converges asymptotically to the MMSE estimator, $\tilde{\varphi}_N^{\mathrm{MMSE}}$, and the average cost $\langle \mathcal{C}(\tilde{\varphi}_N^{\mathrm{opt}}) \rangle$ attains with $N \to \infty$ the *minimal* $\overline{\mathrm{MSE}}$ (3.31). Furthermore, if the data is *independently* distributed, $\tilde{\varphi}_N^{\mathrm{opt}}$ also approaches the ML estimator (3.19) and $\langle \mathcal{C}(\tilde{\varphi}_N^{\mathrm{opt}}) \rangle$ may be related to the frequentist results via Eqs. (3.34) and (3.35). In principle, $\langle \mathcal{C}(\tilde{\varphi}_N^{\mathrm{opt}}) \rangle$ may then be even utilised to derive the local precision bounds and the form of the CRB (3.7).

3.1.3.3 Average *cost* with circular symmetry

We have introduced the concept of general average cost $\langle \mathcal{C}(\tilde{\varphi}_N) \rangle$ (3.36), in order to be able to *globally* solve *phase estimation problems*, in which φ is a circular parameter or more formally an element of the circle group[13] $U(1)$ satisfying $\varphi \equiv \varphi + 2\pi n$ for any $n \in \mathbb{Z}$. In such a case, the squared-distance cost function employed in the $\overline{\mathrm{MSE}}$ (3.26) is not valid, as it does not respect the parameter *topology* which must be taken into account when performing the integral over all parameter values contained in the prior PDF in Eq. (3.36) [Holevo, 1982]. This contrasts the frequentist case of Sec. 3.1.2, where the strategy was optimised to sense only small variations of φ and such issues were completely ignored. Such ignorance, however, can lead within the local approach to estimators that give inconclusive answers, as they disregard any parameter symmetry (see e.g. the later discussed estimator of Eq. (3.46)). On the other hand, when pursuing the Bayesian approach, we may *correctly* account for the parameter topology, what in the case of the circular parameter is achieved by requiring the cost function to be: *symmetric* – $C(\tilde{\varphi}, \varphi) = C(\varphi, \tilde{\varphi})$, *group invariant* – $\forall_{\phi \in U(1)} : C(\tilde{\varphi} + \phi, \varphi + \phi) = C(\tilde{\varphi}, \varphi)$ and *periodic* – $\forall_{n \in \mathbb{Z}} : C(\tilde{\varphi} + 2\pi n, \varphi) = C(\tilde{\varphi}, \varphi)$, so that most generally we may rewrite it as a function of the difference between the estimator value and the parameter realisation, $\theta = \tilde{\varphi} - \varphi$, which reads:

$$C(\tilde{\varphi}, \varphi) = C(\theta) = \sum_{k=0}^{\infty} c_k \cos(k\theta). \tag{3.38}$$

Furthermore, as $C(\theta)$ must rise monotonically from $C(0) = 0$ at $\theta = 0$ to some $C(\pi) = C_{\max}$ at $\theta = \pi$, so that $C'(\theta) \geq 0$, the coefficients c_k must fulfil following constraints:

$$\sum_{k=0}^{\infty} c_k = 0, \qquad \sum_{k=0}^{\infty} (-1)^k c_k = C_{\max}, \qquad \sum_{k=1}^{\infty} k^2 c_k \leq 0, \qquad \sum_{k=1}^{\infty} k^2 (-1)^k c_k \geq 0, \tag{3.39}$$

which may be satisfied by imposing $\forall_{k>0} : c_k \leq 0$ and taking c_k to decay at least quadratically with k. Lastly, as mentioned in the previous section, we require the average cost $\langle \mathcal{C}(\tilde{\varphi}_N) \rangle$ to converge to the $\overline{\mathrm{MSE}}$ $\langle \Delta^2 \tilde{\varphi}_N \rangle$ in the asymptotic N limit, in which the estimator approaches the true parameter and $\theta \to 0$. Such constraint imposes that for small θ: $C(\theta) = \theta^2 + O(\theta^4)$ and hence $\sum_{k=1}^{\infty} k^2 c_k = -2$.

In all the Bayesian estimation problems considered in our work that deal with a circularly symmetric parameter, we will consider the simplest cost function introduced by Holevo [1982], which satisfies all the above-mentioned conditions with $c_0 = -c_1 = 2$, $\forall_{k>1} : c_k = 0$ and explicitly reads:

$$C_{\mathrm{H}}(\tilde{\varphi}, \varphi) = C_{\mathrm{H}}(\tilde{\varphi} - \varphi) = 4 \sin^2\left(\frac{\tilde{\varphi} - \varphi}{2}\right). \tag{3.40}$$

---

[13]Typically denoted by $\mathbb{T}$ or $U(1)$—being also the group of unitary 1×1 matrices.
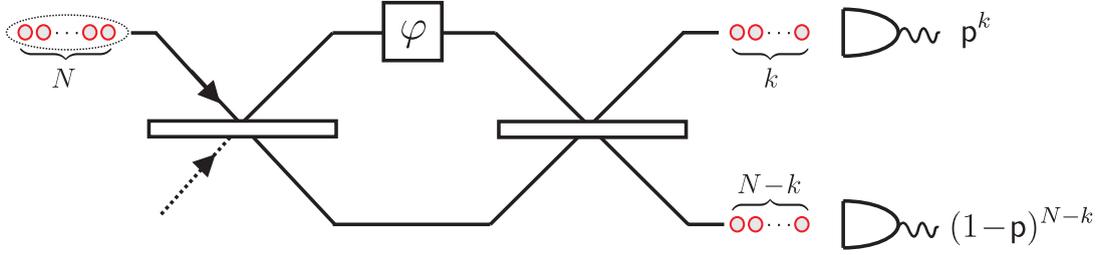
FIGURE 3.1: **Schematic description of a Mach-Zehnder interferometer** which introduces a relative *phase* delay φ in between the light beams travelling in its arms. A light state of N *uncorrelated* photons, i.e. a Fock state $|N\rangle$, is impinged onto one of the input ports. As a result, each of the constituent photons may be treated independently and is detected with probabilities $\mathsf{p} = \cos^2 \frac{\varphi}{2}$ and $1-\mathsf{p} = \sin^2 \frac{\varphi}{2}$ at the output ports. Hence, the overall distribution of registering k photons in one of the ports (and thus $N-k$ in the other) is binomal.

Consequently, it leads to the average cost (3.36) of the form:

$$\langle \mathcal{C}_{\mathrm{H}}(\tilde{\varphi}_N) \rangle = 4 \int d\varphi \, p(\varphi) \int d\mathbf{x} \, p(\mathbf{x}|\varphi) \, \sin^2\left(\frac{\tilde{\varphi}_N(\mathbf{x}) - \varphi}{2} \right). \tag{3.41}$$

Following the same argumentation as in Eq. (3.29) describing the derivation of the MMSE estimator optimal for the $\overline{\mathrm{MSE}}$ (3.26), one may prove that $\langle \mathcal{C}_{\mathrm{H}}(\tilde{\varphi}_N) \rangle$ is minimised if an estimator, $\tilde{\varphi}_N^{\mathrm{H}}$, can be found that for any possible data sample $\mathbf{x}$ collected satisfies the following condition:

$$\int d\varphi \, p(\varphi|\mathbf{x}) \, \sin(\tilde{\varphi}_N^{\mathrm{H}}(\mathbf{x}) - \varphi) = 0. \tag{3.42}$$

3.1.4 EXAMPLE: **Mach-Zehnder interferometry with uncorrelated photons**

Let us consider in detail a physically motivated example of a *Mach-Zehnder interferometer* depicted in Fig. 3.1, for which the estimated φ represents the relative *phase* delay in between the interferometer arms [Demkowicz-Dobrzański *et al.*, 2015], so that the parameter naturally exhibits a circular symmetry, and thus we choose $\varphi \in [-\pi, \pi]$. When a state of light consisting of N uncorrelated photons[14] is shone on the input port, each individual photon may be detected in one of the two output ports with probabilities $\mathsf{p} = \cos^2 \frac{\varphi}{2}$ and $1-\mathsf{p} = \sin^2 \frac{\varphi}{2}$ respectively, what resembles a coin-tossing experiment with binary probabilities p and $1-\mathsf{p}$ dictating the 'heads'/'tails' outcomes. Hence, the overall probability distribution of registering k photons in one port and $N-k$ in the other is binomial and reads

$$p_\varphi^N(k) = \binom{N}{k} \mathsf{p}^k \, (1-\mathsf{p})^{N-k} = \binom{N}{k} \left(\cos^2 \frac{\varphi}{2} \right)^k \left(\sin^2 \frac{\varphi}{2} \right)^{N-k}, \tag{3.43}$$

naturally belonging to exponential family of PDFs (3.14) with $h(k) = \binom{N}{k}$, $T(k) = k$, $\eta(\varphi) = \ln\left(\cot^2 \frac{\varphi}{2} \right)$ and $A(\varphi) = \ln\left(\operatorname{cosec}^2 \frac{\varphi}{2} \right)$. We study the problem of φ-estimation in such a scenario from both frequentist and Bayesian perspectives below.

---

[14]In particular, a Fock state $|N\rangle$, which may be also simulated by impinging a coherent state of light and post-selecting only events when the total number of photons registered at the output ports turns out to be N [Demkowicz-Dobrzański *et al.*, 2015]. Importantly, such a state does not possess any correlations (even classical) in between the constituent photons, as explicitly shown in Note 2.3.

3.1.4.1   Frequentist approach

Calculating the FI (3.8) with the integral $\int d\mathbf{x}$ replaced accordingly by the sum $\sum_{k=0}^{N}$ due to discreteness of the distribution (3.43) or equivalently utilizing Eq. (3.15) for the exponential family of PDFs, we derive the CRB (3.7) via

$$F_{\mathrm{cl}}\big[p_\varphi^N\big] = N\, F_{\mathrm{cl}}[p_\varphi] = N \quad \Longrightarrow \quad \Delta^2\tilde{\varphi}_N \geq \frac{1}{N}\,. \tag{3.44}$$

Although the above CRB is independent of the actual parameter value, investigating the CRB-saturability condition (3.11), we realise that the binomial PDF (3.43) with such a parametrisation does not satisfy the corresponding requirement (3.16) for a global efficient estimator to exist, because

$$\frac{\dot{A}(\varphi)}{\dot{\eta}(\varphi)} = \cos^2\frac{\varphi}{2} \neq \varphi. \tag{3.45}$$

However, for a given parameter true value φ_0, we may always construct a local efficient estimator following the prescription of Eq. (3.18):

$$\tilde{\varphi}_{N,\varphi_0}(k) = \varphi_0 + \cot\left(\frac{\varphi_0}{2}\right) - \frac{2k}{\sin(\varphi_0)\,N}, \tag{3.46}$$

which is correctly unbiased: $\langle \tilde{\varphi}_{N,\varphi_0}\rangle_{\varphi_0} = \varphi_0$, and its MSE indeed saturates the CRB (3.44): $\Delta^2\tilde{\varphi}_{N,\varphi_0}\big|_{\varphi_0} = 1/N$. Yet, for this to be generally true, one must let in the calculation $\tilde{\varphi}_{N,\varphi_0}\in[-\infty,\infty]$ and completely ignore the fact that $\tilde{\varphi}_{N,\varphi_0}$ is utilised to determine a circularly symmetric parameter. Hence, the applicability of the estimator (3.46) becomes doubtful for the true values φ_0, for which $\tilde{\varphi}_{N,\varphi_0}$ exits out of the range of a single period, i.e. $[-\pi,\pi]$, as it does not associate then equivalent parameter values with one another (does not respect the fact that e.g. $\varphi = 3\pi/2 \equiv -\pi/2$). On the other hand, for the parameter values $\varphi_0 = \{0,\pm\pi\}$, the estimator (3.46) diverges even if one allows it to be defining points on the whole real line, but so actually does $\big(\mathrm{d}\ln p_\varphi^N(k)/\mathrm{d}\varphi\big)\big|_{\varphi_0}$, so that the regularity assumption (3.6) fails and the CRB is not valid there at all. As a matter of fact, these pathological points correspond to the cases of $\mathsf{p}=\{0,1\}$, for which the estimation problem becomes deterministic, as by learning then whether respectively $k = \{0,N\}$ we may deduce the parameter value without any error.

Secondly, we consider the ML estimator (3.19) that is guaranteed to saturate the CRB (3.44) independently of φ in the asymptotic N limit. As we have just shown the estimation problem not to allow for a global efficient estimator to exist, the conditions (3.11) and (3.20) are not equivalent and thus $\tilde{\varphi}_N^{\mathrm{ML}}$ must be found by explicitly solving Eq. (3.20) for the PDF (3.43):

$$\tilde{\varphi}_N^{\mathrm{ML}}(k) = \underset{\varphi}{\mathrm{argmax}}\, \ln p_\varphi^N(k) = \pm 2\,\mathrm{arccot}\left(\sqrt{\frac{k}{N-k}}\right). \tag{3.47}$$

The two equivalent maxima arise due to the ambiguity in the sign of φ, as $\forall_{N,k}\colon p_\varphi^N(k) = p_{-\varphi}^N(k)$. As a consequence, before applying $\tilde{\varphi}_N^{\mathrm{ML}}$ as a global estimator, we must possess extra information that allows us to deduce whether the parameter true value φ_0 is positive. The necessary division of the whole range $[-\pi,\pi]$ into halves is dictated by the parametrisation $\mathsf{p} = \cos^2\frac{\varphi}{2}$, as φ may be unambiguously inferred from p only after restricting to parameter sub-ranges in which p is monotonic in φ, i.e. $[-\pi,0]$ and $[0,\pi]$. Thus, the ML estimator (3.47) is non-smooth at $\varphi_0 = \{0,\pm\pi\}$, which consistently are the pathological points at which the regularity condition (3.6)—and hence the CRB (3.44)—fails[15]. Taking $\varphi_0 \geq 0$ we

---

[15]Notice that, in contrast to the local efficient estimator (3.46), the ML estimator (3.47) does not exit the parameter range for which it is defined, i.e. $[-\pi,0]$ or $[0,\pi]$, so that it does *not* violate the circular symmetry of the parameter.
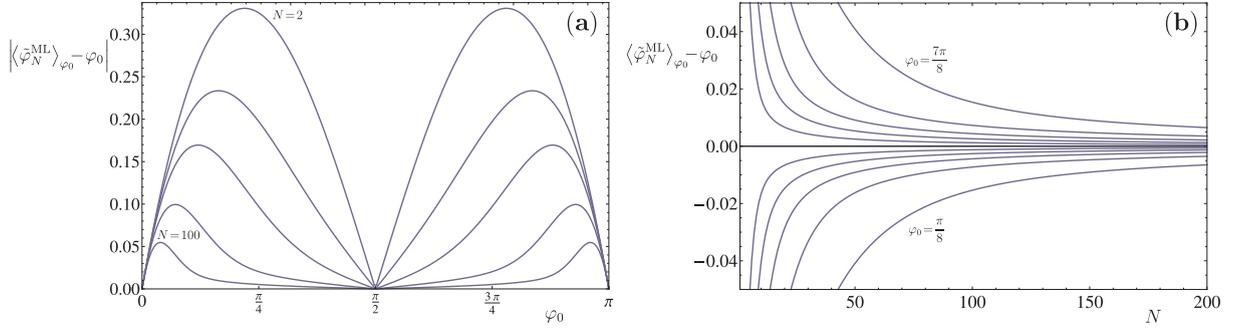
FIGURE 3.2: **Bias of the ML estimator** (3.47) as a function of: (**a**) – the parameter true value φ_0, and (**b**) – the sample size N. The curves in (**a**) are depicted for $N = \{2, 5, 10, 30, 100\}$, whereas the ones in (**b**) correspond to the equally-distributed values: $\varphi_0 = \{\frac{\pi}{8}, \ldots, \frac{\pi}{2}, \ldots, \frac{7\pi}{8}\}$. Both plots clearly indicate a decrease of the bias with an increase in N and the special parameter values $\varphi_0 = \{0, \frac{\pi}{2}, \pi\}$, for which the bias vanishes irrespectively of the sample size. Nevertheless, the ML estimator is *locally unbiased* according to Eq. (3.5) *only* in the asymptotic N limit.

depict the bias of the ML estimator in Fig. 3.2 as a function of consecutively φ_0 and N. As required by the asymptotic unbiasedness property, the bias diminishes gradually with N for any φ_0. Yet, it vanishes regardless of N at the special value $\varphi_0 = \frac{\pi}{2}$ and the pathological $\varphi_0 = \{0, \pm\pi\}$ for all of which thus $\text{Var}[\tilde{\varphi}_N^{\text{ML}}] = \Delta^2 \tilde{\varphi}_N^{\text{ML}}$. However, one should bear in mind that vanishing of the estimator bias is not sufficient for the fulfilment of the local unbiasedness conditions (3.5), which importantly also constrain the "speed" of change of the estimator mean. Hence, the ML estimator, being locally unbiased only in the asymptotic N limit, can in principle surpass the CRB (3.44) for any finite N and φ_0. Yet, this is not the case for the special parameter value $\varphi_0 = \frac{\pi}{2}$, which intuitively should be the optimal one as it corresponds to the point of the steepest variation of p with φ, i.e. $\text{argmax}_\varphi |\mathrm{d}\mathsf{p}/\mathrm{d}\varphi|$, leading to the highest parameter sensitivity. The MSE plots[16] in Fig. 3.3 confirm these facts by indicating that for $\varphi_0 = \frac{\pi}{2}$, represented by the saddle points in Fig. 3.3(**a**), $\Delta^2 \tilde{\varphi}_N^{\text{ML}}$ never surpasses and most rapidly attains the CRB—up to 5% already at $N \approx 50$, as shown in Fig. 3.3(**b**). On the other hand, for the pathological values $\varphi_0 = \{0, \pm\pi\}$ at which CRB is not valid even asymptotically, $\tilde{\varphi}_N^{\text{ML}}$ is errorless with $\Delta^2 \tilde{\varphi}_N^{\text{ML}} = 0$ for any N. Furthermore, for low N and the parameter values close to these pathological points, for which the ML estimator significantly violates the local unbiasedness constraints (3.5), $\Delta^2 \tilde{\varphi}_N^{\text{ML}}$ indeed surpasses the CRB. As a result, the curve in Fig. 3.3(**b**) representing relative percentage excess over the CRB as a function of N for $\varphi_0 = \frac{\pi}{8}$ starts negative below the horizontal line representing the CRB. Nevertheless, the sub-CRB regions lying below the CRB-threshold in Fig. 3.3(**a**) narrow down as the bias evanesces with $N \to \infty$ and, as required, $\Delta^2 \tilde{\varphi}_N^{\text{ML}}$ attains asymptotically the CRB from above for any $\varphi_0 \neq \{0, \pm\pi\}$.

Lastly, we seek for the optimal form of a transformed parameter, i.e. $g(\varphi)$ in Eq. (3.23), such that a global efficient estimator exists. Rewriting the necessary condition (3.25) applicable to exponential PDFs (3.14), we deduce that

$$\frac{\dot{A}(\varphi)}{\dot{\eta}(\varphi)} = \cos^2\frac{\varphi}{2} = \mathsf{p}, \tag{3.48}$$

so that the necessary parameter to be efficiently estimated corresponds simply to the probability of each binary outcome: p. The corresponding transformed CRB (3.23) limiting the MSE and variance of any unbiased estimator of p then reads

$$\Delta^2 \tilde{\mathsf{p}}_N \geq \frac{\mathsf{p}(1 - \mathsf{p})}{N} \tag{3.49}$$

---

[16]We have chosen to plot the MSE, $\Delta^2 \tilde{\varphi}_N$, that adequately quantifies the local performance of any estimator, but one should be aware that the estimator variance $\text{Var}[\tilde{\varphi}_N]$ (for which Fig. 3.3 in the case of $\tilde{\varphi}_N^{\text{ML}}$ changes insignificantly) is the more experimentally relevant quantity, being determinable basing only on a large data sample collected for an unknown φ.
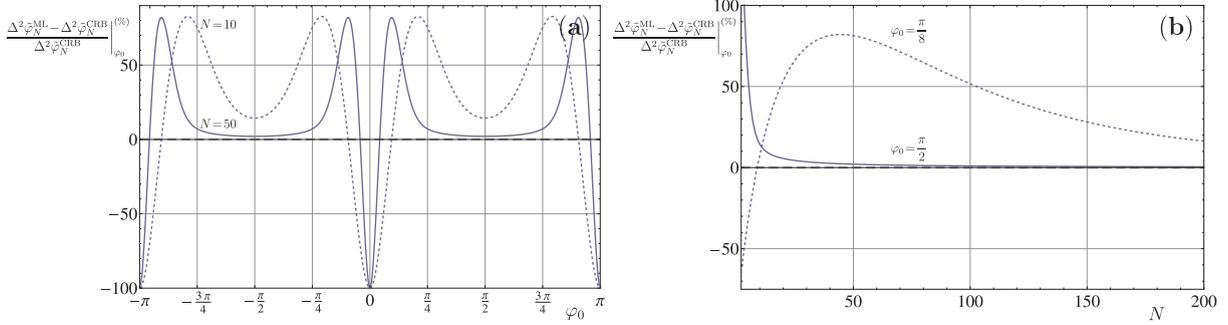
FIGURE 3.3: **MSE of the ML estimator** (3.47) in comparison with the CRB (3.44) as a function of: (**a**) – the parameter true value φ_0, and (**b**) – the sample size N. In both plots the CRB is represented by the horizontal line (*grey dashed*). The curves in (**a**) are depicted for $N = \{10, 50\}$ and show: the pathological points $\varphi_0 = \{0, \pm\pi\}$ at which $\forall_N: \Delta^2\tilde{\varphi}_N^{\mathrm{ML}} = 0$, and the optimal points $\varphi_0 = \pm\frac{\pi}{2}$ for which the CRB is most rapidly attained. The MSE dependence on N, plotted in (**b**), indicates that for parameter values close to the pathological points, e.g. $\varphi_0 = \frac{\pi}{8}$, $\Delta^2\tilde{\varphi}_N^{\mathrm{ML}}$ surpasses the CRB at low N to still attain it asymptotically from above, whereas at the optimal $\varphi_0 = \pm\frac{\pi}{2}$ the CRB is attained within negligible margin of error already for $N \approx 50$.

and is assured to be saturated for any p by the global estimator $\tilde{\mathsf{p}}_N(k) = k/N$, which thus coincides with the ML estimator (3.19) that is now efficient *regardless of* N and not only in the asymptotic N limit. In contrast to the CRB (3.44) on $\Delta^2\tilde{\varphi}_N$, the PDF (3.43) fulfils the regularity condition (3.6) with respect to p for the previously pathological values $\mathsf{p} = \{0, 1\}$, so that the transformed CRB (3.49) still applies to $\Delta^2\tilde{\mathsf{p}}_N$ at these extremal points, where it actually vanishes—correctly indicating the possibility of perfect parameter determination. However, one should note that although the transformed CRB (3.49) corresponds now to a tight inequality, it is parameter dependent. Hence, as discussed in Sec. 3.1.2.3, in any real-life estimation protocol in order to actually compute the CRB, one requires some prior knowledge of the estimated binary probability p that is assured again only in the local estimation regime.

3.1.4.2 Bayesian approach

When solving the above estimation problem within the Bayesian approach, we treat φ as a random variable and thus alter the notation of Eq. (3.43), so that $p^N(k|\varphi) \equiv p_\varphi^N(k)$ represents now the conditional PDF. As the estimated parameter describes the phase delay of the Mach-Zehnder interferometer depicted in Fig. 3.1, we must account for its circular symmetry and therefore minimise the average cost (3.41), $\langle \mathcal{C}_\mathrm{H}(\tilde{\varphi}_N) \rangle$, as the figure of merit rather than the $\overline{\mathrm{MSE}}$ (3.26), $\langle \Delta^2\tilde{\varphi}_N \rangle$. Furthermore, as we ideally do *not assume any prior knowledge* about the estimated phase, we firstly take the prior PDF to correspond to a uniform distribution over the full parameter period, i.e. $p(\varphi) = 1/(2\pi)$ for $\varphi \in [-\pi, \pi]$. However, in such a case, when we compute the posterior PDF $p^N(\varphi|k)$ with help of Eq. (3.30) after substituting for the binomial PDF (3.43), we find the necessary condition (3.42) for the optimal estimator to be trivially satisfied by $\tilde{\varphi}_N^\mathrm{H}(k) = 0$ which completely disregards the data collected: k. Because of choosing the whole parameter range $\varphi \in [-\pi, \pi]$, we are again not able to resolve the sign ambiguity $\pm\varphi$, which previously lead to two distinct solutions of the ML estimator (3.47) within the local approach. Within the Bayesian framework the consequences are even more serious, as it is more beneficial to always output $\varphi = 0$, which is the only non-ambiguous parameter value, rather than infer from the data any other realisation of φ that, due to its equally probable counterpart with opposite sign, introduces on average an overwhelming error. Hence, similarly to the frequentist solution, for the estimation problem not to be "ill-defined", we must possess extra information that allows us to restrict the parameter search to a sub-range in which φ is unambiguously determined by the outcomes, i.e. in which $\mathsf{p} = \cos^2\frac{\varphi}{2}$ is monotonic in φ.
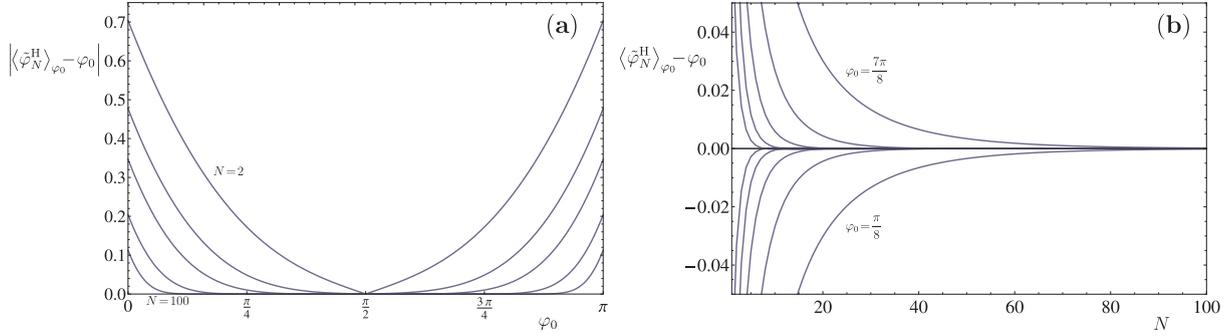
FIGURE 3.4: **Bias of the Bayesian estimator** (3.50) as a function of: (**a**) – the parameter true value φ_0, and (**b**) – the sample size N. The curves in (**a**) are depicted for $N = \{2, 5, 10, 30, 100\}$, whereas the ones in (**b**) correspond to the equally-distributed values: $\varphi_0 = \{\frac{\pi}{8}, \dots, \frac{\pi}{2}, \dots, \frac{7\pi}{8}\}$. Both plots clearly indicate a rapid decrease of the bias with an increase in N. At the special parameter value $\varphi_0 = \frac{\pi}{2}$ the bias vanishes irrespectively of the sample size, what however does *not* guarantee the Bayesian estimator to be *locally unbiased* unless the asymptotic N limit is considered.

That is why, we choose $\varphi \in [0, \pi]$ and the adequate prior distribution $p(\varphi) = 1/\pi$, so that Eq. (3.42) now yields a non-trivial form of the optimal Bayesian estimator:

$$\tilde{\varphi}_N^{\text{H}}(k) = \operatorname{arccot}\left[\frac{\left(k - \frac{N}{2}\right)\Gamma\left(k + \frac{1}{2}\right)\Gamma\left(N - k + \frac{1}{2}\right)}{k!(N-k)!}\right],\tag{3.50}$$

where $\Gamma(x)$ stands for the Euler Gamma function. Consistently with $\langle\mathcal{C}_{\text{H}}(\tilde{\varphi}_N)\rangle \overset{N\to\infty}{=} \langle\Delta^2\tilde{\varphi}_N\rangle$ and discussions of Sec. 3.1.3.2, the estimator (3.50) converges in the asymptotic N limit to the ML estimator (3.47), i.e. $\forall_k: \tilde{\varphi}_N^{\text{H}}(k) \overset{N\to\infty}{=} \tilde{\varphi}_N^{\text{ML}}(k)$. However, one should note that for any finite N due to averaging over a uniform prior distribution the optimal Bayesian estimator—in contrast to the ML estimator (3.47)— does not output with certainty the previously pathological values $\varphi = \{0, \pm\pi\}$ for the extremal outcomes $k = \{0, N\}$, e.g. $\tilde{\varphi}_N^{\text{H}}(N) = \operatorname{arccot}\{\sqrt{\pi}\,\Gamma(N + 1/2)/[2(N-1)!]\} \overset{N\to\infty}{=} 0$ in comparison to $\tilde{\varphi}_N^{\text{ML}}(N) = 0$ for any N. Such behaviour is clearly depicted in Fig. 3.4(**a**) illustrating the bias of the Bayesian estimator to be non-zero at $\varphi = \{0, \pm\pi\}$, which leads to a significant MSE shown in Fig. 3.5(**a**) at these values. Note that Figs. 3.4 and 3.5 illustrate the local performance of $\tilde{\varphi}_N^{\text{H}}$ for a given realisation φ_0, so that the MSE (3.3) is indeed the adequate figure of merit. As a result, Figs. 3.4 and 3.5 are directly comparable to Figs. 3.2 and 3.3 describing the performance of $\tilde{\varphi}_N^{\text{ML}}$. Surprisingly, the Bayesian estimator, derived with the global approach in mind, turns out to be more effective than the ML estimator even from the local perspective. In particular, its bias diminishes much more rapidly with the sample size N and its MSE attains the CRB (3.44) much faster. This can be seen when comparing the sub-plots (**b**) of Figs. 3.4 and 3.5 with respectively Figs. 3.2 and 3.3 after noticing a significant change in the corresponding plot scales. Again, as shown in Fig. 3.5, the optimal parameter value for which the Bayesian estimator $\tilde{\varphi}_N^{\text{H}}$ most quickly saturates the CRB is $\varphi_0 = \frac{\pi}{2}$. Yet—in contrast to $\tilde{\varphi}_N^{\text{ML}}$—$\tilde{\varphi}_N^{\text{H}}$ at this special point surpasses the CRB at low N, so that the corresponding curve in Fig. 3.5(**b**) depicting the relative percentage excess over the CRB starts below the horizontal line representing the CRB. This is however consistent, as the Bayesian estimator, similarly to the ML estimator, only asymptotically fulfils the local unbiasedness conditions (3.5), so that in principle it may surpass the CRB (3.44) for any finite N and φ_0.

Nevertheless, one should not forget that, although Figs. 3.4 and 3.5 describe the local performance of the Bayesian estimator (3.50) for a known true value φ_0, $\tilde{\varphi}_N^{\text{H}}$ does not neglect the circular symmetry of the problem being derived basing on the average cost $\langle\mathcal{C}_{\text{H}}(\tilde{\varphi}_N)\rangle$. In fact, substituting the optimal Bayesian
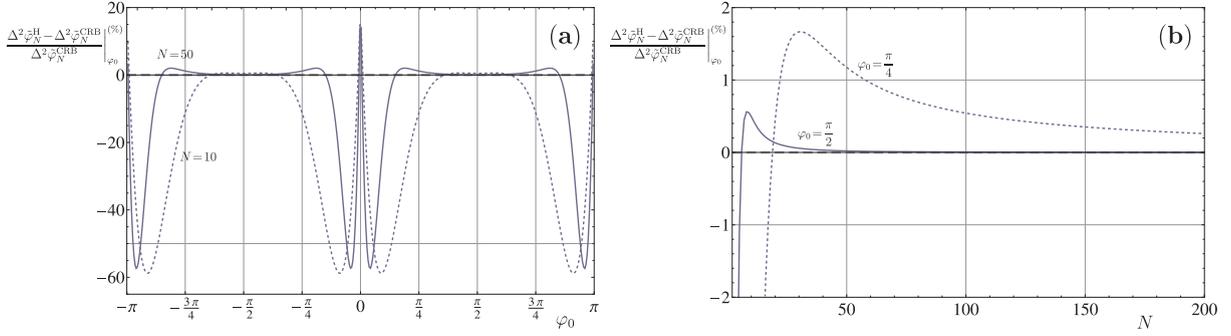
FIGURE 3.5: **MSE of the Bayesian estimator** (3.50) in comparison with the CRB (3.44) as a function of: (**a**) – the parameter true value φ_0, and (**b**) – the sample size N. In both plots the CRB is represented by the horizontal line (*grey dashed*). The curves in (**a**) are depicted for $N = \{10, 50\}$ and show the optimal points $\varphi_0 = \pm\frac{\pi}{2}$ for which the CRB is most rapidly attained. The MSE dependence on N, plotted in (**b**), indicates that $\Delta^2\tilde{\varphi}_N^{\mathrm{ML}}$ surpasses the CRB at low N to still attain it asymptotically from above. For the optimal $\varphi_0 = \pm\frac{\pi}{2}$ the CRB is attained to negligible precision already for $N \approx 50$.

estimator $\tilde{\varphi}_N^{\mathrm{H}}$ into Eq. (3.41), we obtain the expression for the overall average cost:

$$\langle \mathcal{C}_{\mathrm{H}}(\tilde{\varphi}_N^{\mathrm{H}}) \rangle = 2 \left[1 - \frac{1}{\pi(N+1)} \sum_{k=0}^{N} \sqrt{4 + \left(\frac{(N-2k)\Gamma(k+\frac{1}{2})\Gamma(N-k+\frac{1}{2})}{k!(N-k)!} \right)^2} \right], \tag{3.51}$$

which may then be proved to coincide in the asymptotic N limit with the corresponding CRB (3.44), as $\langle \mathcal{C}_{\mathrm{H}}(\tilde{\varphi}_N^{\mathrm{H}}) \rangle \overset{N\to\infty}{=} 1/N$. Revisiting the discussions of Sec. 3.1.3.1, as the average cost (3.51) is asymptotically equivalent to the $\overline{\mathrm{MSE}}$ (3.26), Eq. (3.34) relating local and global results applies. Furthermore, as the corresponding CRB (3.44), $1/N$, is parameter independent, Eq. (3.34) takes its simple form, so that for any regular prior distribution the average cost must indeed converge to the CRB as $N \to \infty$.

For completeness, let us also consider the problem of p-estimation, which we have shown to allow for a global efficient estimator within the frequentist approach. As the binary outcome probability p does not constitute a circular parameter, we may directly minimise the $\overline{\mathrm{MSE}}$ (3.26), $\langle \Delta^2 \tilde{\mathsf{p}}_N^{\mathrm{MMSE}} \rangle$, for which the MMSE estimator (3.29) is always optimal. Thus, assuming a uniform prior distribution $p(\mathsf{p}) = 1$ for $\mathsf{p} \in [0, 1]$, we calculate the posterior PDF defined with help of Eq. (3.30) after substituting for the binomial PDF (3.43), $p_{\mathsf{p}}^N(k) \equiv p^N(k|\mathsf{p})$, to obtain $p^N(\mathsf{p}|k) = (N+1)\, p^N(k|\mathsf{p})$. As a result, we may compute the MMSE estimator defined as the mean of $p^N(\mathsf{p}|k)$:

$$\tilde{\mathsf{p}}_N^{\mathrm{MMSE}}(k) = \int_0^1 \mathrm{d}\mathsf{p}\, \mathsf{p}\, p^N(\mathsf{p}|k) = \frac{k+1}{N+2}, \tag{3.52}$$

and the corresponding minimal $\overline{\mathrm{MSE}}$ (3.31):

$$\left\langle \Delta^2 \tilde{\mathsf{p}}_N^{\mathrm{MMSE}} \right\rangle = \int_0^1 \mathrm{d}\mathsf{p}\, p(\mathsf{p}) \sum_{k=0}^{N} p^N(k|\mathsf{p}) \left(\frac{k+1}{N+2} - \mathsf{p} \right)^2 = \frac{1}{6(N+2)}. \tag{3.53}$$

Again, in accordance with discussions of Sec. 3.1.3.1, the (optimal Bayesian) MMSE estimator (3.52) converges with N to the global unbiased (and thus the ML) estimator $\tilde{\mathsf{p}}_N^{\mathrm{ML}}(k) = k/N$ derived within the frequentist approach. On the other hand, as the corresponding CRB (3.49) is p-dependent, Eq. (3.34) relating local and global precision measures assures the minimal $\overline{\mathrm{MSE}}$ (3.53) to converge to the average

version of the CRB (3.49), so that

$$\left\langle \Delta^2 \tilde{\mathsf{p}}_N^{\text{MMSE}} \right\rangle \stackrel{N \to \infty}{=} \left\langle \Delta^2 \tilde{\mathsf{p}}_N^{\text{ML}} \big|_{\mathsf{p}} \right\rangle_{p(\mathsf{p})} = \int_0^1 d\mathsf{p} \, \frac{\mathsf{p}(\mathsf{p}-1)}{N} = \frac{1}{6N}, \tag{3.54}$$

what is consistent with Eq. (3.53) also yielding $\left\langle \Delta^2 \tilde{\mathsf{p}}_N^{\text{MMSE}} \right\rangle \stackrel{N \to \infty}{=} 1/(6N)$.

Lastly, let us emphasise once more that due to $p_{\varphi/\mathsf{p}}(\mathbf{x}) = \prod_{i=1}^N p_{\varphi/\mathsf{p}}(x_i)$ the asymptotic N limit corresponds to the regime of the infinitely sized, *independently* distributed data. From the frequentist perspective, this means that the locality of estimation is always guaranteed when $N \to \infty$, whereas within the Bayesian framework one is thus able to invoke the asymptotic connection (3.34) with the local results and relate the corresponding estimators and precision measures. In the case of the Mach-Zehnder interferometer of Fig. 3.1, such condition is fulfilled owing to the assumption that the photons employed are uncorrelated between one another and individually measured at the end. As a result, each of them constitutes an *independent statistical object* and the infinite sampling regime may be attained just by increasing the photon number. Importantly, such an assumption is not valid when considering general quantum systems, for which we would like to investigate the positive impact of correlations in between the constituent particles. Thus, in the quantum setting we must not only account for the quantum nature of the process, but also slightly modify the overall estimation scheme, so that the regime of independently distributed data can be clearly identified.

3.2 Quantum estimation theory

3.2.1 The parameter estimation problem in the quantum setting

In the quantum setting, we consider a general estimation scenario [Giovannetti *et al.*, 2004, 2006] depicted in Fig. 3.6, which for compatibility with the classical estimation problem of Sec. 3.1.1 is only constrained so that the estimated parameter φ is encoded *independently* onto each of the N particles contained in the initial, *input state* of the system: ρ_{in}^N (see Sec. 2.1.1 for introduction to the density-matrix description). As a result, φ may in general represent any *latent* variable discussed in Sec. 2.3 parametrising the *single-particle* evolution. Yet, before considering a general channel estimation scenario, we restrict ourselves in Fig. 3.6 to the *phase estimation problem*, in which the parameter specifies the "angle" of a *unitary* rotation[17]: $\mathcal{U}_\varphi[\bullet] = U_\varphi \bullet U_\varphi^\dagger$ with $U_\varphi = \mathrm{e}^{-\mathrm{i}\hat{H}\varphi}$, generated by some single-particle Hamiltonian $\hat{H}$. Such a choice allows us to clearly explain the mechanism behind the quantum enhancement of attainable precision, but the frameworks described in this section apply also to non-unitary estimation tasks, e.g. see the decoherence-strength estimation scenarios discussed in Chap. 4.5. Nevertheless, the scheme of Fig. 3.6 importantly accounts for the noise present in the setup, which we represent by a general quantum channel $\mathcal{D}$ (see Sec. 2.2.1) modelling the decoherence that may affect the particles in an uncorrelated, (**i**), or correlated, (**ii**), manner. One should note that such a picture is an oversimplification for quantum systems in which the parameter-encoding (here unitary) part of the evolution cannot be separated from the noise, i.e. when they do not commute with one another [Chaves *et al.*, 2013]. In the case of uncorrelated noise and the frequentist approach, however, we show later in Chap. 4 that such an issue becomes irrelevant after utilising the language of φ-parametrised quantum channels. Generally for the scenario of Fig. 3.6, we may write the final, *output state* of the whole system both as $\rho_\varphi^N = \mathcal{D}\big[\mathcal{U}_\varphi^{\otimes N}[\rho_{\text{in}}^N]\big] = \mathcal{U}_\varphi^{\otimes N}\big[\mathcal{D}[\rho_{\text{in}}^N]\big]$,

---

[17]Following the notation of Chap. 2, we do not alter the fount when labelling the *operators* acting on quantum wave-functions – e.g. U, $\mathbb{I}$ and K respectively representing unitary, identity and Kraus operators, whereas we utilise the calligraphic fount to denote *super-operators* acting on density matrices – e.g. $\mathcal{U}[\bullet] = U \bullet U^\dagger$, $\mathcal{I}[\bullet] = \bullet$ and $\mathcal{D}[\bullet] = \sum_i K_i \bullet K_i^\dagger$ now corresponding respectively to unitary, identity super-operators and a general quantum channel.
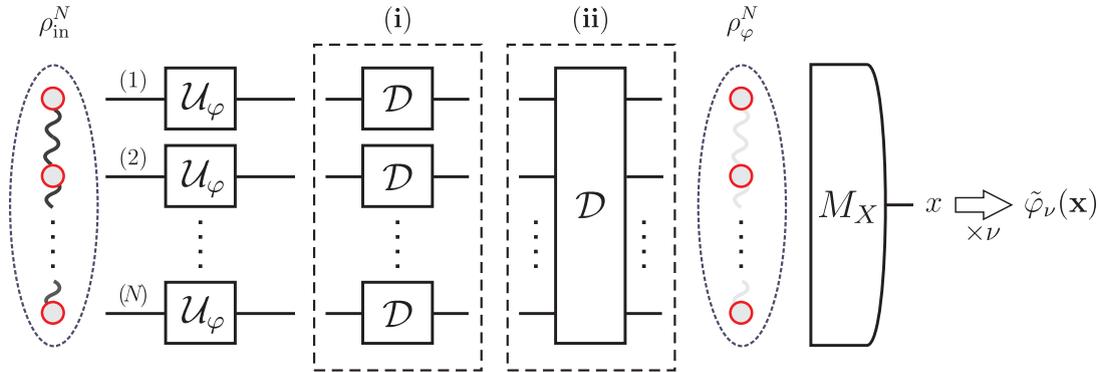
FIGURE 3.6: **Noisy phase estimation scheme in the quantum setting** [Giovannetti *et al.*, 2004, 2006]. The system is prepared in an *input state* ρ_{in}^N consisting of N particles. The parameter is then *independently* encoded onto each of the constituent particles as a *phase* dictated by the unitary transformation $\mathcal{U}_\varphi$. During the process, the system is affected by the *noise* $\mathcal{D}$ that disturbs the particles in an uncorrelated, (**i**), or correlated, (**ii**), fashion. Afterwards, a *quantum measurement* M_X is performed on the system *output state* ρ_φ^N yielding an outcome x. If necessary, the whole procedure is *repeated* ν times, in order to assure the independent character of the sample collected: $\mathbf{x} = \{x_1, x_2, \ldots, x_\nu\}$. Finally, an *estimator* $\tilde{\varphi}_\nu(\mathbf{x})$ is constructed on the data, which performance may be determined and bounded with use of the classical estimation techniques of Sec. 3.1.

where in the case of uncorrelated noise $\mathcal{D}$ also factorises to $\mathcal{D}^{\otimes N}$. A quantum measurement M_X that is performed on the *whole* system yields in principle a *single* outcome described by a random variable X, which is then distributed according to the PDF $p_\varphi^N(x) = \text{Tr}\{\rho_\varphi^N M_x\}$. The quantum measurement operators M_x correspond to the elements of a *Positive Operator Valued Measure* (POVM) introduced in Sec. 2.1.2 that are positive semi-definite, $M_x \geq 0$, and form a complete basis by satisfying either $\int \mathrm{d}x\, M_x = \mathbb{I}$ or $\sum_x M_x = \mathbb{I}$ for respectively continuous or discrete (and hence more experimentally relevant) set of the outcomes[18]. M_X is assumed to be *collective* (acting on all the constituent particles) so that the description encompasses all the potential measurement scenarios, e.g. the ones in which the particles are measured separately and $M_X = \bigotimes_{n=1}^{N} M_{X_n}^{(n)}$, but also the *adaptive schemes* in which after performing measurements on some fraction of the constituent particles further measurements may be adjusted basing on the outcomes already gathered [Wiseman *et al.*, 2009; Wiseman and Killip, 1997, 1998].

Notice that the quantum character of the process has impact only on the outcome statistics and plays no role at the data inference stage, during which all the techniques developed previously in Sec. 3.1 apply. However, if one requires the final data set obtained to be *independently distributed* and thus equivalent to the one introduced in Sec. 3.1.1, the whole estimation procedure must be in principle *repeated* ν *times*, so that $\mathbf{x} = \{x_1, x_2, \ldots, x_\nu\}$. In any case, the procedure is completed as before by constructing an optimal estimator $\tilde{\varphi}_\nu(\mathbf{x})$ most accurately predicting the parameter value. Yet, in general, it is now the repetition number ν that describes the sample size and *not* the particle number N. As a consequence, the asymptotic regime of the frequentist approach corresponds to the $\nu \to \infty$ limit and it is in principle *not* enough to consider sufficiently large N in order to guarantee the locality of estimation, as in the previous case of Sec. 3.1.2.3. Moreover, only in the $\nu \to \infty$ limit the Bayesian results may be associated with the frequentist ones, as only then the corresponding estimators are guaranteed to converge and the minimal $\overline{\text{MSE}}$ (3.31) may be related to the CRB (3.7) via Eq. (3.34).

---

[18]We will use the notation $\oint \mathrm{d}x \ldots$ to represent both the integral and the sum over the outcomes, in order to indicate that the analysis is valid regardless of their continuous or discrete nature.

However, for the special case of: *product input states* $\rho_{\text{in}}^N = \rho_{\text{in}}^{\otimes N}$, *uncorrelated noise* $\mathcal{D}^{\otimes N}$, and *separable measurements* $M_X = \bigotimes_{n=1}^N M_{X_n}^{(n)}$; the PDF dictating the outcome statistics factorises with $p_\varphi(x) = \prod_{n=1}^N p_\varphi(x_n)$ and $p_\varphi(x_n) = \text{Tr}\left\{\mathcal{D}[\rho_{\text{in}}] M_{x_n}^{(n)}\right\}$, so that we recover in "one shot" the classical parameter estimation problem of Sec. 3.1.1 with an N-point, independently distributed data set $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$. Thus, in such a *classical scenario*[19], we may set without loss of generality $\nu = 1$, as it is indeed the particle number N that plays the role of the sample size. In particular, such an observation explicitly proves that we correctly assumed the estimation problem to be classical in the Mach-Zehnder interferometer example of Sec. 3.1.4, where we considered: the input to be in the Fock state $|N\rangle$ that (as shown in Note 2.3) is a product state of photons, no noise to be present in the setup, and a photon-number–counting measurement which does not explore any of the photonic correlations.

In conclusion, the *quantum parameter estimation problem* may always be divided into its quantum and classical parts. In the quantum part, one must find the *optimal input states* leading to the highest parameter sensitivity, but also perform a non-trivial optimisation over the class of all POVMs to find the *best measurement scheme*. Meanwhile, the outcomes collected must be *efficiently interpreted* via the classical estimation methods to yield the maximal precision, which then may be compared with the corresponding classical scenario, in order to quantify the quantum enhancement and, in particular, verify if one is able to surpass the $1/N$ *SQL-like scaling* with the particle number. However, the tools of *quantum estimation theory* often allow to circumvent the classical part of the estimation procedure. Within the frequentist approach, the *Quantum Cramér-Rao Bound* dictates fundamental local limits on precision that are already optimised over both the quantum measurement strategies and the estimators. The quantum Bayesian methods, on the other hand, allow to benefit from the symmetry of a given estimation problem by utilizing the structure of the so-called *covariant POVMs*, which are already designed in a way to incorporate the best data inference strategy and simplify the procedure of the measurement-scheme optimisation. We describe these tools in detail below.

3.2.2 Frequentist approach – *local* estimation of a *deterministic* parameter

3.2.2.1 Quantum Cramér-Rao Bound and Quantum Fisher Information

As shown in the previous section, after specifying the system input state and a particular measurement scheme, any quantum estimation problem becomes fully classical with the only difference being the potential necessity of procedure repetitions to assure the independent character of the data gathered. Hence, the CRB (3.7) of Sec. 3.1.2 naturally applies at this stage and lower-bounds the MSE (3.3) of any locally unbiased estimator constructed on the outcomes. However, one may always determine a further, ultimate lower bound, i.e. the *Quantum Cramér-Rao Bound* (QCRB) [Helstrom, 1976; Holevo, 1982], which importantly is valid regardless of the measurement strategy chosen. Following the notation of Fig. 3.6, it reads:

$$\Delta^2 \tilde{\varphi}_\nu \geq \frac{1}{\nu F_{\text{Q}}[\rho_\varphi^N]}, \qquad \text{where} \qquad F_{\text{Q}}[\rho_\varphi^N] = \text{Tr}\left\{\rho_\varphi^N L_{\text{S}}[\rho_\varphi^N]^2\right\} \tag{3.55}$$

is the *Quantum Fisher Information* (QFI) that is solely determined by the dependence of ρ_φ^N on the estimated parameter. The Hermitian operator L_{S} is the *Symmetric Logarithmic Derivative* (SLD), which

---

[19]In general, the classical setting also allows for *classical correlations* in between the particles and their measurements, which are not accounted for in the above scenario. Yet, we show later in Sec. 3.2.2.3 that these are insufficient to surpass the asymptotic SQL-like scaling of precision with N even for $\nu \to \infty$, so that for simplicity we assume their absence in any scenario that we name later in this work to be *classical*.

can be unambiguously defined for any state ϱ_φ via the relation[20] $\dot{\varrho}_\varphi = \frac{1}{2}\left(\varrho_\varphi L_{\mathrm{S}}[\varrho_\varphi] + L_{\mathrm{S}}[\varrho_\varphi]\varrho_\varphi\right)$, so that in the eigenbasis of $\varrho_\varphi = \sum_i \lambda_i(\varphi)\,|e_i(\varphi)\rangle\langle e_i(\varphi)|$ with $\{|e_i(\varphi)\rangle\}_i$ forming a complete basis $(\forall_i : 0 \le \lambda_i \le 1)$:

$$L_{\mathrm{S}}[\varrho_\varphi] = \sum_{\substack{i,j \\ \lambda_i + \lambda_j \neq 0}} \frac{2\,\langle e_i(\varphi)|\,\dot{\varrho}_\varphi\,|e_j(\varphi)\rangle}{\lambda_i(\varphi) + \lambda_j(\varphi)}\,|e_i(\varphi)\rangle\langle e_j(\varphi)|. \tag{3.56}$$

In other words, for any concrete POVM yielding for ρ_φ^N a PDF $p_\varphi^N(x)$, the QFI upper-limits the corresponding classical FI (3.8), so that[21] $F_{\mathrm{cl}}[p_\varphi^N] \le F_{\mathrm{Q}}[\rho_\varphi^N]$. Most importantly, however, there always exists a measurement strategy [Braunstein and Caves, 1994; Nagaoka, 1989]—a projective measurement in the eigenbasis of the SLD (3.56), $L_{\mathrm{S}}[\rho_\varphi^N] = \sum_i \mu_i(\varphi)\,|f_i(\varphi)\rangle\langle f_i(\varphi)|$, with POVM elements reading $M_i(\varphi) = |f_i(\varphi)\rangle\langle f_i(\varphi)|$—for which the F_{cl} attains the F_{Q}. One should note that such an optimal POVM is in principle not only collective, i.e. acting on *all* the particles, but also, as emphasised by the notation, it depends on the true value of the estimated parameter, what is consistent with the locality assumption of the frequentist approach. On the other hand, one may thus always interpret the QCRB (3.55) as the classical CRB (3.7) evaluated for such an optimal measurement scheme. Hence, all the discussions of Sec. 3.1.2.4 analysing the CRB-saturability issues apply, meaning that the QCRB is assured to be tight only within the regime of local estimation discussed in Sec. 3.1.2.3, which may always be guaranteed in the $\nu \to \infty$ limit of infinitely many procedure repetitions.

Moreover, the QFI may be interpreted similarly to the FI (3.8) as an *information measure* [Barndorff-Nielsen and Gill, 2000], which now is generalised from the space of PDFs to the space of quantum states. In fact, writing explicitly the form of the QFI according to Eq. (3.55) for the above introduced ϱ_φ as

$$F_{\mathrm{Q}}[\varrho_\varphi] = \sum_{\substack{i \\ \lambda_i \neq 0}} \frac{\dot{\lambda}_i(\varphi)^2}{\lambda_i(\varphi)} + 2 \sum_{\substack{i,j \\ \lambda_i + \lambda_j \neq 0}} \frac{[\lambda_i(\varphi) - \lambda_j(\varphi)]^2}{\lambda_i(\varphi) + \lambda_j(\varphi)}\,|\langle \dot{e}_i(\varphi)|e_j(\varphi)\rangle|^2, \tag{3.57}$$

one realizes that F_{Q} splits into its classical and quantum parts. The first, "classical term" in Eq. (3.57) represents the classical FI, $F_{\mathrm{cl}}[\lambda(\varphi)]$, quantifying the information about the parameter encoded in the discrete PDF of the eigenvalues $\lambda_i(\varphi)$, whereas the second, non-negative term accounts for the quantum nature of the state adding a contribution from the rotation of the eigenvectors with the parameter change. Consistently, for any classical state with an invariant eigenbasis, $\varrho_\varphi^{\mathrm{cl}} = \sum_i p_{\varphi,i}\,|e_i\rangle\langle e_i|$, the QFI (3.57) reproduces the FI, i.e. the discrete version of Eq. (3.8), as the "quantum term" in Eq. (3.57) then vanishes and $F_{\mathrm{Q}}[\varrho_\varphi^{\mathrm{cl}}] = F_{\mathrm{cl}}[p_\varphi]$. On the other hand, in the case of pure quantum states, $\varrho_\varphi = |\psi_\varphi\rangle\langle\psi_\varphi|$, for which the "classical term" in Eq. (3.57) is absent, the SLD (3.56) takes a more appealing form: $L_{\mathrm{S}}[|\psi_\varphi\rangle] = 2\left(|\dot{\psi}_\varphi\rangle\langle\psi_\varphi| + |\psi_\varphi\rangle\langle\dot{\psi}_\varphi|\right)$, so that the expression for the QFI simplifies dramatically to:[22]

$$F_{\mathrm{Q}}[|\psi_\varphi\rangle] = 4\left(\langle\dot{\psi}_\varphi|\dot{\psi}_\varphi\rangle - |\langle\dot{\psi}_\varphi|\psi_\varphi\rangle|^2\right). \tag{3.58}$$

Similarly to the FI, the QFI is a *local* quantity, as for a given parameter true value φ_0 it is fully specified by ϱ_{φ_0} and $\dot{\varrho}_{\varphi_0}$, what may be verified by inspecting Eqs. (3.55) and (3.57). Thus, any two quantum states $\varrho_{\varphi,1}$ and $\varrho_{\varphi,2}$ are *equivalent* at a given φ_0 from the QFI perspective, as long as they coincide there up to $O(\delta\varphi^2)$, so that $\varrho_{\varphi_0,1} = \varrho_{\varphi_0,2}$ and $\dot{\varrho}_{\varphi_0,1} = \dot{\varrho}_{\varphi_0,2}$ yielding $F_{\mathrm{Q}}[\varrho_{\varphi,1}]|_{\varphi_0} = F_{\mathrm{Q}}[\varrho_{\varphi,2}]|_{\varphi_0}$. Consequently, the geometrical interpretation of the FI and Eq. (3.10) may also be naturally carried

---

[20]One may show that *non-symmetric* (or in other words non-Hermitian) logarithmic derivatives satisfying $\dot{\varrho}_\varphi = \frac{1}{2}\left(\varrho_\varphi \tilde{L}_{\mathrm{S}}[\varrho_\varphi] + \tilde{L}_{\mathrm{S}}[\varrho_\varphi]^\dagger \varrho_\varphi\right)$ may only lead to weaker bounds, i.e. $F_{\mathrm{Q}}[\rho_\varphi^N] \le F_{\tilde{L}}[\rho_\varphi^N]$ [Hayashi, 2005; Holevo, 1982].

[21]Notice that in contrast to the classical estimation problem of Sec. 3.1.1 we may not assume here $p_\varphi^N(x)$ to be factorisable.

[22]For simplicity, we shorten the notation of functions and super-operators of pure states, so that e.g. $F[|\psi\rangle] \equiv F[|\psi\rangle\langle\psi|]$, $D(|\psi\rangle,|\phi\rangle) \equiv D(|\psi\rangle\langle\psi|,|\phi\rangle\langle\phi|)$ and $\mathcal{U}[|\psi\rangle] \equiv \mathcal{U}[|\psi\rangle\langle\psi|]$.

over into the quantum picture [Barndorff-Nielsen and Gill, 2000; Braunstein and Caves, 1994; Hayashi, 2005][23]. The previously introduced angular distance $D_{\rm cl}$ between PDFs may be directly generalised to quantum states [Wootters, 1981] to obtain the so-called *quantum (Bures) angular distance* [Bengtsson and Życzkowski, 2006]: $D_{\rm Q}(\varrho_1, \varrho_2) = \arccos[{\sf Fid}_{\rm Q}(\varrho_1, \varrho_2)]$ with ${\sf Fid}_{\rm Q}(\varrho_1, \varrho_2) = {\rm Tr}\{\sqrt{\sqrt{\varrho_1}\varrho_2\sqrt{\varrho_1}}\}$ representing now the *quantum fidelity* [Nielsen and Chuang, 2000] defined for any two states $\varrho_{1/2}$. Notice that in the classical case of states sharing a common eigenbasis, i.e. $\varrho_{1/2}^{\rm cl} = \sum_i p_{1/2,i} |e_i\rangle\langle e_i|$, $D_{\rm Q}$ accordingly reproduces the classical angular distance as[24] $D_{\rm Q}(\varrho_1^{\rm cl}, \varrho_2^{\rm cl}) = D_{\rm cl}(p_1, p_2)$. Importantly, expanding the quantum angular distance between the neighbouring states ϱ_{φ_0} and $\varrho_{\varphi_0+\delta\varphi}$ for small $\delta\varphi \geq 0$, we naturally obtain the quantum equivalent of Eq. (3.10):

$$D_{\rm Q}(\varrho_{\varphi_0}, \varrho_{\varphi_0+\delta\varphi}) = \frac{1}{2}\sqrt{F_{\rm Q}[\varrho_\varphi]|_{\varphi_0}}\, \delta\varphi + O(\delta\varphi^2),\qquad(3.59)$$

so that the QFI (3.57) may be similarly interpreted as the the square of the speed, $F_{\rm Q}[\varrho_\varphi] = 4\,({\rm d}D_{\rm Q}/{\rm d}\varphi)^2$, with which the quantum state ϱ_φ is "moving" along the path of φ for a given parameter value[25].

In the case of the phase estimation scheme of Fig. 3.6, in which the parameter-encoding part of the evolution is assumed to be unitary and commuting with the noise, the output state may be written as $\rho_\varphi^N = \mathcal{U}_\varphi^{\otimes N}[\rho_0^N]$ with $\rho_0^N = \mathcal{D}[\rho_{\rm in}^N]$, $U_\varphi^{\otimes N} = \bigotimes_{n=1}^N {\rm e}^{-{\rm i}\hat{H}^{(n)}\varphi} = {\rm e}^{-{\rm i}\hat{H}_N\varphi}$ and the N-particle Hamiltonian defined as $\hat{H}_N = \sum_{n=1}^N \hat{H}^{(n)}$. Hence, the expression for the QFI simplifies, as performing the eigendecomposition of $\rho_0^N = \sum_i \lambda_i |e_i\rangle\langle e_i|$ we only obtain the "quantum term" in Eq. (3.57):

$$F_{\rm Q}[\rho_\varphi^N] = 2 \sum_{\substack{i,j \\ \lambda_i+\lambda_j\neq 0}} \frac{(\lambda_i - \lambda_j)^2}{\lambda_i + \lambda_j} \left|\langle e_i|\hat{H}_N|e_j\rangle\right|^2,\qquad(3.60)$$

which due to the *unitary* parameter-encoding is always parameter independent. Importantly, in order to maximise the precision of the protocol, we must seek the optimal input states that maximise the QFI (3.60) and thus minimise the QCRB (3.55). In general, due to the presence of noise, such an optimisation task is non-trivial. Yet, when the noise is absent and the input state is assumed to be pure, so that $\rho_0^N = |\psi^N\rangle\langle\psi^N|$ and $\rho_\varphi^N = |\psi_\varphi^N\rangle\langle\psi_\varphi^N|$ with $|\psi_\varphi^N\rangle = {\rm e}^{-{\rm i}\hat{H}_N\varphi}|\psi^N\rangle$, Eq. (3.60) simplifies further and the QFI becomes proportional to the *variance of the Hamiltonian* considered:

$$F_{\rm Q}[|\psi_\varphi^N\rangle] = 4\Delta^2\hat{H}_N = 4\left(\langle\psi^N|\hat{H}_N^2|\psi^N\rangle - \langle\psi^N|\hat{H}_N|\psi^N\rangle^2\right).\qquad(3.61)$$

As a result, one may directly see that the optimal input state maximising the variance of $\hat{H}_N$, and hence the QFI (3.61), is an equally weighted superposition of the eigenvectors corresponding to the Hamiltonian minimal and maximal eigenvalues, i.e. $|\psi_{\rm in}^N\rangle_{\rm opt} = \frac{1}{\sqrt{2}}(|\mu_{\rm min}^N\rangle + {\rm e}^{{\rm i}\phi}|\mu_{\rm max}^N\rangle)$ with arbitrary ϕ, where $\hat{H}_N|\mu^N\rangle = \mu^N|\mu^N\rangle$, for which the maximal QFI thus reads $F_{\rm Q}[{\rm e}^{-{\rm i}\hat{H}_N\varphi}|\psi_{\rm in}^N\rangle_{\rm opt}] = (\mu_{\rm max}^N - \mu_{\rm min}^N)^2$ [Giovannetti *et al.*, 2006].

---

[23]For geometrical approach to quantum estimation and information theory also see [Amari and Nagaoka, 2007; Petz, 1996, 2002; Petz and Sudar, 1999].

[24]For pure states, on the other hand, $D_{\rm Q}(|\psi\rangle, |\phi\rangle) = \arccos|\langle\psi|\phi\rangle|$ reproduces the so-called Fubini-Study metric, which is the natural metric used in the geometric approaches to quantum mechanics [Bengtsson and Życzkowski, 2006], for instance widely utilised in study of the *Berry phase* phenomenon [Anandan and Aharonov, 1990].

[25]For completeness, let us comment on the consequences if the *quantum relative entropy* [Bengtsson and Życzkowski, 2006; Nielsen and Chuang, 2000] was used instead in Eqs. (3.10) and (3.59) as an (asymmetric) distance measure between neighbouring PDFs and quantum states, which is the adequate quantity employed in the, complementary to estimation, *hypothesis testing* problems [Bengtsson and Życzkowski, 2006; Helstrom, 1976]. Importantly, it would also locally reproduce the FI at the classical level, but *not* the QFI in the quantum case [Hayashi, 2005]. Such a behaviour indicates that one may interpret any classical local-estimation problem as an (asymmetric) discrimination task between two infinitely close PDFs differing by $\delta\varphi$, but such an interpretation does *not* generalise to quantum states [Hayashi, 2005].

Furthermore, one should also note that for the QFI (3.61) and a single repetition $\nu = 1$, the QCRB (3.55) takes an appealing form of a *"time-energy"–like uncertainty relation* for the latent parameter (see Note 2.5):

$$\Delta^2 \hat{H}_N \, \Delta^2 \tilde{\varphi} \; \geq \; \frac{1}{4}, \tag{3.62}$$

which importantly is *not* inferred from a quantum observable with help of the error-propagation formula (2.22), but rather directly determined by the parametrisation of the quantum state ϱ_φ that dictates the local estimation capabilities, i.e. the speed at which the system is "moving" with variations in φ.

3.2.2.2 Purification-based definitions of the QFI

As indicated by Eq. (3.57), for mixed states the computation of the QFI in principle involves diagonal-isation of the density matrix, what in the case of a general output state ρ_φ^N of Fig. 3.6 may be already infeasible for moderate N due to the dimension of the system Hilbert space, $\mathcal{H}_\mathrm{S}$, growing exponentially with the number of particles. However, in order to potentially circumvent this problem, alternative definitions of the QFI were proposed that do not require the eigen-decomposition, but are specified at the level of state purifications, i.e. any $|\tilde{\Psi}_\varphi\rangle$ such that $\varrho_\varphi = \mathrm{Tr}_\mathrm{E}\{|\tilde{\Psi}_\varphi\rangle\langle\tilde{\Psi}_\varphi|\}$ with E denoting the ancillary part (environment) of the extended system Hilbert space, $\mathcal{H}_\mathrm{S} \otimes \mathcal{H}_\mathrm{E}$, required for the purification. Inter-estingly, in [Escher *et al.*, 2011] the QFI of any ϱ_φ has been proved to be equal to the smallest QFI of its purifications (see also App. C for an alternative proof):

$$F_\mathrm{Q}[\varrho_\varphi] = \min_{\tilde{\Psi}_\varphi} F_\mathrm{Q}[|\tilde{\Psi}_\varphi\rangle] = 4 \min_{\tilde{\Psi}_\varphi} \left\{ \left\langle \dot{\tilde{\Psi}}_\varphi \,\middle|\, \dot{\tilde{\Psi}}_\varphi \right\rangle - \left| \left\langle \tilde{\Psi}_\varphi \,\middle|\, \dot{\tilde{\Psi}}_\varphi \right\rangle \right|^2 \right\}, \tag{3.63}$$

whereas in [Fujiwara and Imai, 2008] another purification-based QFI definition has been constructed:

$$F_\mathrm{Q}[\varrho_\varphi] = 4 \min_{\tilde{\Psi}_\varphi} \left\langle \dot{\tilde{\Psi}}_\varphi \,\middle|\, \dot{\tilde{\Psi}}_\varphi \right\rangle. \tag{3.64}$$

Despite the apparent difference, the above definitions are indeed equivalent, as one may prove (see App. C) that any purification minimizing one of them is likewise optimal for the other and satisfies the condition $\left|\dot{\tilde{\Psi}}_\varphi^\mathrm{opt}\right\rangle = \frac{1}{2} L_\mathrm{S}[\varrho_\varphi] \otimes \mathbb{I}^\mathrm{E} \left|\tilde{\Psi}_\varphi^\mathrm{opt}\right\rangle$ causing the second term in Eq. (3.63) to vanish. The minimisation over all purifications, $\min_{\tilde{\Psi}_\varphi}$, may at first sight incorrectly seem as an abstract and thus a non-computable procedure. However, owing to the local character of the QFI, also the minimised terms on the r.h.s. of Eqs. (3.63) and (3.64) are sensitive only to small—up to $O(\delta\varphi^2)$—variations of φ around a given φ_0. Consequently, they locally depend only on: $|\tilde{\Psi}_{\varphi_0}\rangle$ which is fixed due to $\varrho_{\varphi_0} = \mathrm{Tr}_\mathrm{E}\{|\tilde{\Psi}_{\varphi_0}\rangle\langle\tilde{\Psi}_{\varphi_0}|\}$, and $|\dot{\tilde{\Psi}}_{\varphi_0}\rangle$ which is the only one that may vary between purifications. That is why, the optimisation can always be systematically performed starting from any purification valid at φ_0, say $|\Psi_\varphi\rangle$[26], and searching through purifications that are generated by a unitary rotation of the environment subspace such that $\left|\tilde{\Psi}_\varphi\right\rangle = u_\varphi^\mathrm{E}|\Psi_\varphi\rangle$ with $u_\varphi^\mathrm{E} = \mathrm{e}^{-\mathrm{i}\hat{h}_\mathrm{E}(\varphi - \varphi_0)}$, which cover all the necessary potential shifts of the first derivative at φ_0, as then: $|\tilde{\Psi}_{\varphi_0}\rangle = |\Psi_{\varphi_0}\rangle$ and $\left|\dot{\tilde{\Psi}}_{\varphi_0}\right\rangle = |\dot{\Psi}_{\varphi_0}\rangle - \mathrm{i}\hat{h}_\mathrm{E}|\Psi_{\varphi_0}\rangle$. Crucially, the minimisation over purifications at a given φ_0 is thus equivalent to the optimisation over all Hermitian generators $\hat{h}_\mathrm{E}$, which importantly are of dimension equal to the rank of the density matrix ϱ_{φ_0} and not its size, what leaves room for potential numerical efficiency.

---

[26]In fact, the minimisation at φ_0 requires only the knowledge of $|\Psi_{\varphi_0}\rangle$ and $|\dot{\Psi}_{\varphi_0}\rangle$.

Note 3.4: Purification-based QFI of a pure state when estimating phase.

As an example, let us consider the simplest example of the phase estimation scheme of Fig. 3.6 with the noise absent and a pure input state, for which $\rho_\varphi^N = |\psi_\varphi^N\rangle\langle\psi_\varphi^N|$ with $|\psi_\varphi^N\rangle = \mathrm{e}^{-\mathrm{i}\hat{H}\varphi}|\psi^N\rangle$. As the output state is pure—rank-one—all the relevant purifications are generated by just varying the phase of $|\psi_\varphi^N\rangle$, which from the point of view of the state— but *not* its derivative—is irrelevant, i.e. $|\tilde{\psi}_\varphi^N\rangle = \mathrm{e}^{-\mathrm{i}h_{\mathrm{E}}(\varphi-\varphi_0)}|\psi_\varphi^N\rangle$ with h_{E} now being an arbitrary scalar variable. Eqs. (3.63) and (3.64) then respectively read

$$4\min_{h_{\mathrm{E}}}\left\{\langle\psi_\varphi^N|\big(\hat{H}+h_{\mathrm{E}}\big)^2|\psi_\varphi^N\rangle - \big|\langle\psi_\varphi^N|\big(\hat{H}+h_{\mathrm{E}}\big)|\psi_\varphi^N\rangle\big|^2\right\} \quad\text{and}\quad 4\min_{h_{\mathrm{E}}}\left\{\langle\psi_\varphi^N|\big(\hat{H}+h_{\mathrm{E}}\big)^2|\psi_\varphi^N\rangle\right\}, \tag{3.65}$$

and consistently simplify to $F_{\mathrm{Q}}\big[|\psi_\varphi^N\rangle\big] = 4\Delta^2\hat{H}$ of Eq. (3.61) for the optimal $h_{\mathrm{E}}^{\mathrm{opt}} = -\langle\psi_\varphi^N|\hat{H}|\psi_\varphi^N\rangle$. Moreover, $h_{\mathrm{E}}^{\mathrm{opt}}$ can be equivalently determined by solving the adequate necessary condition $\big|\dot{\tilde{\psi}}_\varphi^N\big\rangle = \frac{1}{2}L_{\mathrm{S}}\big[|\psi_\varphi^N\rangle\big]|\tilde{\psi}_\varphi^N\rangle$ with the SLD (3.56) taking a simpler form: $L_{\mathrm{S}}\big[|\psi_\varphi^N\rangle\big] = 2\mathrm{i}\big[|\psi_\varphi^N\rangle\langle\psi_\varphi^N|, \hat{H}\big]$, for the unitary encoding. Although such an example may seem trivial as h_{E} does not constitute an operator, it gives the correct intuition about the role of the optimal-purification generator. $h_{\mathrm{E}}^{\mathrm{opt}}$ produces a counter-rotation of the state phase, which may be interpreted as an erasure operation that minimises the information about the parameter encoded in the arbitrary phase—more generally the environment. We elaborate more on this issue in Sec. 4.2, where we apply the purifications-based QFI definitions to quantum channels.

On the other hand, we may diminish the number of free variables one has to minimise over, when utilising the purification-based QFI definitions, by restricting to some subclass of e.g. physically motivated generators $\hat{h}_{\mathrm{E}}$. As a result, Eqs. (3.63) and (3.64) may also serve as an effective tool for establishing upper bounds on $F_{\mathrm{Q}}[\varrho_\varphi]$. Although the definition (3.64) can never provide for some suboptimal $|\tilde{\Psi}_\varphi\rangle$ a tighter upper bound on the QFI than Eq. (3.63), it allows for more agility when applied to channel-estimation scenarios discussed in Chap. 4 [Fujiwara and Imai, 2008]. That is why, we utilise Eq. (3.64) explicitly within this work, in particular, when reformulating later the relevant purification-minimisation procedures into *semi-definite programs* (SDPs).

3.2.2.3 Key properties of the QFI and their consequences

We discuss the key properties of the QFI (3.57) that play an important role when studying the ultimate bounds on precision dictated by the QCRB (3.55) and, in particular, when analysing the impact of the correlations (see Secs. 2.1.3 and 2.1.4) in between the constituent particles of the system on the potentially achievable precision-scaling with the particle number N.

Non-negativity and additivity

As shown already in Sec. 3.2.2.1, the QFI may be treated as not only an upper bound on the FI (3.8), but also as the FI itself corresponding to the optimal measurement strategy. Such fact naturally makes the QFI an information measure [Barndorff-Nielsen and Gill, 2000] that must be *non-negative*, what is indeed assured by both ϱ_φ and $L_{\mathrm{S}}[\varrho_\varphi]^2$ in Eq. (3.55) being positive semi-definite matrices. Most importantly, the QFI also generalises the notion of *additivity* of the FI from PDFs to density matrices, what may be verified by considering a general bipartite, parameter-dependent, product state $\varrho_\varphi^{\mathrm{AB}} = \varrho_\varphi^{\mathrm{A}} \otimes \varrho_\varphi^{\mathrm{B}}$, for which then $F_{\mathrm{Q}}\big[\varrho_\varphi^{\mathrm{A}} \otimes \varrho_\varphi^{\mathrm{B}}\big] = F_{\mathrm{Q}}\big[\varrho_\varphi^{\mathrm{A}}\big] + F_{\mathrm{Q}}\big[\varrho_\varphi^{\mathrm{B}}\big]$. Hence, in particular, for tensor product states $F_{\mathrm{Q}}\big[\varrho_\varphi^{\otimes m}\big] = m\,F_{\mathrm{Q}}[\varrho_\varphi]$, indicating that by having access to identical copies of a given state we may at most observe a linear precision improvement of the QCRB leading to the *SQL-like scaling*. Furthermore, this shows that, at the level of the frequentist bound, the scenario of possessing m uncorrelated copies of a system is fully equivalent to the protocol with $\nu = m$ estimation-procedure repetitions accounted for in the scheme of Fig. 3.6. The additivity property thus *proves* that when estimating locally or in the asymptotic m limit, i.e. when the QCRB is guaranteed to be tight, one may *not* benefit from collective measurements that utilise all the available copies, as the same precision must also be achievable performing single-copy measurements that mimic procedure repetitions.

As a result, when investigating the phase estimation scheme of Fig. 3.6 with particles and noise assumed to be uncorrelated—for which most generally the output state reads $\rho_\varphi^N = \otimes_{n=1}^N \rho_\varphi^{(n)}$, so that each particle $\rho_\varphi^{(n)}$ may be effectively treated as a separate "copy"[27]—the additivity of the QFI assures the QCRB to be saturated in the asymptotic N limit without need of procedure repetitions ($\nu = 1$) and, most notably, without use of collective measurements performed on all the particles. However, as soon as we allow for *correlations* in between the particles of the input state such that the output is still *separable* but no longer a product state, i.e. $\rho_\varphi^N = \sum_i p_i \otimes_{n=1}^N \rho_{\varphi,i}^{(n)}$, we may not assume the optimal measurement POVMs to be particle-separable unless we again let $\nu \to \infty$. It is so, because only in the limit of infinitely many repetitions we may rewrite $\rho_\varphi^{N \otimes \nu}$ as a tensor product:

$$\rho_\varphi^{N \otimes \nu} = \left(\sum_i p_i \bigotimes_{n=1}^N \rho_{\varphi,i}^{(n)} \right)^{\otimes \nu} \quad \overset{\nu \to \infty}{\equiv} \quad \bigotimes_i \left(\bigotimes_{n=1}^N \rho_{\varphi,i}^{(n)} \right)^{\otimes p_i \nu}, \tag{3.66}$$

so that the probabilistic nature of ρ_φ^N may be ignored by letting the mixing probabilities p_i represent the fractions (frequencies) of various types (labelled by i) of the system states we possess [Giovannetti *et al.*, 2006]. Eq. (3.66) also shows that any *non-entanglement–like correlations*[28] do *not* allow to surpass the *SQL-like scaling* when $\nu \to \infty$, as due to additivity $F_Q\left[\rho_\varphi^{N \otimes \nu} \right] \overset{\nu \to \infty}{=} F_Q\left[\otimes_i \left(\otimes_{n=1}^N \rho_{\varphi,i}^{(n)} \right)^{\otimes p_i \nu} \right] = \nu \sum_{n=1}^N \sum_i p_i F_Q\left[\rho_{\varphi,i}^{(n)} \right] \le \nu N \, const$ and the QFI must[29] at most scale linearly in N. In fact, such a restrictive conclusion is true irrespectively of the number of repetitions ν, owing to the convexity property of the QFI discussed in the last paragraph of this section.

On the other hand, the argument of Eq. (3.66) fails when ρ_φ^N is *particle-entangled*, what may occur not only when one considers non-separable inputs, but also when dealing with collective noises capable of entangling the system particles during the evolution (e.g. effectively representing particle interactions). In such a situation, in order to reach the QCRB, collective measurements on all the particles are in principle required even in the asymptotic ν limit of many repetitions. Nevertheless, it may turn out that separable measurements are still sufficient in particular cases. For instance, it is so for the optimal scenario of Fig. 3.6 under the idealistic assumption of no noise [Giovannetti *et al.*, 2006] (when maximally entangled input states are optimal), which we discuss in Sec. 3.2.4 with the example of the quantum-enhanced Mach-Zehnder interferometer.

Monotonicity under parameter-independent quantum maps

One should note that the QFI is invariant under any unitary, φ-independent map, so that $\forall_{\mathcal{U}}: F_Q[\varrho_\varphi] = F_Q[\mathcal{U}[\varrho_\varphi]]$. This may be explicitly verified by realizing that the adjoint transformation $\mathcal{U}[\varrho_\varphi] = U \varrho_\varphi U^\dagger$ of the state results in an inverse mapping of the SLD (3.56), i.e. $L_S\left[U \varrho_\varphi U^\dagger \right] = U^\dagger L_S[\varrho_\varphi] U$, so that the expression (3.55) for the QFI is indeed unaffected by any $\mathcal{U}$. On the other hand, such invariance may be intuitively explained by exploring the fact that any unitary rotation is reversible. Hence, it can be always undone and thus also treated as a part of the measurement stage of the protocol, which is known not to have any influence on the QFI. Such notion, however, cannot be generalised to non-unitary quantum maps, i.e. the CPTP maps introduced in Sec. 2.2.1, which generally are irreversible and therefore should somehow affect the QFI. In fact, looking at the space of quantum states from the geometric perspective, any *parameter-independent CPTP map* Λ can only diminish the relative distance between any two states $\varrho_{1/2}$ [Bengtsson and Życzkowski, 2006; Petz, 1996, 2002; Petz and Sudar, 1999],

---

[27]For generality, we allow the particles to be in different states labelled by (n), what encompasses the most natural situation, when $\rho_\varphi^N = \rho_\varphi^{\otimes N}$ and we deal with N identical copies of a particle in a state ρ_φ.

[28]For a review of non-classical correlations without entanglement see for instance [Streltsov, 2015].

[29]Importantly, *const* is independent of N, as it may always be upper-bounded by $\max_n \left\{ \sum_i p_i F_Q\left[\rho_{\varphi,i}^{(n)} \right] \right\}$.

so that $D_Q(\Lambda[\varrho_1], \Lambda[\varrho_2]) \leq D_Q(\varrho_1, \varrho_2)$ and thus according to Eq. (3.59) the QFI may also only decrease under the action of Λ:

$$F_Q[\Lambda[\varrho_\varphi]] \leq F_Q[\varrho_\varphi]. \tag{3.67}$$

On the other hand, the above *monotonicity* property of the QFI can be straightforwardly proved by utilizing the purification-based definition (3.63), as follows

$$\begin{aligned}
F_Q[\Lambda[\varrho_\varphi]] &= F_Q\Big[\mathrm{Tr}_{E_\Lambda}\big\{\mathcal{U}^{SE_\Lambda}[\varrho_\varphi \otimes |\xi\rangle_{E_\Lambda}\langle\xi|]\big\}\Big] = F_Q\Big[\mathrm{Tr}_{E_\varrho E_\Lambda}\big\{\mathcal{U}^{SE_\Lambda} \otimes \mathcal{I}^{E_\varrho}\big[|\Psi_\varphi\rangle_{SE_\varrho}\langle\Psi_\varphi| \otimes |\xi\rangle_{E_\Lambda}\langle\xi|\big]\big\}\Big] \\
&= \min_{\tilde{\Psi}^{\mathrm{ext}}} F_Q\Big[\big|\tilde{\Psi}^{\mathrm{ext}}_\varphi\big\rangle_{SE_\varrho E_\Lambda}\Big] \quad \leq \quad \min_{\tilde{\Psi}} F_Q\Big[\big|\tilde{\Psi}_\varphi\big\rangle_{SE_\varrho}\Big] = F_Q[\rho_\varphi].
\end{aligned} \tag{3.68}$$

Following the prescription of Sec. 3.2.2.2, we have firstly purified the map Λ via the Stinespring dilation theorem (see Thm 2.2.1) and then the input state, so that $F_Q[\Lambda[\varrho_\varphi]]$ eventually equals the minimal QFI of the 'extended' purifications $\big|\tilde{\Psi}^{\mathrm{ext}}_\varphi\big\rangle = u^{E_\Lambda E_\varrho}_\varphi \big|\Psi^{\mathrm{ext}}_\varphi\big\rangle$ with $\big|\Psi^{\mathrm{ext}}_\varphi\big\rangle = (U^{SE_\Lambda} \otimes \mathbb{I}^{E_\varrho})|\Psi_\varphi\rangle_{SE_\varrho}|\xi\rangle_{E_\Lambda}$, which span the system subspace, $\mathcal{H}_S$, but also the ancillary ones $\mathcal{H}_{E_\Lambda}$ and $\mathcal{H}_{E_\varrho}$ introduced due to purifying the channel and the input respectively. Now, by considering 'extended' purifications that are generated *only* by unitary rotations in the subspace $\mathcal{H}_{E_\varrho}$, i.e. $\big|\check{\Psi}^{\mathrm{ext}}_\varphi\big\rangle = u^{E_\varrho}_\varphi\big|\Psi^{\mathrm{ext}}_\varphi\big\rangle$, and realizing that the unitary U^{SE_Λ} may then be ignored by the argument from the beginning of the paragraph, we conclude that such a subclass of the 'extended' purifications represents all the relevant purifications to be considered in the absence of the map Λ, as $F_Q\Big[\big|\check{\Psi}^{\mathrm{ext}}_\varphi\big\rangle_{SE_\varrho E_\Lambda}\Big] = F_Q\Big[\big|\tilde{\Psi}_\varphi\big\rangle_{SE_\varrho}\Big]$ with[30] $\big|\tilde{\Psi}_\varphi\big\rangle = u^{E_\varrho}_\varphi|\Psi_\varphi\rangle$. In other words, in order to reverse the action of Λ, we must impose further constraints on the 'extended' purifications when performing the minimisation in Eq. (3.68). As this may only increase the minimum obtained, the QFI could have been only diminished by the action of Λ in the first place, and therefore it must correspond to quantity that monotonically decreases under parameter-independent CPTP maps. ∎

Note 3.5: Monotonicity of the QFI under partial trace.

As the *partial trace* operation is an example of a quantum map (see Note 2.4), Eq. (3.67) also proves that after tracing-out some parts of a quantum state, e.g. the part supported by $\mathcal{H}_A$ of the state $\varrho^{AB}_\varphi \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the QFI may only decrease, i.e.

$$F_Q[\mathrm{Tr}_A\{\varrho^{AB}_\varphi\}] \leq F_Q[\varrho^{AB}_\varphi], \tag{3.69}$$

what is consistent with the natural intuition that by ignoring some part of a given system we may only lose information about the estimated parameter.

When investigating the estimation scheme of Fig. 3.6, the monotonicity property explicitly proves an intuitive prediction that decoherence may only worsen the precision achieved. Owing to the commutativity of noise with parameter encoding, the map $\mathcal{D}$ may always be assumed to act at the end of the evolution, so that by monotonicity (3.67) it may only decrease the QFI, as $F_Q\big[\mathcal{D}[\mathcal{U}^{\otimes N}_\varphi[\rho^N_{\mathrm{in}}]]\big] \leq F_Q\big[\mathcal{U}^{\otimes N}_\varphi[\rho^N_{\mathrm{in}}]\big]$.

Convexity in quantum states

Lastly, let us note that the QFI (3.57) is a *convex* quantity with respect to density matrices, so that for a given statistical ensemble, $\{p_i, \varrho_{\varphi,i}\}_i$, of quantum states $\varrho_{\varphi,i}$ with p_i representing their corresponding probabilities, the QFI obeys

$$F_Q\Big[\sum_i p_i\,\varrho_{\varphi,i}\Big] \leq \sum_i p_i\,F_Q[\varrho_{\varphi,i}]. \tag{3.70}$$

Intuitively, one may interpret the l.h.s. of Eq. (3.70) as describing the situation in which we are estimating φ basing on an unknown state randomly chosen from the ensemble, whereas the r.h.s. corresponds to the case in which the state is also drawn from the ensemble but is well known before performing the

---

[30]Notice that by additivity $F_Q\big[\big|\check{\Psi}^{\mathrm{ext}}_\varphi\big\rangle\big] = F_Q\big[\big|\tilde{\Psi}_\varphi\rangle|\xi\rangle\big] = F_Q\big[\big|\tilde{\Psi}_\varphi\big\rangle\big] + F_Q\big[|\xi\rangle\big] = F_Q\big[\big|\tilde{\Psi}_\varphi\big\rangle\big]$, as trivially $F_Q[|\xi\rangle] = 0$ being independent of the estimated parameter.

estimation. Hence, in the first case the *same* estimation scheme must be used for all the states, while in the second case we may utilise *different* schemes depending on the state obtained. Consistently, as in the second scenario we can perform only better and extract more information about the parameter, the r.h.s. of Eq. (3.70) cannot be less than the l.h.s.[31].

On the other hand, in order to prove Eq. (3.70) rigorously, it is enough to consider a binary ensemble with two states $\varrho_{\varphi,1/2} \in \mathcal{H}$ distributed according to probabilities p and $1-\mathsf{p}$ respectively. One may then construct a new state in $\mathcal{H} \oplus \mathcal{H}$ as a direct sum of $\varrho_{\varphi,1/2}$, i.e. $\tilde{\varrho}_\varphi = \mathsf{p}\varrho_{\varphi,1} \oplus (1-\mathsf{p})\varrho_{\varphi,2}$, which by the direct-sum structure does not possess any coherences between the two constituent subspaces. As a result, although the states $\varrho_{\varphi,1/2}$ are still randomly distributed, they may be perfectly distinguished before performing the estimation procedure, so that $F_Q[\tilde{\varrho}_\varphi] = \mathsf{p}F_Q[\varrho_{\varphi,1}] + (1-\mathsf{p})F_Q[\varrho_{\varphi,2}]$[32]. Now, as there exists a CPTP map Λ such that $\Lambda[\mathsf{p}\varrho_{\varphi,1} \oplus (1-\mathsf{p})\varrho_{\varphi,2}] = \mathsf{p}\varrho_{\varphi,1} + (1-\mathsf{p})\varrho_{\varphi,2}$ for any $\varrho_{\varphi,1/2}$—consider a quantum channel, $\mathcal{H} \oplus \mathcal{H} \to \mathcal{H}$, with two Kraus operators that respectively select the subspaces 1/2 of a given state—by monotonicity (3.67) $F_Q[\mathsf{p}\varrho_{\varphi,1} + (1-\mathsf{p})\varrho_{\varphi,2}] \leq F_Q[\tilde{\varrho}_\varphi]$, what completes the proof [Fujiwara, 2001]. ∎

Importantly, the convexity (3.70) of the QFI allows to assure that, when considering the phase estimation scheme of Fig. 3.6, the *inter-particle entanglement* is *necessary* to surpass the SQL-like scaling in N. Assuming as before, when discussing the additivity of the QFI, the most general *separable* output state consisting of N particles: $\rho_\varphi^N = \sum_i p_i \otimes_{n=1}^N \rho_{\varphi,i}^{(n)}$, we may always lower-bound its corresponding QCRB (3.55) for any ν by utilizing the convexity property (3.70) as

$$\Delta^2 \tilde{\varphi}_\nu \geq \frac{1}{\nu\, F_Q\left[\sum_i p_i \otimes_{n=1}^N \rho_{\varphi,i}^{(n)}\right]} \quad \geq \quad \frac{1}{\nu\, \sum_i p_i\, F_Q\left[\otimes_{n=1}^N \rho_{\varphi,i}^{(n)}\right]} = \frac{1}{\nu\, \sum_{n=1}^N \sum_i p_i\, F_Q\left[\rho_{\varphi,i}^{(n)}\right]} \geq \frac{const}{\nu N}, \quad (3.71)$$

where the constant factor is again not scalable with[33] N. As a consequence, the QFI may also serve as a *witness* [Bengtsson and Życzkowski, 2006; Horodecki *et al.*, 2009] of entanglement present in between the particles, as by Eq. (3.71) $F_Q[\rho_\varphi^N] > N$ can occur only for non-separable states ρ_φ^N [Pezzé and Smerzi, 2009], what opens doors for theoretical applications of the QFI (3.57) as a quantity sensing and potentially quantifying the multi-particle entanglement [Hyllus *et al.*, 2012; Tóth, 2012].

Moreover, for the phase estimation scheme of Fig. 3.6, the convexity property explicitly proves that it is optimal to use *pure input states*, as Eq. (3.70) implies that given a general $\rho_{in}^N = \sum_i p_i |\psi_i^N\rangle\langle\psi_i^N|$ and defining $\Lambda_\varphi^N = \mathcal{D} \circ \mathcal{U}_\varphi^{\otimes N} = \mathcal{U}_\varphi^{\otimes N} \circ \mathcal{D}$: $F_Q[\Lambda_\varphi^N[\rho_{in}^N]] = F_Q[\sum_i p_i \Lambda_\varphi^N[|\psi_i^N\rangle]] \leq \sum_i p_i F_Q[\Lambda_\varphi^N[|\psi_i^N\rangle]] \leq \max_i F_Q[\Lambda_\varphi^N[|\psi_i^N\rangle]]$. This agrees with the natural intuition that when estimating a latent parameter encoded by the system evolution (or equivalently by a quantum channel, here Λ_φ^N, as described later in Chap. 4) and inputting a random state belonging to an ensemble of pure states, i.e. a mixed state, we may perform only worse than if we had used a perfectly known optimal state from the ensemble.

3.2.3 Bayesian approach – *global* estimation of a *stochastic* parameter

Surely, at the classical level of the estimation scheme of Fig. 3.6 with the input fixed and a particular measurement scheme chosen, nothing prevents us to: reinterpret the established outcome PDF as the conditional distribution, i.e. $p_\varphi^N(x) \equiv p^N(x|\varphi)$, assume some prior distribution of the parameter, $p(\varphi)$, and also apply the Bayesian techniques described in Sec. 3.1.3. Yet, if we incorporate the quantum

---

[31]For completeness, let us also mention a very recent result proving that for a general state $\varrho_\varphi = \sum_i p_i \varrho_{\varphi,i}$ with $\varrho_{\varphi,i} = |\psi_{\varphi,i}\rangle\langle\psi_{\varphi,i}|$ and the unitary encoding $\varrho_\varphi = U_\varphi \varrho_0 U_\varphi^\dagger$, there always exists a decomposition of $\varrho_0 = \sum_i p_i |\psi_{0,i}\rangle\langle\psi_{0,i}|$, i.e. an ensemble $\{p_i, |\psi_{0,i}\rangle\}$ with non-orthogonal $|\psi_{0,i}\rangle$, for which the Eq. (3.70) is tight [Tóth and Petz, 2013; Yu, 2013].

[32]In general, $\forall_{\varrho,\sigma}$: $F_Q[\varrho \oplus \sigma] = F_Q[\varrho] + F_Q[\sigma]$, what may be most easily proved by utilizing the QFI-definition (3.55).

[33]Being lower-limited by $\left(\max_n \sum_i p_i F_Q\left[\rho_{\varphi,i}^{(n)}\right]\right)^{-1}$.

part of the problem into the optimisation procedure and, in particular, try to minimise the *average cost* (3.36) over the choice of input states and quantum measurements, the task in general turns out to be highly non-trivial. Furthermore, as the Bayesian approach is designed to accurately account for the progressive improvement of the knowledge about the estimated parameter with growth of the sample size (see Sec. 3.1.3.1), in order to correctly apply it to the estimation scheme of Fig. 3.6, we should *update the prior* PDF, $p(\varphi)$, after *each* of ν procedure rounds and repetitively minimise the average cost to determine the optimal form of both the input state and the POVM M_X, which may vary from "shot to shot" for the protocol to be most efficient [Demkowicz-Dobrzański, 2011]. This contrasts the case of Eq. (3.33), which in the classical setting at the level of PDFs assured the progressive updating scenario to be equivalent to a single minimisation of the overall $\overline{\text{MSE}}$ (3.26) evaluated on the whole data.

On the other hand, as the independent character of the sample is *not* a necessary requirement within the global approach, we may also set $\nu = 1$ and consider a *single repetition* of the protocol, in which the measurement capabilities are unrestricted and one must thus optimise over *all* M_X satisfying: $M_x \geq 0$, $\oint \mathrm{d}x\, M_x = \mathbb{I}$; and acting potentially on all N particles. Although (as remarked in Sec. 3.2.1) such a general "single-shot" estimation scheme encompasses all classical scenarios of uncorrelated: particles, noise and measurements; the outcomes in principle may not be assumed to be independently distributed, so that the Bayesian solution cannot be always related to the frequentist results in the asymptotic N limit via Eq. (3.34). Nevertheless, we may greatly benefit from the general form of the POVM, especially in the case of the phase estimation scheme of Fig. 3.6, in which the parameter is encoded unitarily and the estimation problem enjoys the *circular symmetry* described in Sec. 3.1.3.3. As a result, if also a complete *lack of prior knowledge* about the parameter, i.e. a uniform prior distribution $p(\varphi) = 1/(2\pi)$, is assumed, such symmetry is maintained at the level of the average cost (3.36) and the estimation task may always be solved after restricting to *covariant measurements* that explore the circular nature of the parameter, but in principle require correlated operations on all the particles [Holevo, 1982].

Considering a single round of the phase estimation protocol we drop for simplicity the repetition-subscript of the estimator, so that $\tilde{\varphi}(x) \equiv \tilde{\varphi}_{\nu=1}(x)$, and adapt the circularly symmetric average cost (3.41) to the quantum setting as

$$\langle \mathcal{C}_{\text{H}}(\tilde{\varphi}) \rangle = \int \frac{\mathrm{d}\varphi}{2\pi} \oint \mathrm{d}x\, \text{Tr}\{\rho_\varphi^N M_x\}\, C_{\text{H}}(\tilde{\varphi}(x) - \varphi) \tag{3.72}$$

with the outcome probability reading $p^N(x|\varphi) = \text{Tr}\{\rho_\varphi^N M_x\}$ for the output state $\rho_\varphi^N = \mathcal{U}_\varphi^{\otimes N}[\rho_0^N]$, where $\rho_0^N = \mathcal{D}[\rho_{\text{in}}^N]$ as before. As shown in App. D following [Holevo, 1982], any POVM utilised in Eq. (3.72) may always be replaced without affecting $\langle \mathcal{C}_{\text{H}}(\tilde{\varphi}) \rangle$ by a *covariant POVM* of the form

$$M_{\tilde{\varphi}} = U_{\tilde{\varphi}}^{\otimes N}\, \Xi^N\, U_{\tilde{\varphi}}^{\dagger \otimes N}, \tag{3.73}$$

which is parametrised by the estimated values and satisfies $\int \frac{\mathrm{d}\tilde{\varphi}}{2\pi} M_{\tilde{\varphi}} = \mathbb{I}$, $M_{\tilde{\varphi}} \geq 0$ with its *seed element* reading $\Xi^N = \oint \mathrm{d}x\, U_{\tilde{\varphi}(x)}^{\dagger \otimes N} M_x\, U_{\tilde{\varphi}(x)}^{\otimes N}$ for a particular M_X and an estimator $\tilde{\varphi}(x)$. Importantly, as covariant measurements constitute a subclass of *all* legal POVMs, the above fact proves that, in order to establish the *minimal* average cost, we may always make Eq. (3.72) independent of the estimator chosen by rewriting it with use of $M_{\tilde{\varphi}}$ (3.73) as

$$
\begin{aligned}
\langle \mathcal{C}_{\text{H}} \rangle &= \iint \frac{\mathrm{d}\varphi}{2\pi} \frac{\mathrm{d}\tilde{\varphi}}{2\pi}\, \text{Tr}\{\rho_\varphi^N M_{\tilde{\varphi}}\}\, C_{\text{H}}(\tilde{\varphi} - \varphi) \\
&= \text{Tr}\{\langle \rho_\varphi^N \rangle_{C_{\text{H}}}\, \Xi^N\},
\end{aligned}
\tag{3.74}
$$

and minimise Eq. (3.74) over all covariant POVMs, i.e. over all possible seed elements that are *positive semi-definite*: $\Xi^N \geq 0$, and satisfy the *completeness constraint*: $\int \frac{\mathrm{d}\varphi}{2\pi} U_\varphi^{\otimes N} \Xi^N U_\varphi^{\dagger \otimes N} = \mathbb{I}$. The second expression for the average cost $\langle \mathcal{C}_{\text{H}} \rangle$ follows from the invariance of the Haar measure—see App. D—with

respect to which both φ and $\tilde{\varphi}$ are integrated. Crucially, it indicates that the average cost (3.74), due to the covariant symmetry, may be interpreted as an outcome of a quantum measurement represented by the seed element, Ξ^N, evaluated on an effective state, which is the output, ρ_φ^N, averaged over the cost function C_H (3.40): $\langle \rho_\varphi^N \rangle_{C_H} = \int \frac{d\varphi}{2\pi} \rho_\varphi^N 4\sin^2\left(\frac{\varphi}{2}\right)$.

Note that, if a given phase estimation problem allows for a *consistent* estimator—see Eq. (3.2)—to exist, then, as Eq. (3.74) is optimised by construction over all estimators, $\tilde{\varphi}$ is bound to approach φ with N for the optimal Ξ^N. Hence, the *minimal* average cost (3.74) converges asymptotically to the $\overline{\text{MSE}}$, i.e. $\langle C_H \rangle \overset{N\to\infty}{=} \langle \Delta^2 \tilde{\varphi} \rangle$, and the MMSE estimator (3.29) representing the mean of the posterior distribution— $p(\tilde{\varphi}|\varphi)$ now being continuously parametrised by the outputs of the covariant POVM—is always optimal in the $N \to \infty$ limit. However, we may not assume the adequate MMSE estimator to converge with N to the ML estimator and directly relate the minimal $\langle C_H \rangle$ to the CRB (3.55) with $\nu = 1$ via Eq. (3.34), unless we consider a classical scenario with uncorrelated noise for which $\rho_\varphi^N = \rho_\varphi^{\otimes N}$ and establish an equivalently-optimal measurement scheme that ignores correlations in between the particles.

Summarising, the phase estimation problem described in Fig. 3.6 is solved for a *single repetition* within the global approach by minimising Eq. (3.74) over the input states ρ_{in}^N and seed elements Ξ^N. Let us emphasise that the input which corresponds to the minimum is optimal from the Bayesian perspective which assumes *no prior knowledge* about the parameter. Hence, it may differ dramatically from the state maximising the QFI (3.60), which is optimal within the complementary frequentist approach designed to indicate the inputs leading to highest sensitivity to small parameter fluctuations from a *known* value— see Sec. 3.1.2.3. On the other hand, the optimal covariant POVM (3.73) may seem unrealistic being continuously parametrised and thus probably not realisable in a real-life experiment. Whence, in order for the solution to be of practical importance, one should attempt to find another POVM with a *finite* number of elements, i.e. $\sum_x M_x = \mathbb{I}$, which also attains the minimal cost (3.74). Such a construction, however, has been shown to be typically feasible for finite dimensional systems with the discrete outcomes of the POVM being then directly associated with particular values of the estimated parameter, what still circumvents the problem of the estimator optimisation [Derka *et al.*, 1998].

3.2.4 EXAMPLE: **Mach-Zehnder interferometry at the Heisenberg Limit**

We revisit the *Mach-Zehnder interferometer* example of Sec. 3.1.4, in order to generalise the previous discussion and fully account for the quantum aspect of the setup. Regarding the interferometer of Fig. 3.1 with a definite number of photons, N, as a special case of the general *phase estimation scheme* of Fig. 3.6, we are able to directly apply the frequentist and Bayesian frameworks developed in Secs. 3.2.2 and 3.2.3. Yet, as shown in Fig. 3.7, in order to exactly match their corresponding notation, one must identify the *input state* as a two-mode, N-photon state (see Eq. (2.6)) of light *inside* the interferometer, i.e.

$$|\psi_{in}^N\rangle = \sum_{n=0}^{N} \alpha_n |n\rangle_a |N-n\rangle_b = \sum_{n=0}^{N} \alpha_n |n, N-n\rangle, \tag{3.75}$$

where a and b label the interferometer arms—modes—and the general coefficients satisfy $\sum_{n=0}^{N} |\alpha_n|^2 = 1$. Although the state (3.75) is written in the occupation number representation, i.e. the second quantisation, in order to study its metrological properties, one should treat its constituent photons as separate particles, what—see Sec. 2.1.4—formally corresponds to returning to the first-quantisation picture and treating photons as distinguishable particles in a permutation invariant state [Demkowicz-Dobrzański *et al.*, 2015]. For example, the *classical strategy* of Sec. 3.1.4, in which N uncorrelated photons, i.e. a Fock $|N\rangle$ state of Note 2.3, are impinged on a balanced beam-splitter, results in binomially distributed α_n-coefficients
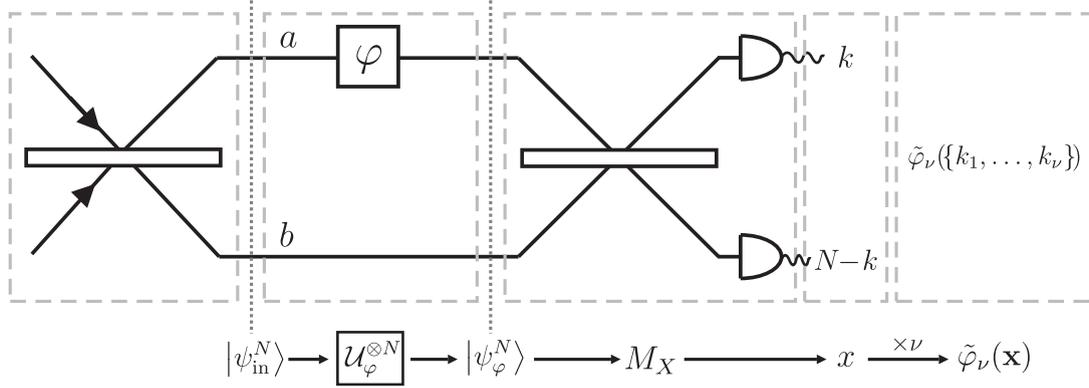
FIGURE 3.7: **Mach-Zehnder interferometer** (Fig. 3.1) **as an instance of a** *noiseless* **phase estimation scheme** (Fig. 3.6). The input state $|\psi_{\text{in}}^N\rangle$ corresponds to the two-mode, N-photon state of light after the first beam-splitter, whereas the output state $|\psi_\varphi^N\rangle$ differs only by the estimated phase, φ, accumulated due to the path difference between the arms a and b. The photon-counting measurement constitutes an example of a general POVM M_X, which outcome x corresponds to the photon number recorded in one of the ports: k (other is then fixed to $N-k$). If necessary, the procedure is repeated ν times to build an independently distributed sample $\mathbf{x} = \{x_1 \equiv k_1, \ldots, x_\nu \equiv k_\nu\}$, from which the parameter value is finally inferred via the estimator $\tilde{\varphi}_\nu$.

in the "modal picture":

$$|\psi_{\text{in}}^N\rangle_{\text{cl}} = \sum_{n=0}^{N} \sqrt{\frac{1}{2^N}\binom{N}{n}}\, |n, N-n\rangle = \left[\frac{1}{\sqrt{2}}\left(|a\rangle + |b\rangle\right)\right]^{\otimes N}. \tag{3.76}$$

Whereas in the "particle picture", as shown in the second expression above, the state inside the interferometer corresponds then to a *product state* of N photons each being in an equally weighted superposition of states $|a\rangle$ and $|b\rangle$, which form a qubit-like (spin-1/2–like) basis of each photon and represent it travelling in either of the arms/modes. Crucially, by considering now a general $|\psi_{\text{in}}^N\rangle$ and varying α_n in the quantum setting, we can introduce the *inter-particle entanglement* that may allow to surpass the $1/N$ SQL-like scaling of precision accounted in Eqs. (3.44) and (3.51), which assessed the performance of the product input state (3.76) within local and global approaches respectively for the classical scenario of Fig. 3.1 with a photon-counting measurement.

As in this section we aim to determine the ultimate quantum-enhancement of precision theoretically attainable in a Mach-Zehnder interferometer, similarly to the classical case, we assume a noiseless scenario in which the photons during the evolution only accumulate the estimated phase that—see Fig. 3.7—is encoded unitarily onto each of them via $U_\varphi^{\otimes N} = \mathrm{e}^{-\mathrm{i}\hat{H}_N\varphi}$. $\hat{H}_N$ is the overall N-photon Hamiltonian that is decomposable in the "particle picture" as $\hat{H}_N = \sum_{n=1}^{N} \hat{H}^{(n)}$ with $\hat{H}^{(n)} = \frac{1}{2}(|a\rangle\langle a| - |b\rangle\langle b|)$ acting on the "n-th photon". Equivalently, $\hat{H}_N$ may be represented in the "modal picture" as $\hat{H}_N = \frac{1}{2}(\hat{n}_a - \hat{n}_b)$ with $\hat{n}_a = \hat{a}^\dagger\hat{a}$ and $\hat{n}_b = \hat{b}^\dagger\hat{b}$ being the photon-number operators of arms a and b respectively. As shown in Fig. 3.7, the *output state* should be associated with $|\psi_\varphi^N\rangle = \mathrm{e}^{-\mathrm{i}\hat{H}_N\varphi}|\psi_{\text{in}}^N\rangle$, as the second beam-splitter can always be incorporated into the measurement part of the general phase estimation scheme of Fig. 3.6.

3.2.4.1   Frequentist approach

As the input state is assumed to be pure (what must be optimal due the convexity property of the QFI described in Sec. 3.2.2.3), the scenario of Fig. 3.7 is noiseless, and the estimated phase is unitarily encoded onto photons, Eq. (3.61) for the QFI directly applies, so that the QFI is proportional to the variance of the N-photon Hamiltonian $\hat{H}_N$ and explicitly reads:

$$F_{\mathrm{Q}}\big[|\psi_\varphi^N\rangle\big] = 4\,\Delta^2 \hat{H}_N = 4\left[\sum_{n=0}^{N} |\alpha_n|^2\, n^2 - \left(\sum_{n=0}^{N} |\alpha_n|^2\, n\right)^2\right]. \qquad (3.77)$$

Classical input states
Let us briefly mention first the performance of inputs consisting of uncorrelated photons employed in the *classical strategy* of Sec. 3.1.4, i.e. the states (3.76), $|\psi_{\mathrm{in}}^N\rangle_{\mathrm{cl}}$, which in the "modal picture" lead to binomially distributed coefficients α_n. Evaluating Eq. (3.77), $F_{\mathrm{Q}}\big[|\psi_\varphi^N\rangle_{\mathrm{cl}}\big] = N$ and one arrives at the QCRB $1/N$ coinciding the classical CRB (3.44) obtained for the photon-counting measurement strategy of Fig. 3.1. Hence, the above fact *proves* that indeed such a measurement scheme is optimal for classical inputs (3.76) from the local perspective.

Optimal input states – NOON states
On one hand, one may perform explicitly the maximisation of Eq. (3.77) over the coefficients α_n, e.g. with use of the method of Lagrange multipliers, in order to show that it is optimal to choose only non-zero $|\alpha_0| = |\alpha_N| = 1/\sqrt{2}$. However, without any calculation, it is straightforward to identify the maximal/minimal eigenvalues of the Hamiltonian $\hat{H}_N$ in the "modal picture" as $\pm N/2$ representing the maximal difference in the photon-number in between the interferometer arms and the corresponding eigenvectors $|N,0\rangle$ and $|0,N\rangle$. Thus, according to the discussion of Sec. 3.2.2.1, the *optimal input state* must be the equally weighted superposition of these eigenvectors[34] [Giovannetti *et al.*, 2006], i.e.

$$|\psi_{\mathrm{in}}^N\rangle_{\mathrm{NOON}} = \frac{1}{\sqrt{2}}\big(|N,0\rangle + |0,N\rangle\big) = \frac{1}{\sqrt{2}}\left(|a\rangle^{\otimes N} + |b\rangle^{\otimes N}\right), \qquad (3.78)$$

what is consistent with the explicit maximisation of Eq. (3.77). The state (3.78) is commonly referred to as the *NOON state* due to its form in the modal representation [Bollinger *et al.*, 1996; Lee *et al.*, 2002], what may be slightly misleading. That is why, we also write it explicitly in Eq. (3.78) in the "particle picture", in order to emphasise that it corresponds to a *maximally entangled state*, i.e. the *Greenberger-Horne-Zeilinger* (GHZ) state [Greenberger *et al.*, 1989], of the particles. Importantly, it is the *inter-particle entanglement* [Demkowicz-Dobrzański *et al.*, 2015], which assures that if one photon travels in a particular arm of the interferometer then so must the others, that leads to an N-fold winding of the off-diagonal terms, e.g. $\mathrm{e}^{iN\varphi}|0,N\rangle\langle N,0|$, so that the effective density matrix of the NOON state resembles the one of a single-photon state but with an N-times greater phase resolution. As a consequence, the sensitivity to φ-variations scales quadratically with the particle number N, what is explicitly manifested by the *maximal* QFI (3.77) reading

$$F_{\mathrm{Q}}\big[|\psi_{\mathrm{in}}^N\rangle_{\mathrm{NOON}}\big] = N^2 \qquad \Longrightarrow \qquad \Delta^2\tilde{\varphi}_\nu \geq \frac{1}{\nu\,N^2} \qquad (3.79)$$

and yielding the ultimate $1/N^2$ scaling of the QCRB (3.55) dictated by the maximal inter-photon correlations allowed by quantum mechanics.

---

[34]For simplicity, we set $\phi = 0$ in $\frac{1}{\sqrt{2}}\big(|N,0\rangle + \mathrm{e}^{i\phi}|0,N\rangle\big)$ and ignore the arbitrary relative phase in between the eigenvectors.

The QCRB (3.79) defines the so-called *Heisenberg Limit* (HL) [Holland and Burnett, 1993] imposed on the precision scaling with the particle number (here *definite* photon number): $1/N^2$. However, when interferometric schemes of Fig. 3.7 with *indefinite* photon number are considered, one must be more rigorous when quantifying the resources [Zwierz *et al.*, 2010, 2012], as naively replacing N in Eq. (3.79) with the average photon number $\bar{N}$ may lead to incorrect conclusions of "surpassing the HL", which must then be properly redefined [Berry *et al.*, 2012; Demkowicz-Dobrzański *et al.*, 2015; Giovannetti and Maccone, 2012; Hall *et al.*, 2012; Hofmann, 2009; Hyllus *et al.*, 2010b]. Furthermore, one should bear in mind that Eq. (3.79) represents a frequentist bound and the HL is guaranteed to be attainable only in the local estimation regime, which is warranted only when $\nu \to \infty$. In fact, for moderate ν, one may construct bounds on the estimator MSE (3.3), $\Delta^2 \tilde{\varphi}_\nu$, of a different type[35] that turn out to be indeed tighter than the QCRB (3.79) [Giovannetti *et al.*, 2012]. On the other hand, the necessity of procedure repetitions in NOON-based scenarios becomes evident when analysing the optimal measurement schemes discussed below that lead to precision saturating the QCRB (3.79).

Optimal measurements

Firstly, let us consider the quantum measurement assured to be optimal by construction, i.e. the projective measurement in the eigenbasis of the SLD (3.56) [Braunstein and Caves, 1994; Nagaoka, 1989]. Utilizing the expression below Eq. (3.65) we derive the SLD for the NOON-based strategy with the output $|\psi_\varphi^N\rangle_{\text{NOON}} = e^{-i\hat{H}_N\varphi}|\psi_{\text{in}}^N\rangle_{\text{NOON}}$ as $L_{\text{S}}\big[|\psi_{\varphi_0}^N\rangle_{\text{NOON}}\big] = iN e^{iN\varphi_0}|0,N\rangle\langle N,0| + h.c.$ for the parameter true value φ_0. In general, the eigenvectors of the SLD read $|E_\pm(\varphi_0)\rangle = \frac{1}{\sqrt{2}}e^{-i\hat{H}_N\varphi_0}(|N,0\rangle \pm i|0,N\rangle)$, but as the QFI (3.79) is parameter independent we may without loss of generality set $\varphi_0 = \pi/(2N)$ to simplify the form of the measurement elements obtained, which then project the output onto states: $|E_\pm\rangle = \frac{1}{\sqrt{2}}(|N,0\rangle \pm |0,N\rangle)$. Notice that these correspond to NOON-like states, which are maximally entangled in between the particles, making such a scheme extremely hard to implement in an experiment. Furthermore, for a single repetition of the protocol ($\nu = 1$), the above quantum measurement yields—independently of the photon number N—only two outcomes that occur with probabilities $p_{\varphi,\pm} = \big|\langle E_\pm|\psi_\varphi^N\rangle_{\text{NOON}}\big|^2$ reading $\cos^2(N\varphi/2)$ and $\sin^2(N\varphi/2)$ respectively. Such behaviour perfectly agrees with the previously mentioned intuition that we may treat the NOON state as a *single statistical object* or in other words a single "entanglement-enhanced photon" which is N-times more sensitive to phase variations. Thus, it should not be surprising that we must require many repetitions ν, in order to gather enough data and attain the QCRB (3.79) and hence the HL. In fact, we may interpret the above scenario as a classical Mach-Zehnder interferometry scheme of Sec. 3.1.4 with ν such "entanglement-enhanced photons" [Higgins *et al.*, 2009, 2007], where each of them independently leads to detection probabilities at the output oscillating N-times quicker with φ, so that now $\mathsf{p} = \cos^2(N\varphi/2)$ in Fig. 3.1. Whence, basing on the binomial PDF (3.43), we may directly write the distribution of registering r "+" outcomes (and hence $\nu - r$ "-" ones) after carrying out ν repetitions of the protocol, as

$$p_\varphi^\nu(r) = \binom{\nu}{r}\left(\cos^2\frac{N\varphi}{2}\right)^r\left(\sin^2\frac{N\varphi}{2}\right)^{\nu-r}, \tag{3.80}$$

which leads to $F_{\text{cl}}\big[p_\varphi^\nu\big] = \nu N^2$ and the CRB indeed coinciding with the QCRB (3.79). Moreover, we may straightforwardly establish the locally efficient and the ML estimators by modifying the corresponding classical-scenario expressions (3.46) and (3.47), as it is enough to substitute $k \to r$ and $N \to \nu$ (the repetitions now stand for the number of uncorrelated photons) and rescale all parameter true values and estimators via $\varphi \to N\varphi$ to account for the N-fold resolution improvement. For example, the ML estimator (3.47) now reads $\tilde{\varphi}_\nu^{\text{ML}} = \frac{2}{N}\text{arccot}\big(\sqrt{r/(\nu-r)}\big)$ and applies to the interval $[0, \pi/N]$. Notice that previously in the classical scenario we suffered from the sign ambiguity of φ, which forced us to consider only to a

---

[35] In particular, the Ziv-Zakai bounds [Tsang, 2012].

π-wide region of parameter values. Now, such an unambiguous sector is further shrank to π/N, because only then φ may be conclusively inferred from $p_{\varphi,\pm}$. Such requirement is equivalent to the possession of prior knowledge about the parameter with MSE $\approx \pi^2/N^2$, which is assumed to be available for free due to locality of the approach. Importantly, it is this *prior knowledge*—which quantification is beyond the capabilities of the frequentist framework—that leads to HL-like scaling in the above scenario!

On the other hand, it has been generally proved in [Giovannetti *et al.*, 2006] that in the *noiseless* version of the phase estimation scenario of Fig. 3.6, in order to achieve the ultimate precision in the local regime of $\nu \to \infty$ with optimal input states employed, it is sufficient to consider only uncorrelated measurements acting separately on the constituent particles, i.e. $M_X = \bigotimes_{n=1}^{N} M_{X_n}^{(n)}$. In case of the quantum-enhanced Mach-Zehnder interferometer of Fig. 3.7 this have been shown in various ways [Bollinger *et al.*, 1996; Gerry and Mimih, 2010; Kok *et al.*, 2002; Lee *et al.*, 2002], but an instructive example of such an efficient uncorrelated scheme is again the photon-counting measurement previously employed in the classical scenario of Fig. 3.1. In fact, it allows to locally attain the corresponding QCRB not only for the optimal NOON-based strategy but also for any path(mode)-symmetric $|\psi_{in}^N\rangle$ [Hofmann, 2009], which possess all coefficients satisfying $|\alpha_n| = |\alpha_{N-n}|$. After propagating the output state through the second beam-splitter[36], $|\psi_\varphi^N\rangle_{NOON} \to |\tilde{\psi}_\varphi^N\rangle_{NOON}$, one may determine the quantum equivalent of Eq. (3.43), i.e. the probability of detecting k and $N-k$ photons at the interferometer output ports in Fig. 3.7, as

$$p_\varphi^N(k) = \left|\langle k, N-k|\tilde{\psi}_\varphi^N\rangle_{NOON}\right|^2 = \frac{1}{2^N}\binom{N}{k}\left[1 + (-1)^k \cos(N\varphi)\right], \qquad (3.81)$$

which indeed yields classical FI, $F_{cl}[p_\varphi^N] = N^2$, saturating the QCRB (3.79) for $\nu = 1$. However, apart from the combinatorial factor which does not carry any information about the parameter, the distribution (3.81) varies only depending on whether k is even or odd. Hence, complete information about φ that may be retrieved from $p_\varphi^N(k)$ resides in the *parity* of the photon-number registered [Anisimov *et al.*, 2010; Chiruvelli and Lee, 2011; Gerry and Mimih, 2010; Seshadreesan *et al.*, 2013]. Furthermore, evaluating thus $p_{\varphi,+} = \sum_{k=0,2,..}^N p_\varphi^N(k) = \cos^2(N\varphi/2)$ and $p_{\varphi,-} = \sum_{k=1,3,..}^N p_\varphi^N(k) = \sin^2(N\varphi/2)$ we reproduce the outcomes of the SLD-based measurement strategy considered above. Therefore, although we have shown that a photon-counting measurement indeed suffices, all the discussions and results from the previous paragraph apply. In particular, we suffer again from the necessity of locality (assured only in the asymptotic limit of many repetitions) and, in particular, the notion of the HL, $1/N^2$, is again contained within the prior knowledge which analysis lies beyond the scope of the frequentist approach.

3.2.4.2 Bayesian approach

In order to study from the Bayesian perspective the performance of the quantum-enhanced Mach-Zehnder interferometer of Fig. 3.7 for a single repetition ($\nu = 1$), we minimise the adequate average cost (3.74), $\langle \mathcal{C}_H \rangle$, over all pure input states (3.75) inside the interferometer, $|\psi_{in}^N\rangle$, and all positive semi-definite seed elements, $\Xi^N \geq 0$, from which the effective measurement schemes based on covariant POVMs (3.73) may be constructed [Berry and Wiseman, 2000; Hradil *et al.*, 1996]. As $\int \frac{d\varphi}{2\pi} 4\sin^2(\varphi/2)e^{-i\varphi(n-m)} = 2\delta_{n,m} - (\delta_{n,m-1} + \delta_{n,m+1})$, the effective matrix $\langle \rho_\varphi^N \rangle_{C_H}$, which describes the output state $\rho_\varphi^N = |\psi_\varphi^N\rangle\langle\psi_\varphi^N|$ averaged over the cost function (3.40), is tridiagonal, i.e. possesses non-zero entries on its main and

---

[36]What corresponds to the transformations on the NOON state: $|N,0\rangle \to \sum_{n=0}^N \sqrt{\frac{1}{2^N}\binom{N}{n}}|n, N-n\rangle$ and $|0,N\rangle \to \sum_{n=0}^N \sqrt{\frac{1}{2^N}\binom{N}{n}}(-1)^n|n, N-n\rangle$.

closest to the main diagonals. Hence, we may most generally write the average cost (3.74) as

$$\langle \mathcal{C}_\mathrm{H} \rangle \; = \; 2\left(\sum_{n=1}^{N} |\alpha_n|^2 \, \Xi_{n,n}^N \right) - 2\,\mathrm{Re}\left\{ \sum_{n=1}^{N} \alpha_n^* \alpha_{n-1} \, \Xi_{n,n-1}^N \right\} \; = \; 2\left(1 - \mathrm{Re}\left\{ \sum_{n=1}^{N} \alpha_n^* \alpha_{n-1} \, \Xi_{n,n-1}^N \right\} \right), \qquad (3.82)$$

where we have acknowledged the fact that the completeness condition $\int \frac{\mathrm{d}\varphi}{2\pi} U_\varphi^{\otimes N} \Xi^N U_\varphi^{\dagger \otimes N} = \mathbb{I}$ forces all the diagonal entries of any seed element to be equal to one, i.e. $\Xi_{n,n}^N = 1$ for all n.

Optimal covariant measurements

As within the Bayesian approach the precision limits are not measurement-strategy independent, we firstly proceed with minimisation of $\langle \mathcal{C}_\mathrm{H} \rangle$ over the choice of seed elements. Because by replacing the input-state coefficients and the seed-element entries by their absolute values one may only decrease the average cost (3.82), it is optimal to choose Ξ^N such that for all n: $\alpha_n^* \alpha_{n-1} \Xi_{n,n-1}^N = |\alpha_n||\alpha_{n-1}||\Xi_{n,n-1}^N|$. Moreover, due to the positive semi-definiteness of the seed element, $\Xi^N \geq 0$, we may upper-bound then the absolute value of its entries as $|\Xi_{n,m}^N| \leq \sqrt{\Xi_{n,n}^N \Xi_{m,m}^N} = 1$. Hence, the subtracted term in Eq. (3.82) can at most read $\sum_{n=1}^{N} |\alpha_n||\alpha_{n-1}|$, what may be always achieved by setting $\Xi_{n,m}^N = \mathrm{e}^{\mathrm{i}(\phi_n - \phi_m)}$, where $\phi_n = \arg(\alpha_n)$. This corresponds to the choice of the *optimal seed element* which may be expressed as: $\Xi_\mathrm{opt}^N = |e^N\rangle\langle e^N|$ with $|e^N\rangle = \sum_{n=0}^{N} \mathrm{e}^{\mathrm{i}\phi_n} |n, N-n\rangle$, what proves that it is positive semi-definite as required [Chiribella *et al.*, 2005; Holevo, 1982]. As a consequence, we obtain the expression for the average cost (3.82) that is optimised over the choice of *all* measurement strategies:

$$\langle \mathcal{C}_\mathrm{H} \rangle \; = \; 2\left(1 - \sum_{n=1}^{N} |\alpha_n||\alpha_{n-1}| \right). \qquad (3.83)$$

and the form of the *optimal covariant POVM* (3.73) reading: $M_{\tilde\varphi} = U_{\tilde\varphi}^{\otimes N} |e^N\rangle\langle e^N| U_{\tilde\varphi}^{\dagger \otimes N}$. Note that the elements of $M_{\tilde\varphi}$ are *not* separable with respect to particles[37], so that they define to a *collective* measurement performed on all the particles.

Classical and locally optimal input states

Firstly, let us consider the performance of inputs employed in the *classical strategy* of Sec. 3.1.4, i.e. the states $|\psi_\mathrm{in}^N\rangle_\mathrm{cl}$ defined in Eq. (3.76) that consist of uncorrelated photons. Substituting their binomial distribution of coefficients α_n into Eq. (3.83), we obtain the corresponding measurement-optimised average cost:

$$\langle \mathcal{C}_\mathrm{H} \rangle_\mathrm{cl} = 2\left(1 - \frac{1}{2^N} \sum_{n=1}^{N} \sqrt{\binom{N}{n}\binom{N}{n-1}} \right) \overset{N\to\infty}{=} \frac{1}{N}, \qquad (3.84)$$

which exhibits an SQL-like scaling for $N \to \infty$, previously observed within the local approach. Moreover, as Eq. (3.84) asymptotically coincides with the QCRB established for the classical states (3.76) in the first paragraph of the previous section, one may *incorrectly* assume that Eq. (3.34) connecting two complementary approaches applies, and thus there must also exist an uncorrelated measurement acting separately on the particles, which asymptotically achieves the average cost (3.84). Notice that it *cannot* be the photon-counting measurement analysed in the classical scenario of Sec. 3.1.4, as, although its corresponding Bayesian cost (3.51) also asymptotically achieves the $1/N$ CRB (3.44) and hence the QCRB, it requires the parameter to be *a priori* restricted to the unambiguous $[0, \pi]$ region. In contrast, the Bayesian strategy described by Eq. (3.84) assumes a *complete* ignorance of the parameter and, as one may numerically verify that the ratio of Eqs. (3.51) and (3.84) is in fact less than one, i.e. $\forall_N$:

---

[37]Notice that $|e^N\rangle$ is a *non-normalised* vector that is *entangled* in between the particles, as for it to be a product state of particles its coefficient would need to be binomially distributed as in Eq. (3.76).

$\langle\mathcal{C}_{\mathrm{H}}(\tilde{\varphi}_N^{\mathrm{H}})\rangle/\langle\mathcal{C}_{\mathrm{H}}\rangle_{\mathrm{cl}}\le 1$, the photon-counting scheme with $\varphi\in[0,\pi]$ performs actually better for finite N than the optimal Bayesian strategy for $\varphi\in[-\pi,\pi]$! The measurement scheme that achieves the average cost (3.84) for the uniform prior distribution, $p(\varphi)=1/2\pi$, is the *adaptive strategy* [Wiseman *et al.*, 2009; Wiseman and Killip, 1997, 1998], in which the measurement of a subsequent photon is adjusted depending on the results previously gathered, what progressively narrows down the region of confidence even when starting from a complete flat prior distribution. Note that for such a measurement strategy to be applicable, the photons must be distinguishable, e.g. arriving at the detector in consecutive time bins, so that they may be targeted individually in a sequence. Although Eq. (3.83) has been derived assuming their indistinguishability, $\langle\mathcal{C}_{\mathrm{H}}\rangle$ still quantifies the minimal cost for distinguishable-particles input states. We return to this issue in Chap. 5, where we consider the lossy Mach-Zehnder interferometer scenario and prove (see App. J) that the average cost cannot be decreased by benefiting from the distinguishability of the particles employed.

For completeness, before optimising Eq. (3.83) over all inputs, let us convince the reader that NOON states, suffering from the π/N ambiguity explained in Sec. 3.2.4.1, are useless when no prior knowledge about parameter is available. Evaluating the measurement-optimised average cost (3.83) for $|\psi_{\mathrm{in}}^N\rangle_{\mathrm{NOON}}$, $\forall_{N>1}:\langle\mathcal{C}_{\mathrm{H}}\rangle_{\mathrm{NOON}}=2$ and the NOON-based estimation scheme does not provide any information about the parameter *regardless of* N, even though general measurements are taken into account. Crucially, being designed for the local regime, in the global setting NOON states are outperformed by the classical inputs (3.76), which (as shown above) still lead to the vanishing cost with N at the $1/N$ SQL-like rate.

Optimal input states – Berry-Wiseman states

In order to minimise the average cost (3.83) over all input states, we rewrite it in a more appealing matrix-form after denoting by $\boldsymbol{\alpha}=\{\alpha_0,\dots,\alpha_N\}$ the vector containing the coefficients α_n, which optimally are taken to be real:

$$\langle\mathcal{C}_{\mathrm{H}}\rangle = 2 - \boldsymbol{\alpha}^T \begin{pmatrix} 0 & 1 & 0 & \dots \\ 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \boldsymbol{\alpha} = 2 - \boldsymbol{\alpha}^T \bar{\mathbf{A}}\,\boldsymbol{\alpha}. \tag{3.85}$$

Thus, it becomes apparent that the *minimal* average cost equals $2-\bar{\lambda}_{\max}$, where $\bar{\lambda}_{\max}$ is the maximal eigenvalue of the matrix $\bar{\mathbf{A}}$, whereas the optimal input state coefficients are determined by the eigenvector corresponding to $\bar{\lambda}_{\max}$. The above problem has been solved independently in [Berry and Wiseman, 2000] and [Bužek *et al.*, 1999], but, as the work of Berry and Wiseman [2000] considered explicitly the interferometric scenario here described, we will refer to the optimal input state as the *Berry-Wiseman* (BW) state[38]:

$$|\psi_{\mathrm{in}}^N\rangle_{\mathrm{BW}} = \sum_{n=0}^{N} \sqrt{\frac{2}{N+2}}\,\sin\left(\frac{n+1}{N+2}\pi\right)|n,N-n\rangle, \tag{3.86}$$

which corresponds to $\bar{\lambda}_{\max}=2\cos[\pi/(N+2)]$ and thus yields the *minimal average cost* now optimised over both input states and measurements:

$$\langle\mathcal{C}_{\mathrm{H}}\rangle_{\mathrm{BW}} = 2\left[1-\cos\left(\frac{\pi}{N+2}\right)\right] \stackrel{N\to\infty}{=} \frac{\pi^2}{N^2}. \tag{3.87}$$

---

[38]Yet, both Eqs. (3.86) and (3.87) have been established for the first time before the Bayesian analysis has been applied to Mach-Zehnder interferometry by Summy and Pegg [1990], who have approached the phase estimation problem by utilising the Hermitian phase operator method of [Pegg and Barnett, 1988].

Importantly, Eq. (3.87) proves that, despite the extra π^2 factor, the $1/N^2$ HL-like scaling is also exhibited when analysing the noiseless phase estimation scheme of Fig. 3.6 from the global perspective. This is assured, because for $N \to \infty$ the average cost (3.87) may be interpreted as the $\overline{\text{MSE}}$, i.e. $\langle \mathcal{C}_H \rangle_{\text{BW}} \overset{N \to \infty}{=} \langle \Delta^2 \tilde{\varphi}^{\text{MMSE}} \rangle_{\text{BW}} = \int \mathrm{d}\varphi \, p(\varphi) \, \Delta^2 \tilde{\varphi}^{\text{MMSE}} \big|_{\varphi} = \pi^2/N^2$ with $p(\varphi) = 1/2\pi$, even though $\langle \mathcal{C}_H \rangle_{\text{BW}}$ does not asymptotically coincide with the QCRB (3.79) (Eq. (3.34) does not hold due to the presence of inter-particle correlations). Furthermore, Eq. (3.87) has been recently generalised in [Jarzyna and Demkowicz-Dobrzański, 2014] to asymptotically apply even when any regular prior distribution, $p(\varphi)$, is considered. Hence, for this to be true, we conjecture that the "local MSE" of the MMSE estimator (3.29) reads $\Delta^2 \tilde{\varphi}^{\text{MMSE}} \big|_{\varphi_0} = \pi^2/N^2$ for any φ_0, what indicates that $\tilde{\varphi}^{\text{MMSE}}$ does *not* converge asymptotically to the ML estimator (3.19). In particular, its variance is always widened by the π^2 factor due to the global character of the approach, even if a very narrow but regular prior PDF is assumed[39].

Lastly, let us comment on the applicability of the Bayesian approach, when one considers the phase estimation problem of the noiseless Mach-Zehnder interferometer, but accounts for the possibility of *many repetitions*, i.e. $\nu > 1$ in the general protocol of Fig. 3.6. In such a case, as mentioned in the first paragraph of Sec. 3.2.3, the strategy that utilises BW-states (3.86) and covariant POVMs (3.73) is optimal only in the first round, for which the above analysis applies. After each run of the protocol, the effective prior distribution must be updated, and the derivation of the optimal input states and measurements minimising the average cost (3.72) must be consecutively repeated. In particular, as the prior is no longer flat after some information about the parameter is gathered during the first round, the circular symmetry of the problem is lost and one may not restrict any more only to covariant POVMs. However, a general procedure has been recently proposed in [Demkowicz-Dobrzański, 2011] that allows for a numerical minimisation of the general average cost (3.72) regardless of the prior distribution assumed. Although the method of Demkowicz-Dobrzański [2011] has been demonstrated to be generally efficient only for relatively low numbers of particles, it was sufficient to indicate for a given moderate N a smooth transition of the optimal inputs from BW-states to NOON-states, as the prior becomes gradually narrowed with increase of the knowledge about the parameter[39]. Hence, we conjecture that such a transition of the optimal input states—which crucially may now be adaptively varied between the procedure repetitions—should also occur with an increase of the repetition number ν. Yet, when focusing only on any of the single protocol repetitions, we suspect the precision to still be limited by the "global HL" (3.87): π^2/N^2, as such bound is independent of the prior considered [Jarzyna and Demkowicz-Dobrzański, 2014] (which just smoothly narrows down with an increase of ν).

---

[39]Let us remind the reader that eventually in the limit of an infinitely narrow prior, i.e. for a Dirac delta distribution that is no longer *regular*, the Bayesian approach ignores the optimisation, as it indicates then that it is optimal just to output the a priori known parameter value, so that $\langle \mathcal{C} \rangle = \langle \Delta^2 \tilde{\varphi} \rangle = 0$, and do not perform any estimation (see Note 3.3).

Chapter 4

Local estimation in the presence of uncorrelated noise

4.1 N-parallel–channels estimation scheme

In this chapter we generalise the phase estimation scheme of Fig. 3.6 with *uncorrelated noise*, so that the quantum estimation problem may be solved without imposing any constraints on the form of the single-particle evolution. We depict the estimation scenario as shown in Fig. 4.1, where now the parameter is encoded independently onto each particle by some general *φ-parametrised quantum channel* Λ_φ. Naturally, such a description encapsulates the scheme of Fig. 3.6 with uncorrelated noise and unitary encoding, for which then $\Lambda_\varphi = \mathcal{D} \circ \mathcal{U}_\varphi = \mathcal{U}_\varphi \circ \mathcal{D}$. As in this chapter we analyse such an *N-parallel–channels estimation scheme* within the *frequentist approach*, we may restrict (by convexity of the QFI discussed in Sec. 3.2.2.3) ourselves already to *pure input states*, $|\psi_{\rm in}^N\rangle$, that can perform only better than mixed inputs. Hence, the output state most generally reads: $\rho_\varphi^N = \Lambda_\varphi^{\otimes N} [|\psi_{\rm in}^N\rangle]$, and a quantum measurement M_X follows
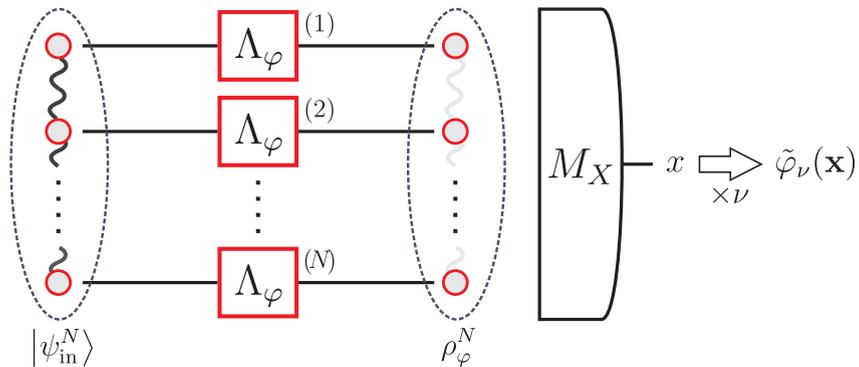
FIGURE 4.1: *N-parallel–channels estimation scheme* as a generalisation of the phase estimation scheme of Fig. 3.6 that includes the *uncorrelated noise*. A pure input state $|\psi_{\rm in}^N\rangle$ consists of N particles, each of which evolves according to a general quantum map Λ_φ that may in principle account for *any* parameter encoding and uncorrelated noise. As before, a quantum measurement on the output state $\rho_\varphi^N = \Lambda_\varphi^{\otimes N} [|\psi_{\rm in}^N\rangle]$ is represented by a POVM M_X and the estimator $\tilde{\varphi}_\nu$ is constructed on the sampled data after ν repetitions of the protocol.

as in the scheme of Fig. 3.6. As before, after repeating the protocol ν times, an unbiased estimator, $\tilde{\varphi}_\nu$, is built on the data gathered and its MSE (3.3) (or equivalently its variance (3.1)) is lower-limited by the QCRB (3.55) which is guaranteed to be attainable for $\nu \to \infty$.

4.2 Local estimation of a *single* quantum channel

Before studying the attainable precision from the local perspective for the general N-particle scheme of Fig. 4.1, we analyse the local properties of a *single*, φ-parametrised quantum channel Λ_φ. In particular, we study the possibilities of generalising the notion of the QFI (3.57) to quantum maps—precisely, families of *Completely Positive Trace Preserving* (CPTP) maps[1] parametrised by φ, $\{\Lambda_\varphi\}_\varphi$—so that the attainable precision of estimation can be quantified by inspecting the form of a given channel, without need to explicitly study its output states with help of the methods described previously in Sec. 3.2.2.1.

4.2.1 Channel QFI

Given a single particle in the scheme of Fig. 4.1, or more generally a quantum system in an initial pure state $|\psi_{\rm in}\rangle$, we identify its final (output) state after the evolution as $\rho_\varphi = \Lambda_\varphi[|\psi_{\rm in}\rangle]$ with the parameter encoded by the action of the CPTP map Λ_φ. The ultimate φ-estimation precision is then dictated by the QCRB (3.55) with the QFI reading: $F_{\rm Q}[\Lambda_\varphi[|\psi_{\rm in}\rangle]]$, and varies depending on the input state chosen. Hence, as shown in Fig. 4.2(**a**), we define the *channel QFI* as the maximal QFI after performing the input optimisation, so that it has a concrete operational and application-like interpretation:

$$\mathcal{F}[\Lambda_\varphi] = \max_{\psi_{\rm in}} F_{\rm Q}[\Lambda_\varphi[|\psi_{\rm in}\rangle]] \, . \tag{4.1}$$

For instance, in case of the noiseless phase estimation scenario previously describing the Mach-Zehnder interferometer of Fig. 3.7, $\Lambda_\varphi = \mathcal{U}_\varphi$ and the output is a pure state: $e^{-i\hat{H}\varphi}|\psi_{\rm in}\rangle$, with $\hat{H}$ being the Hamiltonian generating the phase variation. Hence, as the QFI equals then the variance of the Hamiltonian according to Eq. (3.61), the definition (4.1) just corresponds to

$$\mathcal{F}[\mathcal{U}_\varphi] = \max_{\psi_{\rm in}} F_{\rm Q}\left[e^{-i\hat{H}\varphi}|\psi_{\rm in}\rangle\right] = 4 \max_{\psi_{\rm in}} \Delta^2 \hat{H} = (\mu_{\max} - \mu_{\min})^2 \, , \tag{4.2}$$

where $\mu_{\max/\min}$ are respectively the maximal/minimal eigenvalues of $\hat{H}$ and the maximum occurs for $|\psi_{\rm in}\rangle$ being an equally weighted superposition of the eigenvectors corresponding to $\mu_{\max/\min}$.

Note 4.1: Time-energy uncertainty relation from the channel QFI perspective.

An interesting result is obtained, if one applies Eq. (4.2) in case of the natural latent parameter of the evolution, i.e. the *time* t, for which $U_t = e^{-i\hat{H}t}$ and $\mathcal{F}[\mathcal{U}_t] = 4\Delta^2\hat{H} = (E_{\max} - E_{\min})^2$, where $E_{\max/\min}$ represent the maximal/minimal energies in the spectrum: $\hat{H}|E\rangle = E|E\rangle$. Note that the QCRB-equivalent (3.62) constitutes then exactly the *time-energy uncertainty relation* discussed in Note 2.5:

$$\Delta^2 \hat{H} \, \Delta^2 \tilde{t} \geq \frac{1}{4} \quad \implies \quad \Delta \tilde{t} \geq \frac{1}{E_{\max} - E_{\min}} \, , \tag{4.3}$$

stating that the maximal variance of the Hamiltonian—specified by the energy difference $E_{\max} - E_{\min}$—defines the ultimate resolution with which the duration estimator $\tilde{t}$ may be resolved [Aharonov *et al.*, 2002]. Moreover, as Eq. (4.3) is similarly to Eq. (3.62) optimised over all potential measurement and inference strategies, the time-energy uncertainty relation (4.3) is more general than the Mandelstam-Tamm inequality (2.23), which (as shown in Note 2.5) is derived basing on a particular observable measurement that determines the ultimate sensitivity of the estimator $\tilde{t}$

---

[1]See Secs. 2.2.1 and 2.4 for introduction to quantum channels and discussion of their geometric properties respectively.
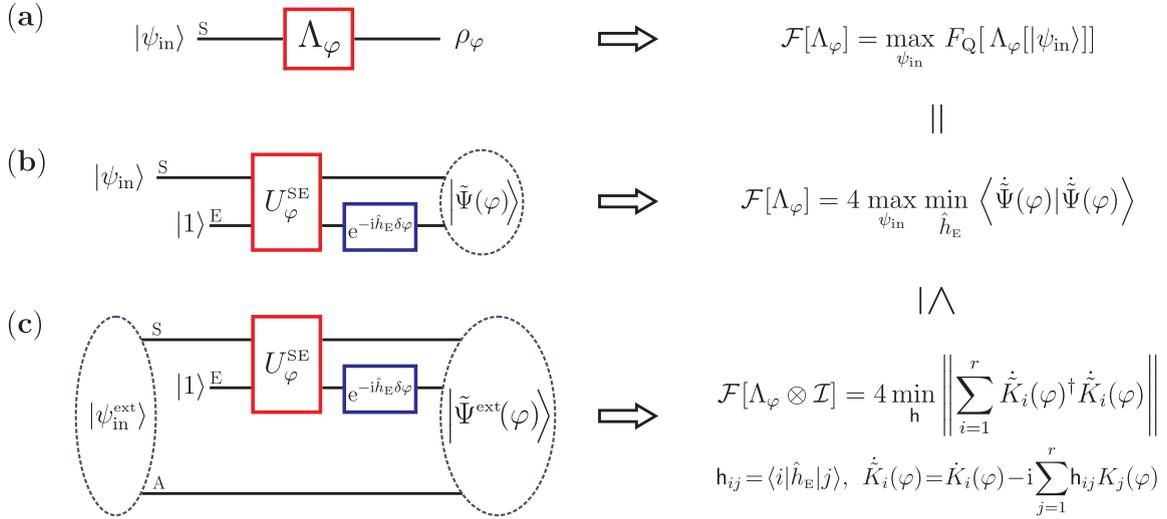
FIGURE 4.2: **Channel QFI evaluated basing on the output state purification**.
(**a**) *Channel QFI* defined as the QFI of the output state maximised over all pure input states.
(**b**) *Channel QFI* equivalently obtained by considering the output state purification after a *local, fictitious, parameter-dependent rotation of the environment*, $u_\varphi^{\rm E} = {\rm e}^{-{\rm i}\hat{h}_{\rm E}\delta\varphi}$ with $\delta\varphi = \varphi - \varphi_0$ for a given φ_0, which hinders as much as possible information about the parameter.
(**c**) *Extended-channel QFI* being independent of the maximisation over the input states. In the purification picture, the environment rotation $u_\varphi^{\rm E}$ can be just understood as a "shift" in the derivatives of the *Kraus operators*, $K_i(\varphi)$, describing the action of the channel.

to the variations of the actual elapsed time t [Braunstein *et al.*, 1996].

Purification-based definition of the channel QFI

In order to express the *channel QFI* (4.1) with help of the *purification-based definition* (3.64) introduced previously in Sec. 3.2.2.2 for quantum states, we utilise the Stinespring dilation theorem (see Thm 2.2.1 and Fig. 2.1) and rewrite a general channel Λ_φ as a unitary map $U_\varphi^{\rm SE}$ acting on the system, S, and the environment, E, disregarded after the evolution. As a result, the combined state $|\Psi(\varphi)\rangle = U_\varphi^{\rm SE}|\psi_{\rm in}\rangle_{\rm S}|1\rangle_{\rm E}$ naturally constitutes the output purification such that $\Lambda_\varphi[|\psi_{\rm in}\rangle] = {\rm Tr}_{\rm E}\{|\Psi(\varphi)\rangle\langle\Psi(\varphi)|\}$, where $|1\rangle$ is an arbitrary fixed state (i.e. $|\xi\rangle$ in Fig. 2.1) chosen to be the first vector in the basis $\{|i\rangle\}_{i=1}^r$ of the environment Hilbert space $\mathcal{H}_{\rm E}^r$. As depicted in Fig. 4.2(**b**), by specifying the dimension r of $\mathcal{H}_{\rm E}^r$ to be equal to the rank of Λ_φ, we may directly follow the recipe of Sec. 3.2.2.2 and construct all the *locally relevant* output purifications, $|\tilde{\Psi}(\varphi)\rangle$, for the parameter true value φ_0 by applying fictitious unitary rotations $u_\varphi^{\rm E} = {\rm e}^{-{\rm i}\hat{h}_{\rm E}(\varphi-\varphi_0)}$ to the environment that are generated by any Hermitian $\hat{h}_{\rm E}$, so that $|\tilde{\Psi}(\varphi)\rangle = \tilde{U}_\varphi^{\rm SE}|\psi_{\rm in}\rangle_{\rm S}|1\rangle_{\rm E}$ with $\tilde{U}_\varphi^{\rm SE} = u_\varphi^{\rm E} U_\varphi^{\rm SE}$. Hence, as this corresponds again to a local shift of the purification first derivative, we may express the channel QFI (4.1) as the "channel version" of Eq. (3.64):

$$\mathcal{F}[\Lambda_\varphi] = 4 \max_{\psi_{\rm in}} \min_{\hat{h}_{\rm E}} \langle \dot{\tilde{\Psi}}(\varphi)|\dot{\tilde{\Psi}}(\varphi)\rangle, \qquad (4.4)$$

where for a given φ_0 we must search only through purifications such that $|\tilde{\Psi}(\varphi_0)\rangle = |\Psi(\varphi_0)\rangle$ and $|\dot{\tilde{\Psi}}(\varphi_0)\rangle = (\dot{U}_{\varphi_0}^{\rm SE} - {\rm i}\hat{h}_{\rm E}U_{\varphi_0}^{\rm SE})|\psi_{\rm in}\rangle_{\rm S}|1\rangle_{\rm E}$. On the other hand, writing the action of the channel Λ_φ in its Kraus representation form introduced in Sec. 2.2.1, i.e. $\Lambda_\varphi[|\psi_{\rm in}\rangle] = \sum_{i=1}^r K_i(\varphi)|\psi_{\rm in}\rangle\langle\psi_{\rm in}|K_i(\varphi)^\dagger$, we can

identify the Kraus operators corresponding to $|\tilde{\Psi}(\varphi)\rangle$ as

$$\tilde{K}_i(\varphi) = \langle i| \tilde{U}^{\mathrm{SE}}_\varphi |1\rangle = \sum_{j=1}^r \mathsf{u}(\varphi)_{ij}\, K_j(\varphi)\,, \tag{4.5}$$

where $\mathsf{u}(\varphi)_{ij} = \langle i| u^{\mathrm{E}}_\varphi |j\rangle$ and $K_j(\varphi) = \langle j| U^{\mathrm{SE}}_\varphi |1\rangle$ are the Kraus operators of the original purification $|\Psi(\varphi)\rangle$. Thus, we may further rewrite Eq. (4.4) as:

$$\mathcal{F}[\Lambda_\varphi] = 4 \max_{\psi_{\mathrm{in}}} \min_{\mathsf{h}} \langle\psi_{\mathrm{in}}| \sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \dot{\tilde{K}}_i(\varphi) |\psi_{\mathrm{in}}\rangle, \tag{4.6}$$

where now a particular choice of $\hat{h}_{\mathrm{E}}$ corresponds to a local shift of first derivatives of the Kraus operators. Whence, at a given φ_0 we must search through Kraus representations generated from the starting one via: $\tilde{K}_i(\varphi_0) = K_i(\varphi_0)$ and $\dot{\tilde{K}}_i(\varphi_0) = \dot{K}_i(\varphi_0) - \mathrm{i} \sum_{j=1}^r \mathsf{h}_{ij} K_j(\varphi_0)$ with $\mathsf{h}_{ij} = \langle i| \hat{h}_{\mathrm{E}} |j\rangle$, what corresponds to the minimisation in Eq. (4.6) over all Hermitian matrices h of size[2] $r \times r$.

Importantly, by inspecting Fig. 4.2(**b**) and Eq. (4.4)/(4.6), it becomes evident that the optimal purification/Kraus representation—which by the previous argumentation of Sec. 3.2.2.2 would be identified for the output state ρ_φ as the one satisfying either $\left|\dot{\tilde{\Psi}}^{\mathrm{opt}}_\varphi\right\rangle = \frac{1}{2} L_{\mathrm{S}}[\rho_\varphi] \otimes \mathbb{I}^{\mathrm{E}} \left|\tilde{\Psi}^{\mathrm{opt}}_\varphi\right\rangle$ or $\dot{\tilde{K}}^{\mathrm{opt}}_i(\varphi) |\psi_{\mathrm{in}}\rangle = \frac{1}{2} L_{\mathrm{S}}[\rho_\varphi] \tilde{K}^{\mathrm{opt}}_i(\varphi) |\psi_{\mathrm{in}}\rangle$ respectively—can now be neatly interpreted as the one for which an artificial environment is chosen that hinders as much as possible information about the estimated parameter due to its extra local rotation u^{E}_φ. As a result, the environment part of the optimal purification, $\left|\tilde{\Psi}^{\mathrm{opt}}_\varphi\right\rangle$, does not carry any locally extractable information about φ, so that consistently with the purification-based definition (3.63): $F_{\mathrm{Q}}\left[\left|\tilde{\Psi}^{\mathrm{opt}}_\varphi\right\rangle\right] = F_{\mathrm{Q}}[\rho_\varphi]$.

4.2.2 Extended-channel QFI

From the geometrical perspective, similarly to the case of quantum states for which the QFI describes the "speed" of variations in φ (see Eq. (3.59)), one may like the channel QFI (4.1) also to quantify the local statistical distance but this time between the maps Λ_φ and $\Lambda_{\varphi+\delta\varphi}$ for small $\delta\varphi$. However, at the level of quantum channels, one should bear in mind that for any variation of φ to be noticeable, there must exist input states which lead to a measurable change of the channel output that is at best in some "orthogonal direction". As a consequence, the quantity $\min_{\mathsf{h}}\{\ldots\}$ in Eq. (4.6) non-trivially varies with the input $|\psi_{\mathrm{in}}\rangle$, as the minimum occurs for the optimal Kraus operators $\left\{K^{\mathrm{opt}}_i(\varphi)\right\}_{i=1}^r$ defined via condition $\dot{K}^{\mathrm{opt}}_i(\varphi) |\psi_{\mathrm{in}}\rangle = \frac{1}{2} L_{\mathrm{S}}[\Lambda_\varphi [|\psi_{\mathrm{in}}\rangle]] K^{\mathrm{opt}}_i(\varphi) |\psi_{\mathrm{in}}\rangle$ that explicitly depends on $|\psi_{\mathrm{in}}\rangle$. As a result, one cannot in principle express the channel QFI (4.1) using only the form of Λ_φ and circumvent the problem of input state maximisation, i.e. $\max_{|\psi_{\mathrm{in}}\rangle}$ in Eq. (4.6), that is difficult in general, as for a given input chosen one must always establish the optimal purification/Kraus representation and thus *cannot* naively interchange the order of max and min in Eq. (4.4)/(4.6) [Fujiwara and Imai, 2008].

Yet, as depicted in Fig. 4.2(**c**), one may construct a natural upper bound on the channel QFI (4.1) by *extending* the input space, $\mathcal{H}_{\mathrm{S}}$, by an equally-large ancillary space, $\mathcal{H}_{\mathrm{A}}$, which is unaffected by the map but measured along with the channel output. In this way, by employing extended input states entangled between these two spaces, $|\psi^{\mathrm{ext}}_{\mathrm{in}}\rangle \in \mathcal{H}_{\mathrm{S}} \otimes \mathcal{H}_{\mathrm{A}}$, one may extract more information about φ by inspecting also the ancilla, A, which—despite not being affected by Λ_φ—due to entanglement with S improves the capabilities of quantum measurements performed on the whole, extended output state $\rho^{\mathrm{ext}}_\varphi = \Lambda_\varphi \otimes \mathcal{I}[|\psi^{\mathrm{ext}}_{\mathrm{in}}\rangle]$

---

[2]Following Sec. 2.2.1, we denote the matrices $\mathsf{u}(\varphi)$ and h with a different fount to indicate that these are *not* operators, and should be just treated as matrices with complex entries specifying linear transformations of Kraus operators.

[Sedlák and Ziman, 2009; Ziman, 2008]. Such a *channel extension* defines then the *extended-channel QFI*:

$$\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] = \max_{\psi_{\rm in}^{\rm ext}} F_{\rm Q}[\Lambda_\varphi \otimes \mathcal{I}[|\psi_{\rm in}^{\rm ext}\rangle]]] \tag{4.7}$$

that is naturally greater than the channel QFI (4.1) and $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] \geq \mathcal{F}[\Lambda_\varphi]$. Note that Eqs. (4.1) and (4.7) coincide when the extension turns out *not* to be beneficial, what is manifested by the optimal extended input state, $|\psi_{\rm in}^{\rm ext}\rangle$, being separable, i.e. $|\psi_{\rm in}^{\rm ext}\rangle = |\psi_{\rm in}\rangle_{\rm S}|\xi\rangle_{\rm A} \Rightarrow \rho_\varphi^{\rm ext} = \rho_\varphi \otimes |\xi\rangle_{\rm A}\langle\xi|$ for any $|\xi\rangle$, and one may freely trace out the ancillary subspace, $\mathcal{H}_{\rm A}$, without affecting the QFI of the extended output state.

Purification-based definition of the extended-channel QFI

The analogue of Eq. (4.6) that specifies the *purification-based definition of the extended-channel QFI* by utilising equivalent Kraus representations (4.5) of the channel reads [Fujiwara and Imai, 2008]:

$$\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] = 4 \max_{\rho_{\rm in}^{\rm S}} \min_{\mathsf{h}} {\rm Tr}_{\rm S}\left\{\rho_{\rm in}^{\rm S} \sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \dot{\tilde{K}}_i(\varphi)\right\} = 4 \min_{\mathsf{h}} \left\|\sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \dot{\tilde{K}}_i(\varphi)\right\|, \tag{4.8}$$

where $\|\ldots\|$ represents the operator norm. The first expression above is obtained by tracing over the ancillary space $\mathcal{H}_{\rm A}$, what leads to the maximisation over all mixed states $\rho_{\rm in}^{\rm S} = {\rm Tr}_{\rm A}\{|\psi_{\rm in}^{\rm ext}\rangle\langle\psi_{\rm in}^{\rm ext}|\}$. Thus, Eq. (4.8) is exactly the purification-based expression for the (unextended) channel QFI (4.6) with pure states $|\psi_{\rm in}\rangle$ replaced by mixed ones $\rho_{\rm in}^{\rm S}$, what crucially allows to interchange the order of max and min above and obtain the second expression [Fujiwara and Imai, 2008]. However, one shall *not* be mistaken that the extended-channel QFI (4.7) can be interpreted as the generalisation of the (unextended) channel QFI (4.1) to mixed-state inputs! Although one may replace without loss of generality pure input states with mixed ones in the primary channel QFI definition (4.1) due to convexity of the QFI, the purification-based definitions (4.4) and (4.6) after such an interchange become invalid. It is so, because the purifications employed in Eqs. (4.4) and (4.6) should then also account for the fact of *purifying the mixed input state* and otherwise lead to an overestimate of the actual output state QFI. As unintentionally shown by Eq. (4.8), the quantity obtained then is the extended-channel QFI (4.7), which consistently constitutes an upper bound: $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] \geq \mathcal{F}[\Lambda_\varphi]$. On the other hand, by inspecting the optimal $\rho_{\rm in}^{\rm S}$ in Eq. (4.8) one may verify if the extension leads to a precision improvement. If there does *not* exist an optimal $\rho_{\rm in}^{\rm S}$ which is *mixed*[3], the optimal extended input $|\psi_{\rm in}^{\rm ext}\rangle$ must be separable, what (as discussed in the previous paragraph) assures $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] = \mathcal{F}[\Lambda_\varphi]$.

Importantly, the expression (4.8) for the extended-channel QFI involves *only* the Kraus representation optimisation and may be reformulated as a *semi-definite program* (SDP) that crucially is *always* efficiently evaluable numerically. In App. I, we demonstrate how to construct the relevant SDP for a more general task of upper-bounding the QFI of N-parallel channels $\Lambda_\varphi^{\otimes N}$, i.e. for the scheme of Fig. 4.1 explicitly analysed from the local perspective later in Sec. 4.3. Yet, because Eq. (4.8) is just a special case of such a more general procedure with $N = 1$, its SDP-reformulation directly follows (see App. I for details).

4.2.3   RLD-based upper bound on the extended-channel QFI

Nevertheless, as the evaluation of the extended-channel QFI via Eq. (4.8) still involves minimisation, which due to many free parameters (h is only constrained to be a Hermitian matrix) is in general not easily solvable analytically, one may want to seek for further upper bounds on $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$, and

---

[3]What occurs only if the maximal eigenvalue of the operator $\sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \dot{\tilde{K}}_i(\varphi)$ in Eq. (4.8) is *non-degenerate* forcing the optimal $\rho_{\rm in}^{\rm S}$ to be pure.

hence also on $\mathcal{F}[\Lambda_\varphi]$, that do not involve any optimisation at all. One possibility for such a construction is to relax the QCRB itself by replacing the SLD in Eq. (3.55) with other logarithmic derivatives of the output state ρ_φ, which are non-Hermitian but still satisfy $\dot{\rho}_\varphi = \frac{1}{2}\left(\rho_\varphi L[\rho_\varphi] + L[\rho_\varphi]^\dagger \rho_\varphi\right)$. As proved in [Hayashi, 2005; Holevo, 1982], for any such $L[\rho_\varphi]$ an upper limit on the QFI (3.57) is obtained: $F_Q[\rho_\varphi] \leq \mathrm{Tr}\{\rho_\varphi L[\rho_\varphi]L[\rho_\varphi]^\dagger\}$, that consistently is guaranteed to be tight for the SLD, i.e. $L[\rho_\varphi] = L[\rho_\varphi]^\dagger = L_S[\rho_\varphi]$ [Nagaoka, 2005]. A commonly utilised example of $L[\rho_\varphi]$ is the *Right Logarithmic Derivative* (RLD): $L_R[\rho_\varphi] = \rho_\varphi^{-1}\dot{\rho}_\varphi$, which exists *if and only if* $\dot{\rho}_\varphi$ *is contained within the support of* ρ_φ, but simplifies often the calculations, as its corresponding upper bound on the QFI, $F_Q[\rho_\varphi] \leq \mathrm{Tr}\{\rho_\varphi^{-1}\dot{\rho}_\varphi{}^2\}$, requires inversion of the output state ρ_φ and not its full eigendecomposition[4].

In [Hayashi, 2011] the applicability of such an RLD-based bound has been addressed in the context of quantum channels. In particular, by defining the Choi-Jamiołkowski (CJ) matrix[5] representing the map Λ_φ, i.e. $\Omega_{\Lambda_\varphi} = \Lambda_\varphi \otimes \mathcal{I}\left[|\mathbb{I}\rangle\right]$ with $|\mathbb{I}\rangle = \sum_{i=1}^{\dim \mathcal{H}_S} |i\rangle_S \otimes |i\rangle_A$, it has been proved that the extended-channel QFI (4.7) can be further upper-limited after replacing the SLD with the RLD via:

$$\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] \quad \leq \quad \mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}] = \left\| \mathrm{Tr}_A\left\{\dot{\Omega}_{\Lambda_\varphi} \Omega_{\Lambda_\varphi}^{-1} \dot{\Omega}_{\Lambda_\varphi}\right\}\right\|, \tag{4.9}$$

where $\|\dots\|$ is again the operator norm and $\Omega_{\Lambda_\varphi}^{-1}$ is the inverse of Ω_{Λ_φ}. Crucially, in contrast to Eqs. (4.1) and (4.7), the above *RLD-based upper bound on the extended-channel QFI* is determined *solely* by the form of Λ_φ (its CJ representation)—and does not require any optimisation neither over the input states accepted by the map nor over its Kraus representations. However, the condition for existence of the RLD defined on quantum states has a direct generalisation to the case of quantum channels, as the bound (4.9) is non-divergent and thus applicable for a given map Λ_φ specified by its CJ matrix Ω_{Λ_φ} *if and only if* $\dot{\Omega}_{\Lambda_\varphi}$ *is contained within the support of* Ω_{Λ_φ}. In App. E, we explicitly prove this requirement stemming from the work of Hayashi [2011] and stress that this is *exactly* the criterion for a given CPTP map to be φ-non-extremal (see Crit. 2.4.4 and App. B). Thus, *the RLD-based bound (4.9) applies to and only to φ-non-extremal channels*, what importantly provides a clear geometric explanation for what kinds of quantum maps, and with what parametrisations, is the RLD-based approach valid.

One may wonder whether the bound (4.9) could be straightforwardly improved by considering the QFI calculated w.r.t. the CJ matrix of the quantum channel considered. However, let us explicitly note that $F_Q[\Omega_{\Lambda_\varphi}] = F_Q[\Lambda_\varphi \otimes \mathcal{I}[|\mathbb{I}\rangle]]$ and thus it *lower*-limits the channel QFI: $F_Q[\Omega_{\Lambda_\varphi}] \leq \mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$, which is maximised over all input states. Hence, $F_Q[\Omega_{\Lambda_\varphi}]$ possesses only an operational meaning when the above lower bound is tight, what occurs *only if* the maximally entangled states, $|\mathbb{I}\rangle$, turn out to be optimal. This, however, happens very rarely especially as Λ_φ is assumed to be noisy and thus *not* unitary.

Last but not least, the bound (4.9) has been proved in [Hayashi, 2011] to be *additive* on channels, so that for any two, φ-non-extremal maps $\Lambda_\varphi^{(1)}$ and $\Lambda_\varphi^{(2)}$:

$$\mathcal{F}^{\mathrm{RLD}}\left[\left(\Lambda_\varphi^{(1)}\otimes\mathcal{I}\right)\otimes\left(\Lambda_\varphi^{(2)}\otimes\mathcal{I}\right)\right] = \mathcal{F}^{\mathrm{RLD}}\left[\Lambda_\varphi^{(1)}\otimes\mathcal{I}\right] + \mathcal{F}^{\mathrm{RLD}}\left[\Lambda_\varphi^{(2)}\otimes\mathcal{I}\right]$$
$$\implies \quad \mathcal{F}\left[(\Lambda_\varphi\otimes\mathcal{I})^{\otimes N}\right] \leq \mathcal{F}^{\mathrm{RLD}}\left[(\Lambda_\varphi\otimes\mathcal{I})^{\otimes N}\right] = N\,\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi\otimes\mathcal{I}]. \tag{4.10}$$

As remarked in the second expression, the RLD-based bound thus constrains not only the QFI of a single extended channel, but also restricts the QFI of N extended channels used in *parallel* to scale at most linearly with N. Crucially, as the extension can only improve the precision, Eq. (4.10) is also a valid upper-bound on the QFI of N parallel uses of an unextended channel, i.e. on the QFI of the output

---

[4]For completeness, let us note as aside that when considering *multi-parameter* estimation schemes, in which the SLD is no more the unique logarithmic derivative defining the multi-parameter QCRB, the RLD may sometimes lead to tighter bounds on the overall achievable precision of simultaneous estimation of multiple parameters [Fujiwara, 1994; Genoni *et al.*, 2012; Hayashi, 2005].

[5]See Sec. 2.4.1 for the description of Choi-Jamiołkowski representation of a quantum channel.

Noise model:	Dephasing	Depolarization	Loss	Spontaneous emission
$\mathcal{F}[\Lambda_\varphi]$ (4.1):	η^2	η^2	η	η
$\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$ (4.7):	η^2	$\frac{2\eta^2}{1+\eta}$	η	$\frac{4\eta}{\left(1+\sqrt{\eta}\right)^2}$
$\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$ (4.9):	$\frac{\eta^2}{1-\eta^2}$	$\frac{2\eta^2(1+\eta)}{(1-\eta)(1+3\eta)}$	n.a.	n.a.

TABLE 4.1: **Channel QFIs**, $\mathcal{F}[\Lambda_\varphi]$, **extended-channel QFIs**, $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$, and **RLD bounds**, $\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$, for the noisy-phase–estimation channels introduced in Sec. 2.2.3 and depicted in Fig. 2.2. RLD bounds cannot be constructed for the *loss* and *spontaneous emission* noise-types, as these correspond to φ-extremal channels (see Sec. 2.4.3).   [n.a.—not applicable]

state in the scheme of Fig. 4.1: $F_\mathrm{Q}\big[\rho_\varphi^N\big] \le N\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$. Strikingly, this proves that the estimation precision attained in the scheme of Fig. 4.1, when considering *any single-particle channel Λ_φ that is φ-non-extremal*, must *at most* asymptotically follow the SQL-like scaling. We give an alternative, but maybe more intuitive, explanation to such a conclusion in Sec. 4.3.1.1, where we show that any locally φ-non-extremal channel can in fact be *classically simulated*, what indeed assures the asymptotic scaling to follow $const/N$. Nevertheless, as $\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$ provides a quantitative measure upper-limiting the maximum achievable quantum enhancement of precision, we utilise it explicitly in Sec. 4.3.1, where we term it for short as the *RLD bound* and the above RLD-based precision-bounding procedure as the *RLD method*, which we then compare with other approaches also allowing to derive asymptotic SQL-like bounds on attainable precision.

4.2.4   EXAMPLE: **Noisy-phase–estimation channels**

In order to apply the notions of *channel QFI* (4.1), *extended-channel QFI* (4.7) and the *RLD bound* (4.9) in a concrete setting, we consider the channels introduced in Sec. 2.2.3 and previously depicted in Fig. 2.2, which model the evolution of a *qubit* with the parameter being encoded as the angle, φ, of rotation around the z axis and the decoherence corresponding to one of the noise-types: *dephasing, depolarisation, loss* and *spontaneous emission*; each of strength η. Notice that these constitute an example of the phase estimation scheme of Fig. 3.6 with $N = 1$ and $\mathcal{D}$ representing one of the above-listed noise models. For each of the channels, we determine the relevant quantities and present them in Tab. 4.1 in an increasing order, as accordingly $\mathcal{F}[\Lambda_\varphi] \le \mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] \le \mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$.

Due to the low dimension of the system, we are able to explicitly calculate the *channel QFI* (4.1) for all of the cases and confirm that, as intuitively expected from Fig. 2.2, it is always optimal to prepare the qubit in any state lying on the *equator* of the Bloch ball, as during the action of any of the noise-types considered it still remains represented by the furthest point from the z axis and thus most sensitive to the parameter variations (rotations around the z axis).

In case of the *extended-channel QFI* (4.7)—due to the presence of the ancilla—the output generally corresponds to a mixed state of two qubits, for which the analytic computation of the QFI (3.57) with a generic input is not straightforward any more. That is why, we utilise Eq. (4.8) and explicitly perform the purification-minimisation for each of the channels of Fig. 2.2, what may be always simplified by decreasing the number of free parameters after inspecting the numerical form of the optimal-purification generator h, which accuracy may be quantified and assured due to the SDP reformulation of Eq. (4.8) presented in

App. I. We list in App. F the analytic forms of the optimal generators h that determine the adequate shift of the first derivatives of Kraus operators, when starting from the canonical Kraus representations stated in Tab. 2.1 for each of the channels considered. Importantly, the results of Tab. 4.1 justify that *extension enhances* the precision only for *depolarisation* and *spontaneous emission* channels, for which the optimal inputs correspond to maximally entangled (Bell-like) states: $\frac{1}{\sqrt{2}}\left(|0\rangle_{\mathrm{S}}|\psi\rangle_{\mathrm{A}} + e^{i\phi}|1\rangle_{\mathrm{S}}|\psi_\perp\rangle_{\mathrm{A}}\right)$. These may be intuitively interpreted to consist of the system qubit prepared again on equator for highest parameter sensitivity and an entangled to it ancillary state with $\langle\psi|\psi_\perp\rangle = 0$, which leads to improved precision of the measurements performed on the overall extended output state[6]. Let us also note that in the absence of noise: $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]^{\eta \to 1} = \mathcal{F}[\Lambda_\varphi]$, as the extension may not be beneficial for a unitary channel [D'Ariano *et al.*, 2001].

Lastly, we compute the corresponding *RLD bounds* (4.9) for the *dephasing* and *depolarisation* noise models, which are the only ones—as shown in Sec. 2.4.3—that lead to $\varphi-non\text{-}extremal$ channels. As $\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$ constitutes also an SQL-like bound on the asymptotic scaling (see Eq. (4.10)), RLD bounds in Tab. 4.1 correctly diverge when $\eta = 1$ for the noiseless estimation scenario, in which the HL must be attainable. However, as a result, the RLD bounds, when interpreted as upper limits on the extended-channel QFI, become useless in the regime of $\eta \to 1$, in which $(\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}] / \mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]) \to \infty$.

4.3 Local estimation of N quantum channels in *parallel*

We now consider explicitly the N-parallel–channels estimation scheme of Fig. 4.1, for which the N-particle output state of the system reads $\rho_\varphi^N = \Lambda_\varphi^{\otimes N}\left[|\psi_{\mathrm{in}}^N\rangle\right]$. Thus, analogously to the single channel measures, we define the *N-channel QFI* as:

$$\mathcal{F}\left[\Lambda_\varphi^{\otimes N}\right] = \max_{\psi_{\mathrm{in}}^N} F_{\mathrm{Q}}\left[\Lambda_\varphi^{\otimes N}\left[|\psi_{\mathrm{in}}^N\rangle\right]\right], \tag{4.11}$$

which linear or quadratic dependence on N dictates respectively the SQL- or HL-like scaling of precision. In case of a classical strategy, for which a further constraint on Eq. (4.11) must be imposed restricting the inputs to product states: $|\psi_{\mathrm{in}}^N\rangle = |\psi_{\mathrm{in}}\rangle^{\otimes N}$, consistently the N-channel QFI becomes N times the single channel QFI (4.1), i.e. $\mathcal{F}\left[\Lambda_\varphi^{\otimes N}\right]\big|_{\mathrm{cl}} = N\,\mathcal{F}[\Lambda_\varphi]$.

4.3.1 SQL-like bounds on the asymptotic precision

As we would like to investigate various methods that allow to prove and quantify the asymptotic SQL-like precision scaling emergent in the scheme of Fig. 4.1 due to the impact of uncorrelated noise, we define the *asymptotic channel QFI* as:

$$\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi] = \lim_{N \to \infty} \frac{\mathcal{F}\left[\Lambda_\varphi^{\otimes N}\right]}{N}, \tag{4.12}$$

which is always bounded from above, unless a given channel Λ_φ allows for an asymptotic super-classical precision scaling. In general, $\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi] \geq \mathcal{F}[\Lambda_\varphi]$ with equality indicating the asymptotic optimality of classical estimation scenarios and no room for any quantum enhancement of precision.

---

[6]The ambiguity of choosing any state of the ancilla may be neatly explained realising the freedom of local unitary rotations that may always be performed on it at the measurement stage.
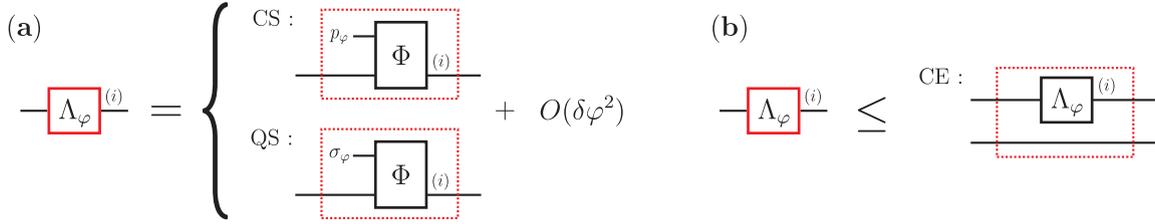
FIGURE 4.3: **SQL-bounding methods on the asymptotic precision scaling** that stem only from the form of a *single* channel, Λ_φ, representing the single-particle evolution in the scheme of Fig. 4.1. (**a**): In the *CS* and *QS methods*, the channel must be *locally* equivalent to a parameter-independent map Φ that is also fed either a classical diagonal state p_φ in CS, or a general, quantum state σ_φ in QS. (**b**): Nevertheless, the tightest bound that is applicable to the widest class of quantum channels is obtained via the *CE method*, in which Λ_φ is replaced by its extension $\Lambda_\varphi \otimes \mathcal{I}$, what trivially can only improve the estimation capabilities.

On the other hand, we may thus quantify with help of Eq. (4.12) the *maximal quantum enhancement of precision*, $\chi[\Lambda_\varphi]$, as the ratio of the asymptotic estimation errors between the classical and optimal-quantum strategies dictated by their corresponding QCRBs (3.55), i.e.

$$\chi[\Lambda_\varphi] = \lim_{N\to\infty} \sqrt{\frac{\Delta^2\tilde{\varphi}_\nu|_{\mathrm{cl}}}{\Delta^2\tilde{\varphi}_\nu|_{\mathrm{Q}}}} = \sqrt{\frac{\nu\,\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi]}{\nu\,\mathcal{F}[\Lambda_\varphi]}} \geq 1, \tag{4.13}$$

where we have left on purpose the repetition number in the last expression to stress that, although Eq. (4.13) is ν-independent, it is guaranteed to be saturable only in the $\nu\to\infty$ limit due to the locality of the frequentist approach. Eq. (4.13) has rather only a formal meaning, as it involves the computation of $\mathcal{F}[\Lambda_\varphi^{\otimes N}]$ for arbitrary large N, what is generally infeasible due to the complexity of the QFI (3.57) rising exponentially with N, not to mention the impossibility of performing the maximisation over all input states in Eq. (4.11).

In what follows we present methods that allow to upper-bound the asymptotic channel QFI (4.12), and hence the maximal quantum precision enhancement (4.13), basing *purely* on the form of a *single* channel Λ_φ—without need of neither considering its tensor-product structure nor performing any optimisation over the input states. Example of such a procedure is the already-mentioned *RLD method* that, due to the additivity property of the RLD bound (4.9) introduced in Sec. 4.2.3, leads to $\mathcal{F}[\Lambda_\varphi^{\otimes N}] \leq N\,\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$ for any φ-non-extremal channel.

In general, we can write

$$\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi] \leq \mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}, \qquad \chi[\Lambda_\varphi] \leq \sqrt{\frac{\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}}{\mathcal{F}[\Lambda_\varphi]}}, \tag{4.14}$$

where for $\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}$ we may substitute not only the RLD-based $\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$, but also: $\mathcal{F}_{\mathrm{as}}^{\mathrm{CS}} \geq \mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} \geq \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$; corresponding to the bounds derived via respectively the *Classical Simulation* (CS), *Quantum Simulation* (QS) and *Channel Extension* (CE) methods schematically explained in Fig. 4.3, but described in detail consecutively below.
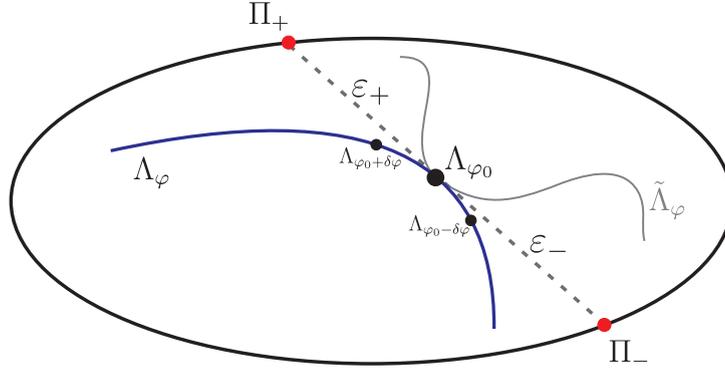
FIGURE 4.4: **Local CS of a φ-non-extremal quantum channel**. The family of φ-parametrised CPTP maps, $\{\Lambda_\varphi\}_\varphi$, is depicted, similarly to Fig. 2.3(**b**), as a trajectory (*blue curve*) in the convex set representing the space of all relevant quantum channels sharing the same input and output Hilbert spaces. As the map Λ_φ is *φ-non-extremal* at φ_0 it is possible to construct its local *Classical Simulation* (CS) there. In case of the *unitary* parameter-encoding [Demkowicz-Dobrzański *et al.*, 2012], the optimal CS corresponds to a probabilistic mixture of two channels lying at the intersection of the tangent and the boundary of the set: $\Pi_\pm$. Note that from the point of view of the QFI, any two channel trajectories, e.g. Λ_φ and $\tilde{\Lambda}_\varphi$ (*grey curve*), are equivalent at a given φ_0 as long as the form of the maps and their first derivatives w.r.t. φ coincide there.

4.3.1.1 Classical Simulation bound

Crucially, the notion of *φ-non-extremality* of a parametrised map Λ_φ (introduced in Sec. 2.4.2) carries a *natural geometric explanation* of why a given channel is bound to asymptotically follow the SQL-like scaling of precision. Returning in Fig. 4.4 to the picture representing the family of CPTP maps parametrised by φ as a trajectory in the convex space of quantum channels, we explain—following [Matsumoto, 2010]—that the channel φ-non-extremality at a given φ_0 assures a local *Classical Simulation* (CS) of the map Λ_φ to be feasible there, so that when considering the parallel action of channels, $\Lambda_\varphi^{\otimes N}$, in Fig. 4.1 and the $N \to \infty$ limit, the precision scaling in N is forced to behave as if the estimation problem was classical.

We term a channel Λ_φ to be *classically simulable* at the parameter value φ_0, if it can be written as a classical mixture of φ-independent channels $\{\Pi_x\}_x$ such that for $\varphi = \varphi_0 + \delta\varphi$ and any ϱ:

$$\Lambda_\varphi[\varrho] = \sum_x p_\varphi(x)\,\Pi_x[\varrho] + O(\delta\varphi^2) = \Phi[\varrho \otimes p_\varphi] + O(\delta\varphi^2), \tag{4.15}$$

so that the action of Λ_φ is mimicked on average by probabilistically choosing a *fixed* channel from the set $\{\Pi_x\}_x$ according to a random variable X distributed with $p_\varphi(X)$. Equivalently, as indicated in the second expression above and Fig. 4.3(**a**), one may write the decomposition (4.15) in a form, in which the channel Λ_φ is expressed as a φ-independent CPTP map, Φ, which apart from the input acts also on a diagonal density matrix, $p_\varphi = \sum_x p_\varphi(x)|e_x\rangle\langle e_x|$, and obeys: $\forall_{\varrho,x} : \Phi[\varrho \otimes |e_x\rangle\langle e_x|] = \Pi_x[\varrho]$. Crucially, for the CS to be valid, Eq. (4.15) must be satisfied only *locally*, as the QFI (3.57) just quantifies the sensitivity to parameter variations (see Eq. (3.59)) what at the level of quantum maps means that, from the point of view of the QFI, all channels are equivalent if they output density matrices that are identical up to the first order in $\delta\varphi$. On the other hand, as the action of any channel may be written with use of its CJ representation (see Eq. (2.25)) a given Λ_φ may thus be defined to be tantamount to another $\tilde{\Lambda}_\varphi$ at φ_0 from the local-estimation perspective, as long as their CJ matrices coincide there up to $O(\delta\varphi^2)$, i.e. $\Omega_{\Lambda_{\varphi_0}} = \Omega_{\tilde{\Lambda}_{\varphi_0}}$ and $\dot{\Omega}_{\Lambda_{\varphi_0}} = \dot{\Omega}_{\tilde{\Lambda}_{\varphi_0}}$, as geometrically shown in Fig. 4.4. Hence, for the CS

(4.15) to be feasible, we must find an ensemble $\{p_\varphi(x), \Pi_x\}_x$ that satisfies $\sum_x p_{\varphi_0}(x)\Omega_{\Pi_x} = \Omega_{\Lambda_{\varphi_0}}$ and $\sum_x \dot{p}_{\varphi_0}(x)\Omega_{\Pi_x} = \dot{\Omega}_{\Lambda_{\varphi_0}}$. This corresponds exactly to the requirement of *φ-non-extremality* of Λ_φ at φ_0 (see Def. 2.4.3), as these conditions can be matched *if and only if* there exist physical maps lying in both directions along the tangent to the trajectory at φ_0, what—as formally specified in Sec. 2.4.2—is true if one may find $\epsilon > 0$ such that both matrices $\Omega_{\Pi_\pm} = \Omega_{\Lambda_{\varphi_0}} \pm \epsilon \dot{\Omega}_{\Lambda_{\varphi_0}}$ are positive semi-definite[7], i.e. their corresponding $\Pi_\pm$ lie within the convex set depicted in Fig. 4.4.

Now, as the maps Λ_φ in the N-parallel–channels estimation scheme of Fig. 4.1 act *independently* on each of the particles, we can simulate the overall action of $\Lambda_\varphi^{\otimes N}$ with help of Eq. (4.15) by just associating N *independent* variables, X^N, with each of the channels. The estimation procedure may thus be described as follows

$$\varphi \to X^N \to \Lambda_\varphi^{\otimes N} \to \Lambda_\varphi^{\otimes N}\big[|\psi_{\text{in}}^N\rangle\big] \to \tilde{\varphi}, \tag{4.16}$$

so that the parameter is firstly encoded by infinite sampling of X^N onto the form of quantum channels acting on the input, $|\psi_{\text{in}}^N\rangle$, and then decoded by performing measurements on the output state, ρ_φ^N. Clearly, such a protocol can only be less efficient than the strategy in which we could infer the parameter directly from X^N, i.e. $\varphi \to X^N \to \tilde{\varphi}$, what corresponds to a *classical* estimation problem with maximal attainable precision dictated by the (classical) FI (3.8): $F_{\text{cl}}\big[p_\varphi^N\big] = NF_{\text{cl}}[p_\varphi]$. As result, we obtain a linearly scaling upper bound on the QFI of the output ρ_φ^N that yields the desired $\mathcal{F}_{\text{as}}^{\text{bound}}$ of (4.14), i.e.

$$F_{\text{Q}}\big[\rho_\varphi^N\big] \le NF_{\text{cl}}[p_\varphi] \qquad \Longrightarrow \qquad \mathcal{F}_{\text{as}}[\Lambda_\varphi] \le F_{\text{cl}}[p_\varphi], \tag{4.17}$$

being fully determined by the single-channel–mixing PDF $p_\varphi(X)$ of Eq. (4.15).

Importantly, the notion of *φ-non-extremality* of a given channel, as introduced in Fig. 2.3, naturally provides a valid CS (4.15) of Λ_φ that corresponds to the mixture of channels $\Pi_\pm$, which lie along the tangent to the trajectory at φ_0 in opposite directions and thus possess CJ matrices: $\Omega_{\Pi_\pm} = \Omega_{\Lambda_{\varphi_0}} \pm \epsilon_\pm \dot{\Omega}_{\Lambda_{\varphi_0}}$ with $\epsilon_\pm > 0$. Consequently, a locally equivalent channel $\tilde{\Lambda}_\varphi = p_{\varphi,+}\Pi_+ + p_{\varphi,-}\Pi_-$ may always be constructed at φ_0, as one can always adequately choose $p_{\varphi,\pm}$, so that $\Omega_{\tilde{\Lambda}_{\varphi_0}} = \Omega_{\Lambda_{\varphi_0}}$ and $\dot{\Omega}_{\tilde{\Lambda}_{\varphi_0}} = \dot{\Omega}_{\Lambda_{\varphi_0}}$. Moreover, one may easily verify that the correct binary PDF then yields $F_{\text{cl}}[p_{\varphi,\pm}] = 1/(\epsilon_+\epsilon_-)$, which constitutes a valid example of the bound (4.17). Importantly, as proved in [Demkowicz-Dobrzański *et al.*, 2012], for channel estimation problems in which the parameter is *unitarily* encoded, it is always *optimal* to employ such a two-point CS that—as depicted in Fig. 4.4—mixes channels $\Pi_\pm$ that lie at the intersection of the tangent line with the boundary of the quantum maps set. Such a choice leads to the tightest bound (4.17): $\mathcal{F}_{\text{as}}^{\text{CS}} = 1/(\varepsilon_+\varepsilon_-)$, which we refer to as the *CS bound*. Notably, the CS bound is thus fully dictated by the "distances" $\varepsilon_\pm$ between Λ_{φ_0} and the boundary along the tangent (see Fig. 4.4) that mathematically correspond to the largest possible values of $\epsilon_\pm$ for which the CJ matrices of $\Pi_\pm$ are still positive semi-definite.

4.3.1.2 Quantum Simulation bound

In [Matsumoto, 2010], a natural generalisation of the channel CS has been proposed, which—as schematically presented in Fig. 4.3(**a**)—corresponds to replacing the classical diagonal state p_φ in the second expression of Eq. (4.15) with a general quantum state σ_φ. Consequently, the so-called *Quantum Simulation* (QS) of channel Λ_φ is obtained[8]:

$$\Lambda_\varphi[\varrho] = \Phi[\varrho \otimes \sigma_\varphi] + O(\delta\varphi^2) = \text{Tr}_{\text{E}_\Phi\text{E}_\sigma}\big\{U\left(\varrho \otimes |\xi_\varphi\rangle\langle\xi_\varphi|\right)U^\dagger\big\} + O(\delta\varphi^2), \tag{4.18}$$

---

[7]Or equivalently, if $\dot{\Omega}_{\Lambda_\varphi}$ is contained within the support of Ω_{Λ_φ} at φ_0—see Crit. 2.4.4.

[8]Note that the notion of *quantum simulability* of a quantum map is equivalent to the *channel programmability* concept introduced in [Ji *et al.*, 2008].

which again must hold only locally for a given φ_0. In order to be able to later utilise the purification-based definitions (3.63)/(3.64) of the QFI, we have written the purified version of the QS in the second expression above, in which we have purified both the channel Φ (by utilising the Stinespring dilation theorem 2.2.1 represented pictorially in Fig. 2.1) and the auxiliary state σ_φ containing the complete information about the parameter, so that $\sigma_\varphi = \text{Tr}_{\text{E}_\sigma}\{|\xi_\varphi\rangle\langle\xi_\varphi|\}$. E_Φ and E_σ thus represent the adequate extra Hilbert spaces required for the purifications to be performed.

By analogous reasoning to the CS case, when Eq. (4.18) holds, we may upper-bound the QFI of the N-particle output state in the scheme of Fig. 4.1 as follows:

$$
\begin{aligned}
F_\text{Q}\big[\rho_\varphi^N\big] = F_\text{Q}\big[\Lambda_\varphi^{\otimes N}\big[|\psi_\text{in}^N\rangle\big]\big] &= F_\text{Q}\big[\Phi^{\otimes N}\big[|\psi_\text{in}^N\rangle\langle\psi_\text{in}^N| \otimes \sigma_\varphi^{\otimes N}\big]\big] = F_\text{Q}\big[\bar{\Pi}\big[\sigma_\varphi^{\otimes N}\big]\big] \\
&\leq F_\text{Q}\big[\sigma_\varphi^{\otimes N}\big] = N F_\text{Q}[\sigma_\varphi],
\end{aligned}
\tag{4.19}
$$

where we have introduced the *φ-independent* map $\bar{\Pi}[\bullet] = \Phi^{\otimes N}\big[|\psi_\text{in}^N\rangle\langle\psi_\text{in}^N| \otimes \bullet\big]$ to make it clear that it may only decrease the overall QFI (see Sec. 3.2.2.3), what leads then to the linearly-scaling upper limit (4.19), and thus the desired $\mathcal{F}_\text{as}^\text{bound} = F_\text{Q}[\sigma_\varphi]$. Notice that consistently, by replacing σ_φ with a diagonal state p_φ, we would recover Eq. (4.17) of the CS method. Hence, the CS may indeed be treated as a special type of the QS and Eq. (4.19) actually serves as an alternative proof of the asymptotic SQL-like precision scaling for classically simulable maps, which we have previously derived basing on the concept of N independent random variables associated with each channel use.

Similarly as in the case of the CS method, a *quantum simulable* channel may admit many decompositions (4.18) and the optimal one must yield the lowest $F_\text{Q}[\sigma_\varphi]$. Therefore, without loss of generality, in the search for the optimal QS, we may take U in Eq. (4.18) to act on the full purified system, i.e. also on the E_σ space. This enlarges the set of all possible QSs beyond the original ones, for which $U = U^{\text{SE}_\Phi} \otimes \mathbb{I}^{\text{E}_\sigma}$, and yields $\mathcal{F}_\text{as}^\text{bound} = F_\text{Q}[|\xi_\varphi\rangle]$, which, due to the purification-based definition (3.63) of the QFI, cannot be smaller than $F_\text{Q}[\sigma_\varphi]$. As a matter of fact, Eq. (3.63) ensures that for any QS employing σ_φ, there exists an "enlarged" decomposition (4.18) leading to the same $\mathcal{F}_\text{as}^\text{bound} = F_\text{Q}[\sigma_\varphi] = F_\text{Q}[|\xi_\varphi\rangle]$ with $|\xi_\varphi\rangle$ being then the minimal purification in Eq. (3.63).

Importantly, we prove in App. H that, in order for the QS (4.18) to be locally feasible at φ_0 and lead to a finite asymptotic bound, Λ_φ of rank r must admit Kraus operators $\{K_i(\varphi)\}_{i=1}^r$ that satisfy at φ_0 conditions:

$$
\text{i}\sum_{i=1}^r \dot{K}_i(\varphi_0)^\dagger K_i(\varphi_0) = 0 \qquad \text{and} \qquad \sum_{i=1}^r \dot{K}_i(\varphi_0)^\dagger \dot{K}_i(\varphi_0) = \frac{1}{4} F_\text{Q}[|\xi_\varphi\rangle]\big|_{\varphi_0} \mathbb{I}.
\tag{4.20}
$$

Hence, by optimising over all locally equivalent Kraus representations of Λ_φ—the ones related to one another by rotations (4.5) generated by any Hermitian h—that satisfy constraints (4.20), we may determine the asymptotic bound given by the optimal local QS, which we refer to as the *QS bound* $– \mathcal{F}_\text{as}^\text{QS}$, as follows:

$$
\mathcal{F}_\text{as}^\text{QS} = \min_\mathsf{h} \lambda \quad \text{s.t.} \quad \alpha_{\tilde{K}} = \frac{\lambda}{4}\mathbb{I}, \ \beta_{\tilde{K}} = 0,
\tag{4.21}
$$

where $\alpha_{\tilde{K}} = \sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \dot{\tilde{K}}_i(\varphi)$, $\beta_{\tilde{K}} = \text{i}\sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \tilde{K}_i(\varphi)$, h represents as before Hermitian generators locally shifting the first derivatives of Kraus operators, and λ has the interpretation of $\mathcal{F}_\text{as}^\text{bound} = F_\text{Q}[|\xi_\varphi\rangle]$.

Yet, one should note that by generalising the CS to QS, we have paid the price of losing the intuitive geometrical interpretation, as the set of quantum simulable channels now contains all maps that locally admit a Kraus representation satisfying conditions (4.20), which, however, cannot be rewritten neatly at the level of the channel CJ representation. On the other hand, as the CS method may be interpreted as the more general QS method with an extra constraint forcing σ_φ in Eq. (4.18) to be diagonal, not only

all φ-non-extremal channels must be quantum simulable, but also the QS bound must be at least as tight for them as the CS one. In other words, whenever the CS bound (4.17) is constructable: $\mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} \leq \mathcal{F}_{\mathrm{as}}^{\mathrm{CS}}$.

4.3.1.3 Channel Extension bound

In the *Channel Extension* (CE) method—as shown in Fig. 4.3(**b**)—each of the N channels in the scheme of Fig. 4.1 is *extended*, i.e. appended an ancillary particle unaffected by the action of Λ_φ exactly in the same fashion as discussed in Sec. 4.2.2 while introducing the extended-channel QFI (4.7). As such an extension can only improve the overall attainable precision, the N-channel QFI (4.11) can then be trivially upper-limited by the corresponding N-*extended-channel QFI*, which importantly can always be upper-bounded as follows [Fujiwara and Imai, 2008]:

$$\mathcal{F}\big[\Lambda_\varphi^{\otimes N}\big] \leq \mathcal{F}\big[(\Lambda_\varphi \otimes \mathcal{I})^{\otimes N}\big] \quad \leq \quad 4 \min_{\mathsf{h}} \left\{ N \left\| \alpha_{\tilde{K}} \right\| + N(N-1) \left\| \beta_{\tilde{K}} \right\|^2 \right\}, \tag{4.22}$$

where as before: $\|\dots\|$ represents the operator norm, $\alpha_{\tilde{K}} = \sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \dot{\tilde{K}}_i(\varphi)$, $\beta_{\tilde{K}} = \mathsf{i} \sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \tilde{K}_i(\varphi)$, and h is the generator of local Kraus-representation rotations (4.5). Crucially, if there exists a Kraus representation for which $\beta_{\tilde{K}} = 0$ so that the second term in Eq. (4.22) vanishes, $\mathcal{F}\big[\Lambda_\varphi^{\otimes N}\big]$ must asymptotically scale at most linearly in N. Hence [Fujiwara and Imai, 2008]:

Definition 4.3.1 ($\beta_{\tilde{K}} = 0$ condition).
The asymptotic SQL-like scaling of precision is assured in the scheme of Fig. 4.1 for a given channel Λ_φ of rank r and the parameter value φ_0, if one may find, for a particular set Kraus operators $\{K_i(\varphi)\}_{i=1}^r$ of Λ_φ, a Hermitian matrix h that satisfies at φ_0:

$$\sum_{i,j=1}^r \mathsf{h}_{ij} K_i(\varphi_0)^\dagger K_j(\varphi_0) = \mathsf{i} \sum_{i=1}^r \dot{K}_i(\varphi_0)^\dagger K_i(\varphi_0). \tag{4.23}$$

Notice that the above requirement, which we term as the $\beta_{\tilde{K}} = 0$ *condition*, turns out to be very effective, as for it to be applicable one needs only to know a particular Kraus representation[9] of a single channel without any further details of the single-particle evolution [Chaves *et al.*, 2013]. Furthermore, according to our best knowledge, no example of a parametrised quantum channel has been found that does not fulfil Eq. (4.23), but still is asymptotically constrained to follow the SQL-like precision scaling[10].

What is more, Eq. (4.22) allows to construct for any channel, that admits a generator h fulfilling the condition (4.23), an upper bound on the *asymptotic extended-channel QFI* as follows[11]:

$$\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi \otimes \mathcal{I}] = \lim_{N \to \infty} \frac{\mathcal{F}\big[(\Lambda_\varphi \otimes \mathcal{I})^{\otimes N}\big]}{N} \quad \leq \quad \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}} = 4 \min_{\substack{\mathsf{h} \\ \beta_{\tilde{K}}=0}} \left\| \sum_{i=1}^r \dot{\tilde{K}}_i(\varphi)^\dagger \dot{\tilde{K}}_i(\varphi) \right\|, \tag{4.24}$$

which also naturally constitutes the required asymptotic bound $\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi] \leq \mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}$ in Eq. (4.14) that we refer to as the *CE bound* – $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$. We have explicitly written the form of $\alpha_{\tilde{K}}$ in Eq. (4.24) to emphasise the similarity between the CE bound and the extended-channel QFI (4.8). The essential difference between Eqs. (4.8) and (4.24) is the $\beta_{\tilde{K}} = 0$ condition (4.23) yielding consistently $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}} \geq \mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] \geq \mathcal{F}[\Lambda_\varphi]$ and

---

[9]In fact, only the form of all K_i and $\dot{K}_i$ for a given, fixed φ_0.

[10]Moreover, Eq. (4.23) has been proved to ensure the asymptotic SQL-like scaling of precision even when one allows for *feedback* in the N-parallel–channels scheme of Fig. 4.1 [Escher *et al.*, 2011].

[11]Yet, we conjecture that the bound (4.24) is actually tight, so that the CE bound coincides with the asymptotic extended-channel QFI, i.e. $\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi \otimes \mathcal{I}] = \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$.

leaving room for potential asymptotic quantum enhancement of precision. Despite the extra constraint (4.23) imposed in Eq. (4.24), $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$ may always be computed similarly to Eq. (4.8) by reformulating the minimisation in Eq. (4.24) into an SDP—see App. I. Although the corresponding SDP is always efficiently solvable only numerically, it may be utilised to identify the non-zero entries of the optimal generator h in Eq. (4.24) and their complex structure, as the numerical accuracy of the SDP solutions may always be quantified. Hence, with help of the SDP we may then construct an ansatz for h in Eq. (4.24), with help of which the minimisation over h may be eventually performed analytically.

Importantly, one should note that the CE bound (4.24) resembles the QS bound (4.21), but *without* the additional constraint in Eq. (4.20) forcing the operator $\alpha_{\tilde{K}}$ to be proportional to identity. Hence, such an observation *proves* that not only the CE method applies to a wider class of parametrised quantum channels than the QS method—and hence also than the CS and RLD methods being further restricted only to the φ-non-extremal channels—but also the QS bound (4.21) can never outperform its CE equivalent, so that most generally: $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}} \leq \mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} \leq \mathcal{F}_{\mathrm{as}}^{\mathrm{CS}}$. On the other hand, we prove in App. G that also the RLD bound (4.9) applied to any φ-non-extremal map can never lead to a tighter bound on the asymptotic QFI, so that also $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}} \leq \mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$ and the CE method is indeed most effective out of the ones presented in this work.

Lastly, let us note that by appending *more than one ancillary particle per channel*, while performing the extension in Eq. (4.22), we could only improve the estimation precision (as trivially for any $k > 1$: $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] \leq \mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}^{\otimes k}]$ implying $\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi \otimes \mathcal{I}] \leq \mathcal{F}_{\mathrm{as}}[\Lambda_\varphi \otimes \mathcal{I}^{\otimes k}]$) and obtain a larger $\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}$, and thus a weaker upper bound on $\mathcal{F}_{\mathrm{as}}[\Lambda_\varphi]$, what worsens the method. However, for a given Λ_φ which satisfies Eq. (4.23), one may wonder whether the asymptotic SQL-like scaling can be beaten by just increasing k at some sufficiently large rate with N. Unfortunately, this can never be the case, as by adequately extending an effective *k-ancilla–channel*: $\Lambda_\varphi \otimes \mathcal{I}^{\otimes k}$, and applying to it the bound (4.22), one may easily verify that the operator norms in Eqs. (4.22) and (4.24) remain unaltered for any $k \geq 1$. Hence, not only Eq. (4.23) still constitutes a sufficient condition for the asymptotic SQL, but also the CE bound takes the form of (4.24) regardless of k. Such a behaviour, might have been expected at least for the φ-non-extremal channels, as a single ancilla is enough to provide a sufficient extension for the CJ representation of the map Λ_φ to be constructable, what is manifested by the RLD bound (4.9) also taking the same form independently of the ancilla number k.

Hence, in general, the *only* way to surpass the CE bound (4.24) is to find a way to alter the form of the channel considered, so that its Kraus representation *ceases* to fulfil the $\beta_{\tilde{K}} = 0$ condition (4.23). For instance, this has been achieved by considering channels that due to decoherence satisfy Eq. (4.23) at finite times, but not in the limit of infinitely short evolution. Such a behaviour has been shown to emerge when accounting for the Non-Markovianity effects in the single-particle evolution [Chin *et al.*, 2012; Matsuzaki *et al.*, 2011], but also when considering dephasing noise perpendicular to the phase-encoding evolution part [Chaves *et al.*, 2013]. In fact, for the second case it has been also shown that, due to the $\beta_{\tilde{K}} = 0$ condition (4.23) being violated at small times, one may actually benefit in this regime from increasing the number of ancillary particles, which may be then utilised to perform the *quantum error-correction* and fully retain the HL [Dür *et al.*, 2014].

4.3.1.4 Example: **Noisy-phase–estimation channels**

Similarly to the discussion of the single channel QFI measures in Sec. 4.2, we consider the noisy-phase–estimation models introduced in Sec. 2.2.3, in which the phase is encoded onto the particle represented by a qubit via the $\mathrm{e}^{-\mathrm{i}\hat{\sigma}_3\varphi}$ rotation and the decoherence is specified by one of the *dephasing, depolarisation, loss* and *spontaneous emission* maps (see Fig. 2.2 and Tab. 2.1). In case of the N-parallel–channel

Noise models:	Dephasing	Depolarization	Loss	Spontaneous emission
$\mathcal{F}_{as}^{CE}$ (4.24)	$\frac{\eta^2}{1-\eta^2}$	$\frac{2\eta^2}{(1-\eta)(1+2\eta)}$	$\frac{\eta}{1-\eta}$	$\frac{4\eta}{1-\eta}$
$\mathcal{F}_{as}^{QS}$ (4.21)	$\frac{\eta^2}{1-\eta^2}$	$\frac{2\eta^2}{(1-\eta)(1+2\eta)}$	$\frac{\eta}{1-\eta}$	n.a.
$\mathcal{F}^{RLD}[\Lambda_\varphi \otimes \mathcal{I}]$ (4.9)	$\frac{\eta^2}{1-\eta^2}$	$\frac{2\eta^2(1+\eta)}{(1-\eta)(1+3\eta)}$	n.a.	n.a.
$\mathcal{F}_{as}^{CS}$ (4.17)	$\frac{\eta^2}{1-\eta^2}$	$\frac{4\eta^2}{(1-\eta)(1+3\eta)}$	n.a.	n.a.
$\chi[\Lambda_\varphi]$ (4.13)	$= \sqrt{\frac{1}{1-\eta^2}}$	$\leq \sqrt{\frac{2}{(1-\eta)(1+2\eta)}}$	$= \sqrt{\frac{1}{1-\eta}}$	$\leq \sqrt{\frac{4}{1-\eta}}$

TABLE 4.2: **CE, QS, RLD and CS bounds on the asymptotic channel QFI (4.12)** for the noisy-phase–estimation channels introduced in Sec. 2.2.3 and depicted in Fig. 2.2. For each noise model, we also present an upper bound on the *maximal quantum enhancement of precision* (4.13), which is obtained by utilising the corresponding CE bound and the channel QFI presented in Tab. 4.1, so that $\chi[\Lambda_\varphi] \leq \sqrt{\mathcal{F}_{as}^{CE}[\Lambda_\varphi]/\mathcal{F}[\Lambda_\varphi]}$. Yet, in the case of dephasing and loss channels the corresponding values of the limits on $\chi[\Lambda_\varphi]$ have been shown to be attainable [Caves, 1981; Ulam-Orgikh and Kitagawa, 2001].
[n.a.—not applicable]

estimation scheme of Fig. 4.1, we model the evolution of each of the N particles by one of the above models, which thus determines the single-particle channel Λ_φ. In Tab. 4.2, we present in an increasing order the corresponding bounds on the asymptotic QFI (4.12) obtained via the methods analysed in this section, where in the last row we list the upper limits on the maximal quantum enhancement of precision (4.13) specified by the ratios of the adequate CE bounds and the single-channel QFIs presented in Tab. 4.1, i.e. $\chi[\Lambda_\varphi] \leq \sqrt{\mathcal{F}_{as}^{CE}[\Lambda_\varphi]/\mathcal{F}[\Lambda_\varphi]}$.

Again, as only the phase estimation scenarios with *dephasing* and *depolarisation* noise-types lead to the effective φ-*non-extremal* channels, only for these we may limit the asymptotic precision by utilising the CS and RLD methods. In case of the *dephasing* noise, which yields the simplest, rank-2 Λ_φ channel (see Secs. 2.2.3 and 2.4.3) all the asymptotic SQL-like bounds in Tab. 4.2 take the same form, proving that the geometric CS method is sufficient to determine the maximal quantum enhancement of precision. It is so, as the CS bound (4.17) has indeed been shown to coincide with the asymptotic channel QFI (4.12) [Demkowicz-Dobrzański *et al.*, 2015; Ulam-Orgikh and Kitagawa, 2001]. As the *depolarisation* noise yields a full-rank channel that does not lie on any of the boundaries of the CPTP-maps space—see Sec. 2.4.3—the RLD bound (4.17) proves for it to be tighter than the CS one. Nevertheless, the more general QS method provides an even better bound, but most importantly also applies to the φ-*extremal* channel obtained accounting for the *loss* noise-type. Furthermore, for all these three channels the QS bounds (4.21) turn out to be as accurate as the CE ones, what may be also verified by inspecting the relevant optimal Kraus-representation generators h of the CE method listed in App. F, which indeed yield $\alpha_{\tilde{K}}$ to be proportional to identity—satisfying the extra QS constraint (4.20). However, in the case of *spontaneous-emission* noise, the QS method ceases to work, as the $\beta_{\tilde{K}} = 0$ condition (4.23) fixes $\alpha_{\tilde{K}}$ to be disproportional to identity. It is so, because—as explained in Sec. 2.4.3—the spontaneous-emission channel is *extremal* and the CE method is the only one able to deal with it.

One should also note that the CE bounds, $\mathcal{F}_{as}^{CE}$, indeed provide the best limits on the asymptotic QFI for all the cases and, that is why, we employ them to upper-bound the maximal quantum enhancement of precision, $\chi[\Lambda_\varphi]$, presented in the last row of Tab. 4.2. Let us remark that, for *dephasing* and *loss* noise models, phase estimation strategies have been found that asymptotically attain the corresponding

CE bounds [Caves, 1981; Demkowicz-Dobrzański *et al.*, 2015; Ulam-Orgikh and Kitagawa, 2001]. Hence in these two cases, not only $\mathcal{F}_{\text{as}}[\Lambda_\varphi] = \mathcal{F}_{\text{as}}^{\text{CE}}$, but also one may not asymptotically benefit from the channel extension and $\mathcal{F}_{\text{as}}[\Lambda_\varphi] = \mathcal{F}_{\text{as}}[\Lambda_\varphi \otimes \mathcal{I}]$ (see Eq. (4.24)). On the other hand, as shown in Tab. 4.1, these two channels are also examples of ones for which the extension does not improve the precision at the single-channel level, i.e. $\mathcal{F}[\Lambda_\varphi] = \mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$, but the problem of relating the two regimes we leave open for future research.

Lastly, let us comment that all the bounds presented in Tab. 4.2 correctly diverge as $\eta \to 1$, when we return to the noiseless unitary phase estimation problem, in which the $1/N^2$ HL limit must be attainable and any SQL-bounding methods of Sec. 4.3.1 must fail. Geometrically, for φ-non-extremal channels (here dephasing and depolarisation noises) such limit corresponds to decreasing the distances $\varepsilon_\pm$ to the boundary of the quantum channels set in Fig. 4.4, so that the φ-non-extremality is eventually lost and both CS and RLD methods cease to apply (i.e. the CS (4.17) and RLD (4.9) bounds diverge, as respectively $\mathcal{F}_{\text{as}}^{\text{CS}} = 1/(\varepsilon_-\varepsilon_+) \overset{\varepsilon_\pm \to 0}{=} \infty$ and $\dot{\Omega}_{\mathcal{U}_\varphi}$ is no longer contained within the support of $\Omega_{\mathcal{U}_\varphi}$). On the other hand, in the case of the QS and CE methods, the necessary $\beta_{\tilde{K}} = 0$ condition (4.23) cannot be satisfied for the noiseless unitary evolution, so that (in contrast to extended-channel QFI (4.7) which definition (4.8) crucially lacks the $\beta_{\tilde{K}} = 0$ constraint) $\|\alpha_{\tilde{K}}\|$ diverges in both Eqs. (4.21) and (4.24) as $\eta \to 1$, and the QS and CE bounds become unbounded.

4.3.2 Finite-N CE bound

In Sec. 4.3.1, we have presented the *CE method* as the most effective one that applies to the broadest class of parametrised channels (i.e. ones satisfying the $\beta_{\tilde{K}} = 0$ condition (4.23)) and provides the tightest upper limits on the asymptotic QFI (4.12) and the maximal quantum enhancement of precision (4.13). On the other hand, when moderate numbers of particles, N, are considered in the scheme of Fig. 4.1, the CE bound (4.24) derived for the asymptotic regime, despite still being valid, is far too weak to be useful. Although for very low values of N the precision can be quantified numerically, for instance, by brute-force–type methods computing explicitly the QFI, in the intermediate-N regime—being beyond the reach of computational power, yet with N too low for the asymptotic methods to be effective—more accurate bounds should play an important role.

We propose the *finite-N CE method* which, despite still being based *only* on the properties of a *single* channel, provides bounds on precision that are relevant in the intermediate-N regime. We utilise the upper limit (4.22) on the N-extended-channel QFI and construct the *finite-N CE bound*, $\mathcal{F}_N^{\text{CE}}$, as follows

$$\frac{\mathcal{F}\left[(\Lambda_\varphi \otimes \mathcal{I})^{\otimes N}\right]}{N} \quad \leq \quad \mathcal{F}_N^{\text{CE}} = 4 \min_{\mathsf{h}_N} \left\{ \|\alpha_{\tilde{K}}\| + (N-1)\|\beta_{\tilde{K}}\|^2 \right\}, \tag{4.25}$$

where in contrast to the CE bound (4.24) we do not impose the SQL-bounding condition $\beta_{\tilde{K}} = 0$ (4.23). We rather search for the minimal Kraus representation at each N, so that the optimal generator, h_N, now varies depending on the particle number. Importantly, we show in App. I that $\mathcal{F}_N^{\text{CE}}$ can similarly to $\mathcal{F}_{\text{as}}^{\text{CE}}$ be efficiently evaluated numerically by recasting the minimisation over h_N in Eq. (4.25) into an SDP. Note that for $N = 1$, the finite-N CE bound (4.25) coincides with the extended-channel QFI (4.8), i.e. $\mathcal{F}_{N=1}^{\text{CE}} = \mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$, whereas for $N \to \infty$, (if a given channel allows the $\beta_{\tilde{K}} = 0$ condition to be fulfilled) $\mathcal{F}_N^{\text{CE}}$ attains the CE bound (4.24), i.e. $\mathcal{F}_{N\to\infty}^{\text{CE}} = \mathcal{F}_{\text{as}}^{\text{CE}}$, and the optimal h_N in Eq. (4.25) converges to h

minimising Eq. (4.24). As a consequence, $\mathcal{F}_N^{\mathrm{CE}}$ varies smoothly between these two regimes and it may provide much more accurate bounds on precision than its asymptotic version[12].

On the other hand, let us emphasise that the finite-N CE bound (4.25) also applies to channels for which it is impossible to fulfil the $\beta_{\tilde{K}} = 0$ condition (4.23) and thus the CE method fails. Moreover, as Eq. (4.25) is solved independently for each N, the finite-N CE method may also be utilised in scenarios in which the single-particle evolution depends on the number of particles employed, or in other words, in which the overall scheme may be viewed as the one of Fig. 4.1 but with each channel now denoted as $\Lambda_{\varphi, N}$ to indicate that its form may change with N. Physically, for instance, such a description is valid when analysing schemes in which the strength of decoherence varies depending on the number of particles involved in an experiment [Wasilewski *et al.*, 2010], or in frequency estimation scenarios discussed in Sec. 4.4, in which one may vary the time duration of each experimental "shot" given a particular N [Huelga *et al.*, 1997]. As a matter of fact, $\mathcal{F}_N^{\mathrm{CE}}$ has been explicitly used in [Chaves *et al.*, 2013], where a frequency estimation scheme was considered and it has been numerically demonstrated that the finite-N CE bound (4.25) allows to prove the correct asymptotic super-classical precision scaling, $1/N^{5/3}$, reaching *beyond* the SQL.

Returning again to the noisy-phase–estimation channels of Fig. 2.2 with uncorrelated noise modelled by respectively: *dephasing, depolarisation, loss* and *spontaneous emission* maps; we study numerically via SDPs the form of their corresponding finite-N CE bounds (4.25). Surprisingly, we observe that in all four cases $\mathcal{F}_N^{\mathrm{CE}}$ may be simply related to its asymptotic form[13] as

$$\mathcal{F}_N^{\mathrm{CE}} = \frac{N\,\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}}{N + \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}}, \tag{4.26}$$

where one should substitute for $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$ the corresponding CE bounds presented in Tab. 4.2. Notice, that for all but the spontaneous emission noise-models the form of the finite-N CE bound (4.26) allows us to establish a connection between the extended-channel QFI (4.7) and the CE bound (4.24), as by writing Eq. (4.26) for $N = 1$ we obtain the relation $\mathcal{F}[\Lambda_{\varphi} \otimes \mathcal{I}] = \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}/(1 + \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}})$ that may be verified for dephasing, depolarisation and loss noise-types by substituting their $\mathcal{F}[\Lambda_{\varphi} \otimes \mathcal{I}]$ and $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$ listed in Tabs. 4.1 and 4.2.

4.3.3 EXAMPLE: N-qubit phase estimation in the presence of loss and dephasing

For *dephasing* and *loss* decoherence models of Fig. 2.2, we show explicitly in Fig. 4.5 both the asymptotic (4.24) and the finite-N (4.25) CE bounds accompanied by the plots of the actual estimator variances and QCRBs (3.55) evaluated for particular phase estimation strategies that are optimal either in the small or large particle-number regime.

In the case of *dephasing noise*, we consider a Ramsey spectroscopy setup of [Bollinger *et al.*, 1996; Wineland *et al.*, 1994, 1992] in which the particles (atoms) are prepared in a spin-squeezed state [Ma *et al.*, 2011; Ulam-Orgikh and Kitagawa, 2001]. The parameter is then encoded in the phase of a unitary rotation generated by the total angular momentum of the atoms that independently experience dephasing [Huelga *et al.*, 1997], what constitutes an example of the phase estimation scheme of Fig. 3.6

---

[12]Nevertheless, we conjecture that, in order to actually construct a tight bound in Eq. (4.25) that when optimised coincides with $\mathcal{F}\big[(\Lambda_{\varphi} \otimes \mathcal{I})^{\otimes N}\big]$ (or even $\mathcal{F}\big[\Lambda_{\varphi}^{\otimes N}\big]$) for finite N, one may not restrict to *single*-channel generators h_N as in the case of Eq. (4.25), but must also account for purifications that are generated by operations performed collectively on the output of many channels acting in parallel.

[13]In case of the spontanous-emission–noise the formula (4.26) is valid only for $N \geq 2$, what we suspect to be a consequence of the spontaneous emission channel being an extremal map (see Sec. 2.4.3 and Tab. 2.2).
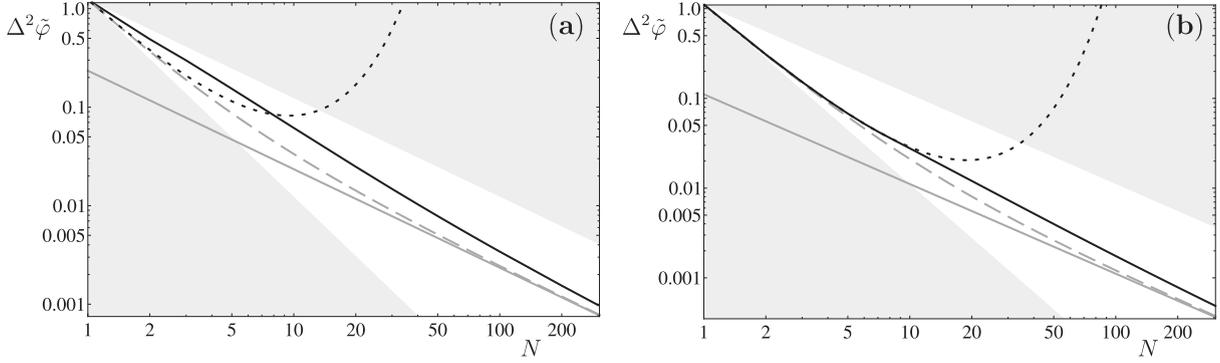
FIGURE 4.5: **Applicability of the CE bounds in noisy-phase–estimation scenarios.**
Asymptotic (*solid grey line*, 4.24) and finite-N (*dashed grey line*, 4.25) CE bounds compared with $\Delta^2\tilde{\varphi}$ achieved by various strategies in noisy phase estimation. Shaded areas represent regions in which the estimator MSE is either worse than SQL, $1/N$, or surpasses the HL, $1/N^2$.
(**a**) *Dephasing noise* of strength $\eta = 0.9$: (*solid black line*) – spin-squeezed states and Ramsey-type measurements [Ma *et al.*, 2011], (*dotted black line*) – QCRB (3.55) evaluated for GHZ/NOON states (3.78).
(**b**) *Loss model* with $\eta = 0.9$ particle survival probability: (*solid black line*) – QCRB (3.55) minimised over input states, (*dotted black line*) – QCRB (3.55) evaluated for GHZ/NOON states (3.78).

with uncorrelated noise. By measuring the total angular momentum perpendicular to the one generating the estimated phase change, the parameter may be reconstructed by registering the atomic population distribution. Such a scenario may be interpreted exactly as the Mach-Zehder interferometry setup of Fig. 3.1 with a photon-number difference measurement, where the atoms are represented by a finite number of photons involved and the uncorrelated dephasing noise accounts for the modal mismatch leading to imperfect visibility of the interferometer [Demkowicz-Dobrzański *et al.*, 2015]. We plot the MSE achieved in such a protocol as the *solid black line* in Fig. 4.5(**a**), which importantly saturates the dephasing CE bound of Tab. 4.2 (*solid grey line*) proving the scheme to be asymptotically optimal. For comparison, the maximal precision theoretically achievable by the GHZ input states (or equivalently the NOON states when viewed in the modal picture—see Eq. (3.78)) is also shown (*dotted black line*), i.e. the QCRB (3.55) with QFI: $\mathcal{F}_N^{\mathrm{NOON}} = \eta^{2N}N^2$, which for low N attains the finite-N CE bound (*dashed grey line*). Such an observation proves that in experiments with only few particles involved, such as [Leibfried *et al.*, 2004], it has been correctly assumed to use the GHZ states and completely ignore the uncorrelated dephasing present.

In Fig. 4.5(**b**), we depict the precision achieved in the phase estimation scenario in the presence of losses. The *loss noise-model* of Fig. 2.2 acting independently on each particle simulates a process in which each particle may survive and sense the estimated parameter with probability η. In case of the Mach-Zehnder interferometer of Fig. 3.1, this effectively corresponds to setting the power transmittance of both interferometer arms to η, what we explicitly show in Chap. 5 while discussing in detail the lossy interferometer scenario. Here, we plot the QCRB (3.55) for the numerically optimised N-particle (or equivalently N-photon) input states (*solid black line*) and for GHZ/NOON states (3.78) (*dotted black line*), which for losses lead to the QFI: $\mathcal{F}_N^{\mathrm{NOON}} = \eta^N N^2$. Crucially, for the loss noise-model one may prove that it is optimal to consider only inputs consisting of indistinguishable, bosonic particles (see Chap. 5 and [Demkowicz-Dobrzański *et al.*, 2009]), so that it is possible to numerically compute and minimise the QCRB, and thus explicitly perform the optimisation over the N-photonic input states. As a result, we are able to assess the tightness of the finite-N CE bound, i.e. $\mathcal{F}\left[\Lambda_\varphi^{\otimes N}\right] \leq \mathcal{F}_N^{\mathrm{CE}}$, by inspecting the gap between the *solid black* and *dashed grey* lines in Fig. 4.5(**b**). Although evidently being a good approximation to $\mathcal{F}\left[\Lambda_\varphi^{\otimes N}\right]$ for low and very large particle numbers, $\mathcal{F}_N^{\mathrm{CE}}$ ceases to be tight in the moderate-N regime,

in which more accurate bounds would require more information about the estimation capabilities than is contained in the form of a single channel. As in the previous case, for very low N the effects of decoherence may be ignored and the GHZ/NOON-state–based strategy (*dotted black line*) saturates the finite-N CE bound (*dashed grey line*).

4.4   Local frequency estimation in atomic models

In this section, we study the precision measures as well as the bounds developed in Secs. 4.2 and 4.3 in the case of *frequency estimation* problems of *atomic spectroscopy*. The general Ramsey spectroscopy setup—as introduced in [Bollinger *et al.*, 1996; Wineland *et al.*, 1994, 1992]—describes N identical two-level atoms, i.e. spin-1/2 particles or qubits, and the estimated parameter is typically the *detuning*, ω, of an external oscillator frequency from the atoms transition frequency dictated by their energy spacing. In general, the evolution of such an N-qubit system is modelled by a master equation (2.18) (in the LGKS form (2.16)) introduced in Sec. 2.2.2:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_\omega^N(t) = \sum_{n=1}^{N} -\mathrm{i}\frac{\omega}{2}\left[\hat{\sigma}_z^{(n)},\rho_\omega^N(t)\right] + \mathcal{L}^{(n)}\left[\rho_\omega^N(t)\right], \tag{4.27}$$

where $\hat{\sigma}_z^{(n)}$ is the Pauli operator generating a unitary rotation of the n-th atom around the z axis in its Bloch-ball representation[14]. The *uncorrelated noise* is represented by the Liouvillian part $\mathcal{L}^{(n)}$ acting independently on each (here the n-th) particle, so that the integral of Eq. (4.27) for a given input state $\rho_\omega^N(0) = |\psi_{\mathrm{in}}^N\rangle\langle\psi_{\mathrm{in}}^N|$ may be interpretted as the t-dependent output state of the *N-parallel–channels estimation scheme* depicted in Fig. 4.1: $\rho_\omega^N(t) = \Lambda_{\omega,t}^{\otimes N}\big[|\psi_{\mathrm{in}}^N\rangle\big]$. As generally the master equation (4.27) accounts for the fact that the decoherence process occurs simultaneously and is non-separable from the Hamiltonian part of the evolution, the form of the single-particle channel $\Lambda_{\omega,t}$ may be highly non-trivial. Nevertheless, it is in priciple always computable, as both the CJ and the Kraus representations of $\Lambda_{\omega,t}$ may be established basing on Eq. (4.27) either by rewriting the action of the channel with use of the CJ matrix introduced in Sec. 2.4.1 [Bengtsson and Życzkowski, 2006], or directly from the LGKS form by a more general construction [Andersson *et al.*, 2007]. As a consequence, this means that in principle all the techniques described in Secs. 4.2 and 4.3 that assess the estimation capabilities stemming from the geometry of maps $\Lambda_{\omega,t}$ parametrised by ω or from the Kraus-representation–optimisation procedures should be applicable.

However, in order to correctly quantify the resources [Huelga *et al.*, 1997], we must assume the complete spectroscopy experiment to take a *fixed* overall time[15] T, during which the estimation procedure is repeated $\nu = T/t$ times, so that by varying the single-shot duration, t, the repetition number, ν, in the scheme of Fig. 4.1 is altered. Consequently, the QCRB (3.55) on the MSE (3.3) of an estimator $\tilde{\omega}$ of the detuning frequency can be conveniently rewritten as

$$\Delta^2\tilde{\omega} \quad \geq \quad \frac{1}{T\,f_t[\rho_\omega^N(t)]} \quad \geq \quad \frac{1}{T}\frac{1}{\max_{0\leq t\leq T} f_t\big[\Lambda_{\omega,t}^{\otimes N}\big[|\psi_{\mathrm{in}}^N\rangle\big]\big]}, \tag{4.28}$$

where $f_t\big[\rho_\omega^N(t)\big] = F_\mathrm{Q}\big[\rho_\omega^N(t)\big]/t$ is now the effective '*QFI per shot duration*'. As remarked in the second expression above, in order to establish the *minimal* QCRB for a given input, one must also optimise the single-shot duration, which importantly may change depending on the number of the particles employed.

---

[14]In case of channels specified in Tab. 2.1 this just corresponds to setting $\varphi = \omega t$.

[15]For an alternative approach, where rather than the overall time the rate of particle production is fixed, see [Shaji and Caves, 2007].

Noise models:	Dephasing	Depolarisation	Loss	Spontaneous emission
$\mathfrak{f}[\Lambda_\omega]$	$\frac{1}{2\,\mathrm{e}\gamma}$	$\frac{3}{4\,\mathrm{e}\gamma}$	$\frac{1}{\mathrm{e}\gamma}$	$\frac{1}{\mathrm{e}\gamma}$
$\mathfrak{f}[\Lambda_\omega\otimes\mathcal{I}]$	$\frac{1}{2\,\mathrm{e}\gamma}$	$\approx 1.27\times\frac{3}{4\,\mathrm{e}\gamma}$	$\frac{1}{\mathrm{e}\gamma}$	$\frac{4\tilde{w}}{\gamma\left(1+\mathrm{e}^{\tilde{w}/2}\right)^2}$
$\mathfrak{f}_N^{\mathrm{CE}}$ $(N\geq 2)$	$\frac{N}{2\gamma}\frac{w_1[N]}{1+(\mathrm{e}^{w_1[N]}-1)N}$	$\frac{3N}{4\gamma}\frac{\alpha\,w_\beta[N]}{2+\left(\mathrm{e}^{\frac{\alpha}{4}w_\beta[N]}-1\right)\left(\mathrm{e}^{\frac{\alpha}{4}w_\beta[N]}+2\right)N}$	$\frac{N}{\gamma}\frac{w_1[N]}{1+(\mathrm{e}^{w_1[N]}-1)N}$	$\frac{N}{\gamma}\frac{4\,w_4[N]}{4+(\mathrm{e}^{w_4[N]}-1)N}$
$\mathfrak{f}_{\mathrm{as}}^{\mathrm{CE}}$	$\frac{1}{2\gamma}$	$\frac{1}{\gamma}$	$\frac{1}{\gamma}$	$\frac{4}{\gamma}$
$\chi[\Lambda_\omega]$	$=\sqrt{\mathrm{e}}$	$\leq\sqrt{\frac{4\mathrm{e}}{3}}$	$=\sqrt{\mathrm{e}}$	$\leq\sqrt{2\mathrm{e}}$

TABLE 4.3: **Channel QFIs, CE bounds and quantum enhancements in frequency estimation.** In frequency estimation tasks, the precision is maximised by adjusting the single experimental shot duration t. The t-optimised (extended) channel QFIs as well as their finite-N and asymptotic CE bounds are presented, where $w_x[N] = 1 + W\left[\frac{x-N}{\mathrm{e}N}\right]$, $\tilde{w} = 1 + 2\,W\left[\frac{1}{2\sqrt{\mathrm{e}}}\right]$ and $W[x]$ is the Lambert W function. As in the case of depolarizing channel not all the solutions possess an analytic form, only their numerical approximations are shown with $\alpha\approx 2.20$ and $\beta\approx 1.32$. In the last row, upper bounds on the maximal quantum enhancements of precision, $\chi[\Lambda_\varphi]\leq\sqrt{\mathfrak{f}_{\mathrm{as}}^{\mathrm{CE}}[\Lambda_\varphi]/\mathfrak{f}[\Lambda_\varphi]}$, are listed that now account for the t-optimisation of the most efficient classical and quantum strategies. As before in Tab. 4.2 when estimating phase, for depolarisation and spontaneous emission maps these limits may possibly be not achievable due to the channel extension assumed by the CE bound.

As a result, the form of the single-particle evolution map $\Lambda_{\omega,t}$ in the picture of Fig. 4.1 varies with N via $t=t_{\mathrm{opt}}(N)$, so that one must be careful when applying the precision-bounding techniques of Sec. 4.3, which have been derived assuming a fixed form of the single-particle channel. On the other hand, notice that the total time T in Eq. (4.28) plays now the role of ν in Eq. (3.55), so that the local estimation regime (see Sec. 3.1.2.3) may always be guaranteed for a given t by letting $T\gg t$. Furthermore, this means that for protocols in which $t_{\mathrm{opt}}(N)\to 0$ as $N\to\infty$, the QCRB (4.28) is always assured to be saturable in the asymptotic N limit, which then additionally warrants $\nu\to\infty$.

As now $f_t[\varrho_\omega(t)]$ in Eq. (4.28) plays the role of the QFI in Eq. (3.55), we may define with its help the frequency estimation equivalents of all the single-channel precision measures discussed in Sec. 4.2, as well as the N-parallel–channel quantities and the bounds derived in Sec. 4.3. For instance, the channel QFI (4.1) may be reformulated as:

$$\mathfrak{f}[\Lambda_\omega] = \max_{0\leq t\leq T}\ \max_{\psi_{\mathrm{in}}}\ f_t[\Lambda_{\omega,t}[|\psi_{\mathrm{in}}\rangle]] \tag{4.29}$$

that represents now the *maximal channel QFI per shot duration*, which is optimised over not only the single-particle inputs $|\psi_{\mathrm{in}}\rangle$ but also the shot duration t. In a similar manner, we may define: $\mathfrak{f}[\Lambda_\omega\otimes\mathcal{I}]$ for (4.7), $\mathfrak{f}^{\mathrm{RLD}}[\Lambda_\omega\otimes\mathcal{I}]$ for (4.9), $\mathfrak{f}_{\mathrm{as}}[\Lambda_\omega]\leq\mathfrak{f}_{\mathrm{as}}^{\mathrm{bound}}$ for (4.14) with CE, QS and CS bounds respectively then reading: $\mathfrak{f}_{\mathrm{as}}^{\mathrm{CS}}\geq\mathfrak{f}_{\mathrm{as}}^{\mathrm{QS}}\geq\mathfrak{f}_{\mathrm{as}}^{\mathrm{CE}}$, and the finite-$N$ CE bound (4.25) equivalent $\mathfrak{f}_N^{\mathrm{CE}}$, which now needs to be maximised over t independently for each N. Consequently, a t-optimised upper bound on the maximal quantum enhancement of precision (4.13) may also be constructed as $\chi[\Lambda_\varphi]\leq\sqrt{\mathfrak{f}_{\mathrm{as}}^{\mathrm{CE}}[\Lambda_\varphi]/\mathfrak{f}[\Lambda_\varphi]}$.

As before, we consider the decoherence models represented by the *dephasing*, *depolarisation*, *loss* and *spontaneous emission* qubit maps, which corresponding Liouvillians of Eq. (4.27) have been specified in Sec. 2.2.3. However, as these noise-types *commute* with the unitary evolution part, the overall process

may still be described with use of the noisy-phase–estimation channels depicted in Fig. 2.2 after just setting $\varphi = \omega t$ and[16] $\eta \to \eta(t)$. The adequate substitutions for the time-dependent strength of decoherence, which importantly agree with the corresponding Liouvillians, can also be found in Sec. 2.2.3. Crucially, the commutativity property allows us to directly establish the t-optimised frequency estimation equivalents of all quantities, $\mathcal{F}$, listed in Tabs. 4.1 and 4.2 by just substituting in each of them for $\eta(t)$ according to Eq. (4.27) and computing $\mathfrak{f} = \max_{0 \leq t \leq T} \mathcal{F} t$, which form becomes evident after realising that the 'QFI per shot duration' in Eq. (4.29) may always be rewritten in terms of the QFI computed w.r.t. φ, as for $\varphi = \omega t$ $F_{\mathrm{Q}}[\varrho_\omega] = F_{\mathrm{Q}}[\varrho_\varphi] t^2$.

In Tab. 4.3, we present the channel QFIs relevant for frequency estimation, their asymptotic and finite-N CE bounds, as well as the upper bounds on the maximal quantum enhancements of precision for each noise-model considered. In the case of dephasing, we correctly recover the results of Escher *et al.* [2011]; Huelga *et al.* [1997]. Let us emphasize that all the quantities listed are independent of the total experimental time T, what is really a consequence of the noise present, which forces the optimal single-shot duration, t_{opt}, to be finite, as by letting the system evolve for too long the decoherence effects dominate making the ω-estimation impossible. Moreover, by increasing the number of particles involved the noise has even more severe impact on the attainable precision, so that one must conduct shorter experimental shots as N grows and $t_{\mathrm{opt}}(N) \overset{N \to \infty}{\longrightarrow} 0$. As a result, one may always set large enough T, so that $\forall_N : t_{\mathrm{opt}}(N) \ll T$, and the time characteristics become indeed fully determined by the decoherence process. This dramatically contrasts the noiseless case, in which it is always optimal to estimate as long as possible to just gather the maximal information about ω, and thus set $t = T$ independently of[17] N. Such a sudden change of the characteristics causes the channel and extended-channel measures: $\mathfrak{f}[\Lambda_\omega]$ and $\mathfrak{f}[\Lambda_\omega \otimes \mathcal{I}]$ in Tab. 4.3, to diverge in the limit of vanishing decoherence $\gamma \to 0$, what contrasts the case of their phase-estimation equivalents listed in Tab. 4.1 that remain finite when evaluated for $\eta = 1$ (*not* to be confused with the asymptotic bounds of Tab. 4.2 that also diverge when $\eta \to 1$).

4.5 Local estimation of the decoherence-strength parameter

Lastly, we would like to explicitly demonstrate that the SQL-bounding methods introduced in Sec. 4.3.1, i.e. the CS, RLD, QS and CE methods, may also be applied to estimation tasks in which the estimated parameter is *not* unitarily encoded, so that the estimation problem may *not* be any more described by the noisy phase estimation scenario of Fig. 3.6, but rather *only* by the *N-parallel–channels estimation scheme* of Fig. 4.1. In order to do so, we study the qubit decoherence models previously employed in the noisy-phase–estimation channels of Fig. 2.2, but this time with the parameter to be estimated being the *decoherence strength* η, so that $\Lambda_\eta = \mathcal{D}_\eta$ (see Tab. 2.1). This kind of a problem has been widely considered not only in the estimation theory [Adesso *et al.*, 2009; Frey *et al.*, 2011; Hotta *et al.*, 2005; Monras and Paris, 2007], but also when examining issues of channel discrimination [D'Ariano *et al.*, 2005; Sacchi, 2005] with particular application in quantum reading [Nair, 2011; Pirandola, 2011; Pirandola *et al.*, 2011].

As compared to unitary rotations, the nature of the estimated parameter is then dramatically different. In case of a phase-like parameter estimation (see Sec. 3.2.4.1) the use of maximally entangled input state of N particles (3.78) results in an effective N-times greater "angular speed" of rotation leading to the HL-like scaling in the absence of noise. In decoherence-strength estimation tasks, a change in the parameter value

---

[16]For an example of the application of the precision-bounding methods to a frequency estimation model described by Eq. (4.27) with a non-commutative dissipative part, see [Chaves *et al.*, 2013].

[17]Note that due to the presence of noise also the locality of estimation is assured in the asymptotic N limit, whereas in the noiseless scenario we run again into the problem of correctly quantifying the resources, as we must still repeat the whole experiment enough times for the QCRB (4.28) to be meaningful.

Noise models:	*Dephasing*	*Depolarisation*	*Loss*	*Spontaneous emission*
$\mathcal{F}[\Lambda_\eta]$ (4.1)	$\frac{1}{1-\eta^2}$	$\frac{1}{1-\eta^2}$	$\frac{1}{\eta(1-\eta)}$	$\frac{1}{\eta(1-\eta)}$
$\mathcal{F}[\Lambda_\eta \otimes \mathcal{I}]$ (4.7)	$\frac{1}{1-\eta^2}$	$\frac{3}{(1-\eta)(1+3\eta)}$	$\frac{1}{\eta(1-\eta)}$	$\frac{1}{\eta(1-\eta)}$
$\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}$ (4.14)	$\mathcal{F}_{\mathrm{as}}^{\mathrm{CS}} = \frac{1}{1-\eta^2}$	$\mathcal{F}_{\mathrm{as}}^{\mathrm{CS}} = \frac{3}{(1-\eta)(1+3\eta)}$	$\mathcal{F}_{\mathrm{as}}^{\mathrm{CS}} = \frac{1}{\eta(1-\eta)}$	$\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}} = \frac{1}{\eta(1-\eta)}$

TABLE 4.4: **Channel QFIs and asymptotic bounds in decoherence-strength estimation scenarios** for the noise-models introduced in Sec. 2.2.3—the ones of Tab. 2.1 with $\Lambda_\eta = \mathcal{D}_\eta$ after setting $\varphi = 0$—. Owing to the different nature of the estimated parameter, the geometrical CS method is enough to provide the tightest asymptotic SQL-like bounds on precision for all but the *spontaneous emission* map, which being *extremal*—and thus η-extremal—allows only for the CE bound (4.24) to be applicable. In *all* the cases, as $\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}} = \mathcal{F}[\Lambda_\eta \otimes \mathcal{I}]$, the asymptotic bounds (4.14) are saturable *classically* by employing *extended* channels. Moreover, only when considering the *depolarisation* map there is room for quantum enhancement in the scheme of Fig. 4.1, as in all other cases also $\mathcal{F}[\Lambda_\eta] = \mathcal{F}[\Lambda_\eta \otimes \mathcal{I}]$, so that $\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}$ may be attained classically without need of extending each channel.

can be geometrically interpreted in the picture of Fig. 2.3 as a "movement" in the direction away from the boundary, i.e. *inside* the space of all relevant quantum channels, so that unless such a trajectory coincides with another non-flat boundary face (as in the case of spontaneous emission map that is *extremal*) the channel must be trivially η-*non-extremal* for any $\eta < 1$, as depicted in Fig. 2.3(**b**). Yet, the "speed" along such a trajectory cannot be naively amplified N-times by just employing N channels in parallel[18], so that in contrast to the noiseless phase estimation scenario the optimal entangled inputs must *not* lead to a scaling but at most to a *constant-factor* quantum enhancement, which, however, can be quantified by the methods of Sec. 4.3.1.

Moreover, due to the above interpretation it should not be surprising that for the noise-models: *dephasing*, *depolarisation* and *loss* the purely geometrical notion of *classical simulability* is enough to bound most tightly the asymptotic precision of estimation, so that $\mathcal{F}_{\mathrm{as}}^{\mathrm{CS}} = \mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} = \mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}] = \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$. Remarkably, in all three cases the distances from the boundary of the channels set are generally not symmetric, i.e. $\varepsilon_+ \neq \varepsilon_-$, and vary with the decoherence-strength parameter. In the CS picture of Fig. 4.4, we obtain for *dephasing*: $\varepsilon_\pm = 1 \pm \eta$; for *depolarisation*: $\varepsilon_+ = \eta + \frac{1}{3}$, $\varepsilon_- = 1 - \eta$; and for *loss*: $\varepsilon_+ = \eta$, $\varepsilon_- = 1 - \eta$; so that the corresponding CS bounds (4.17), $\mathcal{F}_{\mathrm{as}}^{\mathrm{CS}} = 1/(\varepsilon_+ \varepsilon_-)$, explicitly depend on η. On the other hand, the *spontaneous emission* map is special as it is extremal (and thus η-*extremal* for all η), so that both CS and RLD methods fail. Hence, we approach it by means of the CE method, for which it turns out that the necessary $\beta_{\tilde{K}} = 0$ condition (4.23) is only satisfied trivially by setting $\mathsf{h} = \mathbf{0}$. Therefore, the CE bound (4.24) may be constructed, but with no room for h-optimisation in Eq. (4.24), what also constrains $\alpha_{\tilde{K}} \not\propto \mathbb{I}$ disallowing the QS method to be applicable. The results are summarised in Tab. 4.4 together with (unextended-) (4.1) and extended- (4.7) channel QFIs for each of the noise-models (evaluated now w.r.t. the decoherence-strength parameter in contrast to phase estimation analysis of Tab. 4.1).

Importantly, for *all* the noise-models considered the asymptotic bounds (4.14) coincide with the corresponding extended-channel QFIs (4.7), i.e. $\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}} = \mathcal{F}[\Lambda_\eta \otimes \mathcal{I}]$, so that the maximal quantum enhancement of precision (4.13) in each case must be attainable by a *classical* estimation strategy that employs *extended* channels. The fact that the product input states—uncorrelated in between the extended N

---

[18]From another point of view, in case of the noiseless *unitary* parameter encoding, one may translate the parallel-channel scenario of Fig. 4.1 into a sequential one, what naturally explains the N-fold "speed" in parameter variations [Maccone, 2013]. Such a picture, however, fails when considering *decoherence-strength* estimation.

channels but possibly requiring entanglement between each single particle and its ancilla—are optimal for noise-strength estimation with extended channels, has already been noticed for the low-noise channels [Hotta *et al.*, 2006] and for generalised Pauli channels [Fujiwara and Imai, 2003], of which the latter contain the dephasing and depolarisation maps studied here.

Furthermore, in case of *dephasing*, *loss* and *spontaneous emission* maps, we realise that the corresponding CS bounds are actually attainable *classically* without need of the ancillae, as also the extension at the single-channel level is unnecessary, i.e. $\mathcal{F}[\Lambda_\eta] = \mathcal{F}[\Lambda_\eta \otimes \mathcal{I}]$. For *dephasing* noise, each single qubit must be optimally prepared in any pure state lying on the equator of the Bloch sphere, whereas while estimating the strength of *losses* the form of the input state (even mixed) is completely irrelevant. The latter fact explicitly proves that in the optical interferometry experiments sensing the power-transmission parameter of the interferometer arms the entanglement between the photons entering the interferometer is unnecessary and the total photon-number fluctuations are really the ones that limit the precision [Adesso *et al.*, 2009; Monras and Paris, 2007]. These may be reduced by employing Gaussian states [Monras and Paris, 2007] or in principle fully eliminated with use of any definite photon-number states that indeed attain the CS bound of Tab. 4.4 [Adesso *et al.*, 2009]. In order for the scheme to be most sensitive to variations of the strength—'decay rate'—of the *spontaneous emission*, one should indeed intuitively prepare all qubits in the 'excited', $|1\rangle$, state represented by the south pole in Fig. 2.2(**d**). As a result, one then achieves the coinciding unextended- and extended- channel QFIs of Tab. 4.4, which have been derived for the first time in [Fujiwara, 2004].

The example of *depolarisation* map is different, as it is known that for qubits [Frey *et al.*, 2011; Fujiwara, 2001] (see also Tab. 4.4), the precision of estimation may be improved by extending the channel, i.e. $\mathcal{F}[\Lambda_\eta] < \mathcal{F}[\Lambda_\eta \otimes \mathcal{I}] = \mathcal{F}_{\mathrm{as}}^{\mathrm{CS}}$ for $\eta < 1$. This leaves space for potential quantum enhancement, so that if the CS bound of Tab. 4.4 is tight then it may be asymptotically attained in the N-parallel–channels scheme of Fig. 4.1 only with use of entangled input particles. Consistently, this notion has been confirmed when considering two depolarisation channels acting in parallel, for which it is optimal to input maximally entangled two-qubit (Bell or equivalently $|\psi_{\mathrm{in}}^{N=2}\rangle_{\mathrm{GHZ}}$) states for $1/\sqrt{3} < \eta < 1$ and separable pure states otherwise [Fujiwara, 2001]. Thus, generalising such an observation, we consider the GHZ input states[19] (3.78) of arbitrary particle number, $|\psi_{\mathrm{in}}^N\rangle_{\mathrm{GHZ}}$, and utilise the results of [Simon and Kempe, 2002], in which the evolution of GHZ states undergoing uncorrelated depolarisation has been studied, in order to obtain the symbolic expression for the QFI (3.57):

$$F_{N,\eta}^{\mathrm{GHZ}} = F_{\mathrm{Q}}\big[\Lambda_\eta^{\otimes N}|\psi_{\mathrm{in}}^N\rangle_{\mathrm{GHZ}}\big] = \frac{N^2}{4}\frac{\eta^{2(N-1)}\alpha_{N-1}^2}{\alpha_N(\alpha_N^2 - \eta^{2N})} + \frac{2}{(1-\eta^2)^2}\sum_{k=0}^N \binom{N}{k}\frac{\big(N\beta_{N-k,k+1}^- - k\,\beta_{N-k,k}^-\big)^2}{\beta_{N-k,k}^+}, \quad (4.30)$$

where $\alpha_n = \left(\frac{1+\eta}{2}\right)^n + \left(\frac{1-\eta}{2}\right)^n$ and $\beta_{m,n}^\pm = \left(\frac{1+\eta}{2}\right)^m\left(\frac{1-\eta}{2}\right)^n \pm \left(\frac{1-\eta}{2}\right)^m\left(\frac{1+\eta}{2}\right)^n$. Unfortunately, Eq. (4.30) proves that $F_{N,\eta}^{\mathrm{GHZ}}/N$ is not monotonically rising with N and, in fact, $F_{N,\eta}^{\mathrm{GHZ}}/N \overset{N\to\infty}{=} \mathcal{F}[\Lambda_\varphi]$, so in the asymptotic N limit the GHZ input states do *not* lead to any quantum enhancement. Yet, for moderate particle numbers and $\eta > 1/\sqrt{3}$, as depicted in Fig. 4.6(**a**), the GHZ inputs outperform the classical strategies attaining their best precision per particle-number always for $N = 2$. Hence, in order to reach the maximal quantum enhancement (4.13) allowed by the GHZ-based strategies, it is optimal to group particles into $N/2$ pairs of Bell states, $|\psi_{\mathrm{in}}^{N=2}\rangle_{\mathrm{GHZ}}$, what then allows to attain $\chi_{\mathrm{GHZ}}[\Lambda_\eta]$ that is indeed greater than one if $\eta > 1/\sqrt{3}$ [Fujiwara, 2001], and reads:

$$\chi_{\mathrm{GHZ}}[\Lambda_\eta] = \sqrt{\frac{F_{N=2,\eta}^{\mathrm{GHZ}}/2}{\mathcal{F}[\Lambda_\eta]}} = \frac{6\eta^2}{1+3\eta^2} \qquad \leq \qquad \sqrt{\frac{\mathcal{F}_{\mathrm{as}}^{\mathrm{CS}}}{\mathcal{F}[\Lambda_\eta]}} = \sqrt{1 + \frac{2}{1+3\eta}}, \qquad (4.31)$$

---

[19]As the depolarisation noise does not preserve *bosonicity* of the particles, we must assume them to be distinguishable from the beginning, so we avoid naming the state (3.78) as NOON, reserving such name to the modal description.
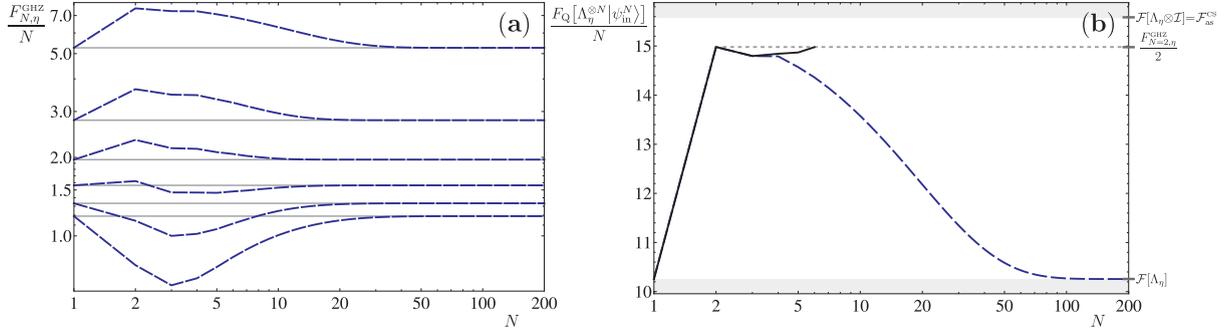
FIGURE 4.6: N-**parallel–channels depolarisation-strength estimation with GHZ inputs.**
(**a**) *QFI per particle* attained by the GHZ inputs (*blue dashed lines*) for depolarisation strengths: $\eta = \{0.4, 0.5, \dots, 0.9\}$, from bottom to top. (*Gray horizontal lines*) – maximal precision achieved by classical strategies, i.e. $\mathcal{F}[\Lambda_\eta]$ for each η. GHZ-based inputs outperform classical strategies provably only for moderate N and $\eta > 1/\sqrt{3} \approx 0.58$.
(**b**) *QFI per particle* attained by the GHZ inputs (*blue dashed line*) and optimal inputs (*solid black line*) for $\eta = 0.95$. Results indicate that when considering GHZ-based strategies, it is optimal to employ uncorrelated *pairs* of particles as inputs with each doublet in $|\psi_{\text{in}}^{N=2}\rangle_{\text{GHZ}}$ (*horizontal grey dotted line*). *Shaded regions* represent 'worse than classical' (*bottom*) and 'better than the CS bound' (*top*) regimes.

where we have adequately upper-limited $\chi_{\text{GHZ}}[\Lambda_\eta]$ with help of the the CS bound of Tab. 4.4. The gap between the two is clearly shown in Fig. 4.6(**b**) corresponding to the spacing between the horizontal *dotted grey line* and the *upper shaded region*. Notice that despite very low depolarisation-strength, $\eta = 0.95$, the GHZ input states (*blue dashed line*) rapidly lose their quantum advantage attaining the classically achievable regime (*lower shaded region*) already for $N \approx 100$. In order to benchmark their performance, we perform brute-force numerical optimisation over the input states, what we are only able to do for $N \leq 6$ (*solid black line*). Yet, we already observe significant improvement over the GHZ states, what indicates their sub-optimality and suggests the potential of attaining the CS-based bound on quantum enhancement of Eq. (4.31). Nevertheless, being beyond the scope of this work, we leave the problem of finding the optimal entangled input states attaining the maximal precision for future research.

Chapter 5

Mach-Zehnder interferometry with photonic losses

5.1   Lossy Mach-Zehnder interferometer

In the last chapter of this work, we revisit again the Mach-Zehder interferometer setup previously discussed at the classical and quantum levels in Secs. 3.1.4 and 3.2.4 respectively. Yet, this time, we thoroughly analyse the corresponding phase estimation problem but accounting for the *photonic losses* that lead to potentially different power-transmission coefficients, $\eta_{a/b}$, of the interferometer arms labelled by a/b. As depicted in Fig. 5.1, such a process is generally modelled by introducing fictitious beam-splitters of transmittances $\eta_{a/b}$ with vacuum states, $|0\rangle$, impinged on their auxiliary input ports. Such a loss model is relatively general, as due to the commutativity of the noise with the phase accumulation [Demkowicz-Dobrzański *et al.*, 2009], it accounts for photonic losses occurring at any time during the phase-sensing stage. Moreover, any losses taking place during the preparation and detection stages, provided they are equal in both arms, are also included, as these may be commuted through the input and output beam-splitters pictured in Fig. 5.1, and thus accommodated into the transmittances of the fictitious ones. As a result, the above description finds its applicability in typical quantum-enhanced interferometry experiments [Kacprowicz *et al.*, 2010; Spagnolo *et al.*, 2012; Vitelli *et al.*, 2010], and most notably (as shown by Demkowicz-Dobrzański *et al.* [2013]), when describing the gravitational-wave detectors [LIGO Collaboration, 2011, 2013]. Notice that when following the notation of Sec. 2.1.4 and treating each photon as a distinguishable particle in an overall permutation invariant state, then in the special case of equal losses, $\eta_a = \eta_b$, the channel representing the single-photon evolution through the interferometer of Fig. 5.1 is exactly the noisy-phase–estimation *loss map* illustrated in Fig. 2.2(**c**).

As before when analysing the Mach-Zehnder interferometer in the quantum setting in Sec. 3.2.4, we consider as the input a general pure, N-photon, two-mode (arm) state (3.75), which we write again for convenience:

$$|\psi_{\text{in}}^N\rangle = \sum_{n=0}^{N} \alpha_n |n\rangle_a |N-n\rangle_b = \sum_{n=0}^{N} \alpha_n |n, N-n\rangle. \tag{5.1}$$

Then, the output state of Fig. 5.1 most generally reads

$$\rho_\varphi^N = \sum_{l_a=0}^{N} \sum_{l_b=0}^{N-l_a} p_{l_a,l_b} |\xi_{l_a,l_b}(\varphi)\rangle\langle\xi_{l_a,l_b}(\varphi)| = \bigoplus_{N'=0}^{N} \sum_{l_a=0}^{N-N'} p_{l_a,N-N'-l_a} |\xi_{l_a,N-N'-l_a}(\varphi)\rangle\langle\xi_{l_a,N-N'-l_a}(\varphi)|, \tag{5.2}$$
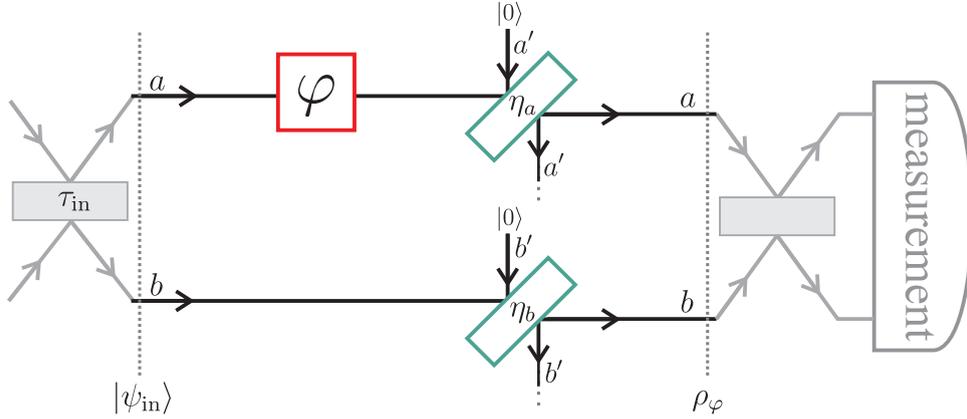
FIGURE 5.1: **Lossy interferometer** corresponding to the generalisation of the Mach-Zehnder of Fig. 3.7 that accounts for photonic losses leading to non-perfect power transmittances $\eta_{a/b}$ of the interferometer arms a/b. As in Sec. 3.2.4, an optimal N-photon input state (3.76), $|\psi_{\rm in}\rangle = |\psi_{\rm in}^N\rangle$, is sought that despite losses leads to the maximal precision of phase estimation. Its performance is compared against a *classical strategy*, in which a coherent state $|\alpha\rangle$ of light with mean photon-number $|\alpha|^2 = N$ is impinged on the input beam-splitter, so that then $|\psi_{\rm in}\rangle = |\sqrt{\tau_{\rm in}}\,\alpha\rangle_a \,|\sqrt{1-\tau_{\rm in}}\,\alpha\rangle_b$.

The extra labelling of the environmental modes a' and b' is introduced to match the later notation of Fig. 5.4.

where p_{l_a,l_b} is the φ-independent, binomially distributed probability of losing l_a and l_b photons in arms a and b respectively, so that $p_{l_a,l_b} = \sum_{n=0}^{N} |\alpha_n|^2 \, b_n^{(l_a,l_b)}$ with

$$b_n^{(l_a,l_b)} = \begin{cases} \binom{n}{l_a} \eta_a^{n-l_a} \left(1-\eta_a\right)^{l_a} \binom{N-n}{l_b} \eta_b^{N-n-l_b} \left(1-\eta_b\right)^{l_b} & , \; l_a \leq n \leq N - l_b \\ 0 & , \; \text{otherwise} \end{cases} . \tag{5.3}$$

The direct sum in the second expression of Eq. (5.2) indicates that the output states of different total number of surviving photons, N', belong to orthogonal subspaces, which in principle could be distinguished by a non-demolition, photon-number–counting measurement. The pure states which are outputted according to p_{l_a,l_b} read[1]:

$$|\xi_{l_a,l_b}(\varphi)\rangle = \frac{1}{\sqrt{p_{l_a,l_b}}} \sum_{n=0}^{N} \alpha_n \, e^{-in\varphi} \sqrt{b_n^{(l_a,l_b)}} \, |n - l_a, N - n - l_b\rangle \tag{5.4}$$

and obey $\langle \xi_{l_a,l_b} | \xi_{l'_a,l'_b} \rangle = \delta_{l_a+l_b, l'_a+l'_b}$, so that indeed no coherences exist between their versions for various total numbers of surviving photons: $N' = N - l_a - l_b$.

Motivated by typical optical implementations [Hariharan, 2003], in order to benchmark the quantum enhancement of precision attained by the N-photon input states (5.1), we do *not* compare them directly with the *classical input states* (3.76) of uncorrelated photons (previously considered in Sec. 3.1.4), but rather with the natural optical strategy employing a *coherent state* of light $|\alpha\rangle$ that despite not possessing a definite number of photons still has them all uncorrelated with one another. Although the state $|\alpha\rangle$ in principle leads to the input $\left|\psi_{\rm in}^{|\alpha\rangle}\right\rangle = \left|\sqrt{\tau_{\rm in}}\,\alpha\right\rangle_a \left|\sqrt{1-\tau_{\rm in}}\,\alpha\right\rangle_b$, when split on the input beam-splitter of transmittance $\tau_{\rm in}$, one should not forget that due to the lack of global phase-reference $\left|\psi_{\rm in}^{|\alpha\rangle}\right\rangle$ should also be phase-averaged [Bartlett *et al.*, 2007; Jarzyna and Demkowicz-Dobrzański, 2012; Mølmer, 1997].

---

[1] The phase delay factor $e^{-in\varphi}$ in Eq. (5.4) naturally arises if the losses occur *after* the phase accumulation, as drawn in Fig. 5.1. However, as only the relative phase delay in between the modes a and b possesses physical significance, all the later calculations in this chapter can be equivalently performed having the order of the phase delay and losses reversed.

Notice that the presence the global phase-reference would need to be accounted for by introducing a third mode of light[2] containing photons that would have to be also counted then as a part of the total resources N. As such an an extra beam cannot be assumed, the actual input state must be really written as:

$$\rho_{\text{in}}^{|\alpha\rangle} = \bigoplus_{N=0}^{\infty} p_N \left|\xi_{|\alpha\rangle}^N\right\rangle\!\left\langle\xi_{|\alpha\rangle}^N\right|, \tag{5.5}$$

representing a mixture of N-photonic states that are Poissonian distributed according to $p_N = e^{-\bar{N}}\frac{\bar{N}^N}{N!}$ and read:

$$\left|\xi_{|\alpha\rangle}^N\right\rangle = \sum_{n=0}^{N} \sqrt{\binom{N}{n}\kappa_a^n\,\kappa_b^{N-n}}\,\left|n, N-n\right\rangle, \tag{5.6}$$

where $\bar{N} = |\alpha|^2$ is the mean number of photons present in the interferometer that are split between the arms a and b in fractions $\kappa_a = \tau_{\text{in}}$ and $\kappa_b = 1 - \tau_{\text{in}}$ respectively. Importantly, notice that each state (5.6) is exactly the generalisation of the previously considered classical N-photon input (3.76), which would be obtained after setting $\tau_{\text{in}} = \kappa_a = \kappa_b = \frac{1}{2}$. Hence, the photons are indeed uncorrelated with one another with each of them being in the state $\sqrt{\kappa_a}\,|a\rangle + \sqrt{\kappa_b}\,|b\rangle$ that in comparison to Eq. (3.76) differs only due to the presence of the weights $\kappa_{a/b}$, which are additionally introduced to compensate for the potential unequal losses in the interferometer arms, i.e. $\eta_a \neq \eta_b$ in Fig. 5.1. However, in order to adequately compare the precision attained by $\left|\psi_{\text{in}}^N\right\rangle$ and $\rho_{\text{in}}^{|\alpha\rangle}$, we must take the mean number of photons in the state (5.5), $\bar{N}$, to equal the exact photon number of (5.1), N, so that $|\alpha|^2 = N$. On the other hand, as the global-phase–averaging may be equivalently performed after propagating the coherent state through the interferometer of Fig. 5.1, the form of the output state for the input (5.5) may be easily obtained:

$$\rho_{\varphi}^{|\alpha\rangle} = \bigoplus_{N=0}^{\infty} p_N \left|\xi_{|\alpha\rangle}^N(\varphi)\right\rangle\!\left\langle\xi_{|\alpha\rangle}^N(\varphi)\right|, \tag{5.7}$$

where $\left|\xi_{|\alpha\rangle}^N(\varphi)\right\rangle$ is the straightforward generalization of Eq. (5.6) with: extra phase-factor $e^{-in\varphi}$ introduced, modified coefficients $\kappa_{a/b}$ that now due to losses read $\eta_a\tau_{\text{in}}$ and $\eta_b(1 - \tau_{\text{in}})$ respectively, and the overall average photon number diminished to $|\alpha|^2 \left[\eta_a\tau_{\text{in}} + \eta_b(1 - \tau_{\text{in}})\right]$.

5.2   Frequentist approach – local phase estimation

The lossy interferometric setup of Fig. 5.1 has been explicitly studied within the *frequentist* approach in [Demkowicz-Dobrzański *et al.*, 2009; Dorner *et al.*, 2009], where the corresponding form of the QFI (3.57) with respect to the estimated phase φ for the output state (5.2), ρ_{φ}^N, has been investigated. In the most general case of unequal losses ($\eta_a \neq \eta_b$) an upper bound on the QFI has been proposed which directly follows from the convexity of the QFI (see Sec. 3.2.2.3):

$$F_Q\left[\rho_{\varphi}^N\right] \le \bar{F}_Q\left[\rho_{\varphi}^N\right] = \sum_{l_a=0}^{N}\sum_{l_b=0}^{N-l_a} p_{l_a,l_b}\,F_Q\left[|\xi_{l_a,l_b}(\varphi)\rangle\right], \tag{5.8}$$

where

$$\bar{F}_Q\left[\rho_{\varphi}^N\right] = 2\sum_{l_a=0}^{N}\sum_{l_b=0}^{N-l_a} \frac{\mathbf{x}^T\mathbf{R}^{(l_a,l_b)}\mathbf{x}}{\mathbf{x}^T\mathbf{b}^{(l_a,l_b)}} \tag{5.9}$$

---

[2]One may interpret then such a setup as an "interferometer inside another interferometer", what clearly overcomplicates and unnecessarily modifies the phase estimation scenario.

with $\mathbf{x}$ and $\mathbf{b}^{(l_a,l_b)}$ being vectors containing variables $x_n = |\alpha_n|^2$ and $b_n^{(l_a,l_b)}$ respectively. The elements of the matrix $\mathbf{R}^{(l_a,l_b)}$ read $R_{nm}^{(l_a,l_b)} = b_n^{(l_a,l_b)}(n-m)^2 b_m^{(l_a,l_b)}$. Intuitively, the coefficients $b_n^{(l_a,l_b)}$ arise from the binomial distributions dictating the probability of l_a and l_b photons being lost in either of the interferometer arms for every $|n, N-n\rangle$ constituent of the N-photon input state (5.1). Although inequality (5.8) is not generally tight, for the case of equal losses ($\eta = \eta_a = \eta_b$) in both arms $\bar{F}_Q$ approximates the actual QFI within a negligible error margin [Demkowicz-Dobrzański *et al.*, 2009]. Yet, when losses occur in a single arm ($\eta = \eta_a$, $\eta_b = 1$), only the direct sum in Eq. (5.2) is present, and hence $F_Q[\rho_\varphi^N] = \bar{F}_Q[\rho_\varphi^N]$, so that Eq. (5.9) becomes the expression for the QFI.

On the other hand, one may in principle construct the optimal POVM saturating the real QFI (5.8) by the general recipe of Sec. 3.2.2.1, which for ϱ_φ^N (5.2) leads to a non-demolition measurement—allowing to learn the number of surviving photons[3]—followed by a projection onto the eigenvectors of the corresponding SLD (3.56). Yet, such a measurement is not very practical, as due to locality of the approach its form generally depends on the particular value of the estimated phase φ [Demkowicz-Dobrzański *et al.*, 2009]. It has been also proved in [Demkowicz-Dobrzański *et al.*, 2009] that $\bar{F}_Q$ (5.9) cannot increase by allowing the N photons in the input state (5.1) to be *distinguishable* (see Sec. 2.1.4 and Eq. (2.8)), e.g. by sending them in non-overlapping time bins. In fact, the knowledge of which photon was lost additionally harms the quantum superposition, whereas the distinguishability does not provide any advantage in terms of phase sensitivity. On the other hand, the ability to target each photon individually allows for the adaptive measurement schemes [Wiseman *et al.*, 2009; Wiseman and Killip, 1997, 1998], which being φ-independent turn out to be much more experimentally approachable. Nevertheless, from the point of view of the QFI it is optimal to restrict only to the bosonic input states of the form (5.1), as there always exist a measurement scheme, e.g. the one described above, that attains the QCRB without utilising the adaptivity.

Classical input states

In case of the coherent-state–based *classical strategy*, one may directly calculate the QFI of the output state $\rho_\varphi^{|\alpha\rangle}$ due the direct-sum structure in Eq. (5.7): $F_Q[\rho_\varphi^{|\alpha\rangle}] = \sum_N p_N F_Q[|\xi_{|\alpha\rangle}^N(\varphi)\rangle]$, and after maximising it over τ_{in} construct the minimal QCRB (3.55) on the achievable precision, which correctly exhibits the SQL-like–scaling behaviour [Demkowicz-Dobrzański *et al.*, 2009]:

$$\Delta^2\tilde{\varphi}_{cl} \geq \frac{1}{4}\left(\frac{1}{\sqrt{\eta_a}} + \frac{1}{\sqrt{\eta_b}}\right)^2 \frac{1}{N}, \tag{5.10}$$

where we have adequately fixed $|\alpha|^2 = N$, so that Eq. (5.10) may be later compared with $\Delta^2\tilde{\varphi}_Q$ attained by the optimal N-photon input states (5.1). Notably, Eq. (5.10) is obtained after setting the input beam-splitter transmittance to $\tau_{in} = 1/(1+\sqrt{\eta_a/\eta_b})$, what may be intuitively explained, as such a choice yields the intensities of light-beams in arms a and b described by the output state (5.7) to be equal, and thus leads to the highest *visibility* of the interferometer.

5.2.1 Numerical solution for moderate N

As $\bar{F}_Q$ in Eq. (5.9) is a concave function with respect to the input-state coefficients $x_n = |\alpha_n|^2$ [Demkowicz-Dobrzański *et al.*, 2009], one may efficiently seek numerically for the optimal input states maximizing $\bar{F}_Q$ for moderate values of N (≤ 100). An example of an optimal $|\alpha_n|^2$-distribution as a function of η is illustrated in Fig. 5.2(**a**), where we have assumed single-arm losses ($\eta = \eta_a$, $\eta_b = 1$) and set the

---

[3]What in real-life experiments is typically achieved by post-selection.
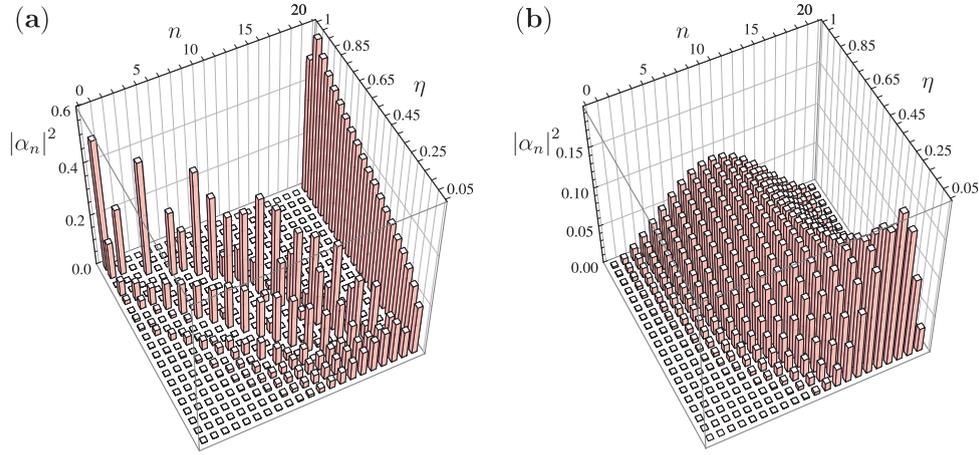
FIGURE 5.2: **Coefficients, $|\alpha_n|^2$, of the optimal N-photon input states** (5.1), $\left|\psi_{\text{in}}^N\right\rangle$, plotted as a function of the transmittance coefficient η of the arm a, which is the only one subjected to photonic losses ($N = 20$). As η decreases, more weight must be appointed to coefficients in the mid-range of n-s to increase the robustness of $\left|\psi_{\text{in}}^N\right\rangle$ against losses. Due to the asymmetry of noise a bias towards higher values of n appears, which does not occur for the equal-losses scenario. Such overall, qualitative behaviour is observed both when the *frequentist* (**a**) and *Bayesian* (**b**) approaches are considered, despite the contrasting structures of $|\alpha_n|$-distributions. In the Bayesian case (**b**), the lack of prior knowledge effectively "smooths out" the optimal distribution. At $\eta = 1$, the numerical results correctly reproduce the noiseless solutions of Sec. 3.2.4: for (**a**) the NOON states (3.78), for (**b**) the BW states (3.86).

photon number to $N = 20$. The numerical results of Fig. 5.2(**a**) indicate that although in accordance with Sec. 3.2.4.1 the *NOON input states* (3.78) are optimal in the noiseless scenario ($\eta = 1$), they rapidly cease to be efficient in the presence of noise, as other coefficients than $|\alpha_0|$ and $|\alpha_N|$ must be gradually introduced with increase of losses, in order to improve the robustness of the input. The fact that NOON states are extremely fragile should not be surprising, bearing in mind that a loss of a single photon is enough for them to erase all the information about the estimated phase. Mathematically, such behaviour is manifested by their QFI being determined solely by the subspace $N' = N$ in Eq. (5.2), so that when utilising Eq. (5.8) only the non-zero term with $p_{l_a=0, l_b=0}$ contributes and[4]:

$$F_{\text{Q}}\big[\rho_\varphi^{\text{NOON}}\big] = p_{l_a=0, l_b=0}^{\text{NOON}} \, F_{\text{Q}}\big[\big|\xi_{l_a=0, l_b=0}^{\text{NOON}}(\varphi)\big\rangle\big] = 2 \, \frac{(\eta_a \eta_b)^N}{\eta_a^N + \eta_b^N} \, N^2, \tag{5.11}$$

what is consistent with the expression already noted in Sec. 4.3.3 for equal-losses ($\eta_a = \eta_b = \eta$): $\mathcal{F}_N^{\text{NOON}} = \eta^N N^2$. Importantly, Eq. (5.11) shows that, although the N^2 term (which previously lead to the HL-like scaling in Eq. (3.79)) is preserved, the losses introduce an exponentially decaying factor that yields vanishing QFI and thus divergent precision in the asymptotic N limit. In other words, the losses degrade exponentially in N the probability of *not* losing any photon in a NOON state, i.e. $p_{l_a=0, l_b=0}^{\text{NOON}} = (\eta_a^N + \eta_b^N)/2$, so that no information about the phase may be retrieved for sufficiently large N.

Crucially, as the numerical optimisation over the input states in Eq. (5.8) may be performed only for moderate N, one may not be sure about answering the question whether there exist inputs that despite losses lead to asymptotic precision scaling beyond the SQL. Yet, we have shown explicitly in Sec. 4.3.1.4

---

[4]For unequal-losses ($\eta_a \neq \eta_b$) scenarios, one may further improve Eq. (5.11) by considering unbalanced NOON input states: $\alpha_N|N, 0\rangle + \alpha_0|0, N\rangle$, for which it is optimal to set $\alpha_0 = \sqrt{\eta_a^{N/2}/(\eta_a^{N/2} + \eta_b^{N/2})}$ what yields $p_{l_a=0, l_b=0} = (\eta_a \eta_b)^{N/2}$ and $F_{\text{Q}} = 4 \frac{(\eta_a \eta_b)^N}{\left(\eta_a^{N/2} + \eta_b^{N/2}\right)^2} N^2$ [Demkowicz-Dobrzański *et al.*, 2009].

that it cannot be so in the equal-losses scenario, for which the QS bound (4.21) is sufficient to prove the asymptotic SQL-like scaling of precision. Such an observation actually proves that for any $\eta_a < \eta_b < 1$ in Fig. 5.1 the asymptotic precision scaling must also be SQL-like, as by increasing $\eta_a \rightarrow \eta_b$ and thus decreasing the overall noise, we may always construct an equal-losses scenario which is known to be SQL-bounded and may perform only better. Such an argument, however, does not hold when considering single-arm losses, but we show explicitly in the following section that the techniques of Sec. 4.3.1 are then also applicable, and in fact yield asymptotic SQL-like bounds for any $\eta_{a/b}$ in Fig. 5.1 (with the only exception being naturally the noiseless scenario ($\eta_a = \eta_b = 1$) discussed in Sec. 3.2.4.1 that attains the HL).

5.2.2 Asymptotic SQL-like bound on precision

In the following paragraphs we show that the *QS method* of Sec. 4.3.1.2 is sufficient to determine an asymptotic SQL-like bound on precision not only for equal losses (as already discussed in Sec. 4.3.1.4), but also in the general lossy interferometry setting of Fig. 5.1. Furthermore, we demonstrate that such a bound may be independently derived by utilising the concept of *minimisation over purifications* introduced in Sec. 3.2.2.2, but returning to the *modal*- rather than *particle*-picture of the photonic state employed as an input. Nevertheless, both approaches yield the same upper bound on precision that, however, has been just recently proved to be tight [Knysh *et al.*, 2014], i.e. to be exactly the QCRB (3.55) calculated for the optimal N-photon inputs (5.1) in the asymptotic N limit:

$$\Delta^2 \tilde{\varphi}_{\mathrm{Q}} \geq \frac{1}{4} \left(\sqrt{\frac{1 - \eta_a}{\eta_a}} + \sqrt{\frac{1 - \eta_b}{\eta_b}} \right)^2 \frac{1}{N}, \tag{5.12}$$

which thus also quantifies the *maximal quantum enhancement of precision* (4.13) corresponding to the ratio of Eqs. (5.10) and (5.12):

$$\chi\!\left[\Lambda_\varphi^{(\eta_a, \eta_b)}\right] = \lim_{N \to \infty} \sqrt{\frac{\Delta^2 \tilde{\varphi}_{\mathrm{cl}}}{\Delta^2 \tilde{\varphi}_{\mathrm{Q}}}} = \frac{\sqrt{\eta_a} + \sqrt{\eta_b}}{\sqrt{(1 - \eta_a)\,\eta_b} + \sqrt{(1 - \eta_b)\,\eta_a}}, \tag{5.13}$$

where by $\Lambda_\varphi^{(\eta_a, \eta_b)}$ we label the *lossy interferometer channel*—the effective quantum map describing the evolution of each photon while propagating through the interferometer of Fig. 5.1. Notice that, for *equal losses* ($\eta = \eta_a = \eta_b$), Eq. (5.13) correctly coincides with the expression: $\sqrt{1/(1 - \eta)}$, previously stated in Tab. 4.2, whereas for *single-arm losses* it simplifies to $\sqrt{\left(1 + \sqrt{\eta}\right) / \left(1 - \sqrt{\eta}\right)}$ originally derived in [Knysh *et al.*, 2011].

In Fig. 5.3(**a**), we present a plot of the maximal achievable precision in the lossy interferometer with *single-arm* losses ($\eta_a = \eta$, $\eta_b = 1$) after choosing $\eta = 0.7$. The HL attained in the *noiseless* setting by the NOON states (3.78) (*long-dashed black line*) is shown for comparison. The presence of losses forces the asymptotic scaling to be SQL-like and follow the ultimate precision dictated by the bound (5.12) (*solid grey line*). The optimal N-photon input states (5.1) with coefficients distributed according to Fig. 5.2(**a**) approach the bound, yet we are restricted by the numerical capabilities of maximising the QFI (5.8), $F_{\mathrm{Q}}\!\left[\rho_\varphi^N\right] = \bar{F}_{\mathrm{Q}}\!\left[\rho_\varphi^N\right]$, to relatively low photon numbers ($N \lesssim 150$). Nevertheless, apart from the explicit calculation of the asymptotic form of the $F_{\mathrm{Q}}\!\left[\rho_\varphi^N\right]$ in [Knysh *et al.*, 2014], the bound (5.12) has been shown to be attainable by utilising particular indefinite-photon–number input states and measurements, e.g. inputs employing light-squeezing accompanied by the photon-counting detection [Caves, 1981; Demkowicz-Dobrzański *et al.*, 2015]. For completeness, the performance of coherent-state–based *classical strategy* is also plotted (*short-dashed blue line*) that is fully determined by Eq. (5.10).
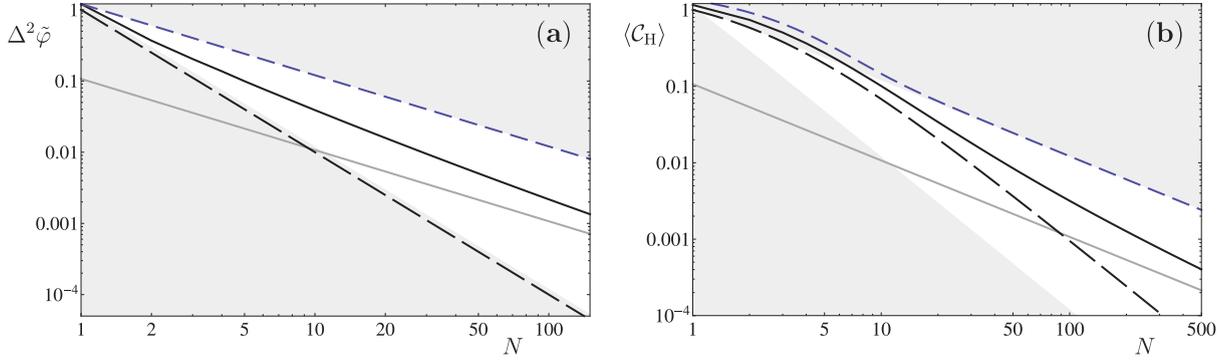
FIGURE 5.3: **Precision of phase estimation for a lossy interferometer** when analysed from the *frequentist* (**a**) and *Bayesian* (**b**) perspectives. For a single-arm losses scenario, i.e. $\eta_a = 0.7$, $\eta_b = 1$, the precision attained by the optimal N-photon input states of Fig. 5.2 (*solid black lines*) and the coherent-state–based *classical strategy* (*short-dashed blue lines*) is depicted. Within both approaches the ultimate precision is dictated by the bound (5.12) (*solid grey lines*). For comparison, the performance of the optimal N-photon strategy in the *noiseless* scenario ($\eta_a = \eta_b = 1$), i.e. NOON states (3.78) in (**a**) and BW states (3.86) in (**b**), is plotted (*long-dashed black lines*) that in both cases attains the HL.

Lastly, let us emphasise that due to the lack of global-phase reference any SQL-like bound, e.g. the one of Eq. (5.12), must also apply to *indefinite-photon–number input states* of light with the mean number of photons, $\bar{N}$, fixed. Any such bound is derived basing on the previously discussed upper limit (4.14) on the QFI that scales linearly with the number of particles involved, i.e. $F_Q[\rho_\varphi^N] \leq N \mathcal{F}_{\text{as}}^{\text{bound}}$. On the other hand, due to global-phase averaging, any indefinite-particle(photon)–number input takes the form: $\sum_N p_N \rho_{\text{in}}^N$ with $\sum_N p_N = \bar{N}$. Thus, the output state may always be written as $\sum_N p_N \rho_\varphi^N$, and its QFI may always be upper-bounded stemming from the convexity property of the QFI (see Sec. 3.2.2.3) as:

$$F_Q\left[\sum_N p_N \rho_\varphi^N\right] \leq \sum_N p_N F_Q\left[\rho_\varphi^N\right] \leq \sum_N p_N N \mathcal{F}_{\text{as}}^{\text{bound}} = \bar{N} \mathcal{F}_{\text{as}}^{\text{bound}}. \tag{5.14}$$

Hence, most generally, for any input that we may regard as an incoherent mixture of states occupying different particle-number sectors, we may apply all the SQL-like bounding techniques discussed in Sec. 4.3.1 after accordingly replacing the definite particle number, N, with the input mean particle number, $\bar{N}$, in all the eventually obtained bounds on precision[5].

Quantum Simulation of the lossy interferometer channel

Let us show that the precision bound (5.12) may be derived by means of the *QS method* discussed in Sec. 4.3.1.2, which previously turned to be already sufficient when considering the scenario of equal losses (see Tab. 4.2), i.e. for the loss noise model depicted in Fig. 2.2(**c**). In order to apply the QS bound (4.21) to the more general setting of the interferometer in Fig. 5.1, we establish the form of the corresponding channel $\Lambda_\varphi^{(\eta_a,\eta_b)}$ acting on each photon within the input state (5.1), so that the output state (5.2) may be then written as $\rho_\varphi^N = \Lambda_\varphi^{(\eta_a,\eta_b)\otimes N}[|\psi_{\text{in}}^N\rangle]$ allowing us to regard the interferometer of Fig. 5.1 as an instance of the *N-parallel–channels estimation scheme* of Fig. 4.1. Fortunately, this may be easily achieved by just trivially generalising the canonical Kraus operators specified for the equal-loss model described in

---

[5]Yet, notice that the argumentation of Eq. (5.14) may not be applied when considering the *finite-N CE bound* of Sec. 4.3.2 that does *not* yield a linear, i.e. concave in N, upper limit on the QFI. As a result, the finite-N CE bound (4.25) may only be confidently applied when dealing with systems consisting of a definite number of particles.

Tab. 2.1, so that they now read:

$$K_1 = \begin{pmatrix} \sqrt{\eta_a} & 0 \\ 0 & \sqrt{\eta_b} \\ 0 & 0 \end{pmatrix}, \; K_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \sqrt{1-\eta_a} & 0 \end{pmatrix}, \; K_3 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \sqrt{1-\eta_b} \end{pmatrix}, \quad (5.15)$$

and thus account for different transmittances of the arms a and b. As before, the overall action of the single-photon map may be written as $\Lambda_\varphi^{(\eta_a,\eta_b)}[\bullet] = \sum_{i=1}^3 K_i(\varphi) \bullet K_i(\varphi)^\dagger$, after setting all $K_i(\varphi) = K_i U_\varphi$ in accordance with Fig. 5.1, but—as the losses and phase-accumulation commute with one another—we may have equivalently chosen $K_i(\varphi) = \tilde{U}_\varphi K_i$ with $\tilde{U}_\varphi$ representing an enlarged $U_\varphi = \mathrm{e}^{-\mathrm{i}\frac{\varphi}{2}\hat\sigma_z}$ that also trivially acts on the vacuum state $|0\rangle$. Notice that, in order to emphasise the symmetry of the problem and match the previous notation of noisy-phase–estimation models of Fig. 2.2, we have without loss of generality assumed again the phase to accumulate as[6] $(\varphi/2,-\varphi/2)$ in the arms (a,b) of the interferometer, and *not* as $(\varphi,0)$, as originally drawn in Fig. 5.1.

Having established the Kraus representation of the lossy interferometer channel $\Lambda_\varphi^{(\eta_a,\eta_b)}$, we may directly compute[7] the *QS bound* (4.21) following the recipe of Sec. 4.3.1.2:

$$\mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} = \frac{4}{\left(\sqrt{\frac{1-\eta_a}{\eta_a}} + \sqrt{\frac{1-\eta_b}{\eta_b}}\right)^2} \quad (5.16)$$

corresponding to the minimum in Eq. (4.21) that occurs after optimally choosing

$$\mathsf{h}_{\mathrm{opt}} = -\frac{1}{8}\,\mathrm{diag}\left\{\chi, \frac{\eta_a}{1-\eta_a}\left(\frac{4}{\eta_a} - \chi\right), -\frac{\eta_b}{1-\eta_b}\left(\frac{4}{\eta_a} + \chi\right)\right\} \quad \text{with} \quad \chi = \mathcal{F}_{\mathrm{as}}^{\mathrm{QS}}\,\frac{\eta_a - \eta_b}{\eta_a\eta_b}. \quad (5.17)$$

Hence, we indeed recover the ultimate quantum limit (5.12) as $\Delta^2\tilde{\varphi}_\mathrm{Q} \geq 1/(\mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} N)$, which leads to the correct maximal quantum enhancement of precision (5.13). Consistently, for equal losses ($\eta = \eta_a = \eta_b$), the QS bound (5.16) coincides with the one quoted in Tab. 4.2 for the loss noise-model and the optimal generator (5.17) becomes then, as necessary, the adequate $\mathsf{h}_{\mathrm{opt}}$ (F.6) noted in App. F.

Optimal purification of the output state

Interestingly, we show that in the case of the lossy interferometer of Fig. 5.1 the ultimate bound on precision (5.12) may also be independently established after moving away from the N-parallel–channel picture of Fig. 4.1, in which the properties of the single-particle map are investigated to determine the SQL-like bounds on precision. We demonstrate that Eq. (5.12) may be equivalently derived by considering the overall N-particle output state, ρ_φ^N, and by directly utilising the *purification-based definition of its QFI* (3.63) introduced in Sec. 3.2.2.2. We follow the analysis of [Escher *et al.*, 2011] after adopting the *modal description* of light propagating through the lossy interferometer of Fig. 5.1, but in addition we explicitly make use of the so-called *Jordan-Schwinger* (JS) *map* [Jordan, 1935; Schwinger, 1965], what greatly facilitates the intuition behind the search for the optimal purification of the output. It allows us to benefit from the notion previously introduced for finite-dimensional spaces (purification-based methods introduced for states in Sec. 3.2.2.2 and for channels in Sec. 4.2.1) stating that, in order to establish the tightest upper bound on the output state QFI, one must choose a local rotation of the environmental subspace that erases most information about the estimated parameter potentially

---

[6]As aside, let us note that such a choice not only simplifies the calculations, but also, in case the output state ρ_φ is path(mode)-symmetric, assures the redundancy of a global-phase reference [Jarzyna and Demkowicz-Dobrzański, 2012].

[7]Actually, it is more efficient to compute the *CE bound* (4.24), what can be achieved with help of the SDP discussed in App. I, and then realise that the extra condition, $\alpha_{\tilde{K}_{\mathrm{opt}}} \propto \mathbb{I}$, of the QS method is also naturally fulfilled.

encoded in the environment. As shown below, such picture naturally generalises to the modal light-description, in case of which one must optimise over local rotations of the environmental *modes* in order to establish the best possible output purification, $|\tilde{\Psi}_\varphi\rangle$ satisfying $\rho_\varphi^N = \text{Tr}_{\text{E}}\{|\tilde{\Psi}_\varphi\rangle\langle\tilde{\Psi}_\varphi|\}$, that yields the tightest $F_{\text{Q}}[\rho_\varphi^N] \le F_{\text{Q}}[|\tilde{\Psi}_\varphi\rangle]$ in accordance with the purification-based definition (3.63).

Importantly, the JS map allows us to express the action of the interferometer of Fig. 5.1 in terms of the algebra of the angular momentum operators [Yurke *et al.*, 1986]. Denoting by $\hat{a}^\dagger$ and $\hat{a}$ respectively the creation and annihilation operators of the mode describing the light travelling in arm a, and similarly for arm b, we define the angular momentum operators:

$$\hat{J}_x^{ab} = \frac{1}{2}(\hat{a}^\dagger\hat{b} + \hat{b}^\dagger\hat{a}), \quad \hat{J}_y^{ab} = \frac{\text{i}}{2}(\hat{b}^\dagger\hat{a} - \hat{a}^\dagger\hat{b}), \quad \hat{J}_z^{ab} = \frac{1}{2}(\hat{a}^\dagger\hat{a} - \hat{b}^\dagger\hat{b}), \tag{5.18}$$

which fulfil the commutation relations[8]: $[\hat{J}_i^{ab}, \hat{J}_j^{ab}] = \text{i}\,\epsilon_{ijk}\hat{J}_k^{ab}$. As later we consider such operators acting on various pairs of modes, we explicitly label by a superscript the modes, e.g. m_1 and m_2, on which a given $\hat{J}_i^{m_1 m_2}$ is defined. Moreover, we may construct unitary operators generated by the angular momentum operators (5.18) that thus describe the rotations in the abstract spin space: $U_i^{m_1 m_2}(\theta) = \text{e}^{-\text{i}\theta\hat{J}_i^{m_1 m_2}}$. Notably, a beam-splitter of transmittance η, acting on light modes m_1 and m_2 and transforming a given state $|\text{in}\rangle$ into $|\text{out}\rangle$, may then be represented as such a spin rotation around the y axis [Campos *et al.*, 1989; Yurke *et al.*, 1986]:

$$\begin{pmatrix} \hat{m}_1 \\ \hat{m}_2 \end{pmatrix}_{\text{out}} = \begin{pmatrix} \sqrt{\eta} & -\sqrt{1-\eta} \\ \sqrt{1-\eta} & \sqrt{\eta} \end{pmatrix} \begin{pmatrix} \hat{m}_1 \\ \hat{m}_2 \end{pmatrix}_{\text{in}} \quad \Longleftrightarrow \quad |\text{out}\rangle = U_y^{m_1 m_2}(\theta)\,|\text{in}\rangle \quad \text{with} \quad \theta = 2\arccos\sqrt{\eta}. \tag{5.19}$$

Similarly, the accumulation of phase φ in between two modes corresponds just to a rotation around the z axis: $U_z^{m_1 m_2}(\varphi)$. As a consequence, we are able to fully re-express the action of the lossy interferometer of Fig. 5.1 employing the JS representation, which we schematically depict in Fig. 5.4 after already ignoring the input and output beam-splitters of Fig. 5.1 that do not contribute to the phase-sensing process. On the other hand, in order to verify such description, one may inspect the overall input-output relation between $|\psi_{\text{in}}^N\rangle$ and ρ_φ^N (previously decomposed into independent channels acting on the constituent photons to match the N-parallel–channel scheme of Fig. 4.1), but this time working in the *second-quantisation* picture, so that:

$$\rho_\varphi^N = \sum_{l_a, l_b = 0}^\infty \hat{K}_{l_a, l_b}(\varphi)\,|\psi_{\text{in}}^N\rangle\langle\psi_{\text{in}}^N|\,\hat{K}_{l_a, l_b}^\dagger(\varphi), \tag{5.20}$$

where we have employed the natural Kraus operators expressible with help of the modal bosonic operators and parametrised by the numbers of photons lost in each arm [Dorner *et al.*, 2009], i.e.

$$\hat{K}_{l_a, l_b}(\varphi) = \sqrt{\frac{(1-\eta_a)^{l_a}}{l_a!}}\,\sqrt{\eta_a}^{\hat{a}^\dagger\hat{a}}\,\hat{a}^{l_a}\,\sqrt{\frac{(1-\eta_b)^{l_b}}{l_b!}}\,\sqrt{\eta_b}^{\hat{b}^\dagger\hat{b}}\,\hat{b}^{l_b}\,\text{e}^{-\text{i}\frac{\varphi}{2}(\hat{a}^\dagger\hat{a} - \hat{b}^\dagger\hat{b})}. \tag{5.21}$$

In accordance with Fig. 5.1, each $\hat{K}_{l_a, l_b}(\varphi)$ accounts first for the phase accumulation and afterwards for the losing l_a and l_b photons in arms a and b respectively. Furthermore, after simple algebra one may also show that each Kraus operator (5.21) may be importantly written utilising the spin rotations in the JS picture, so that

$$\hat{K}_{l_a, l_b}(\varphi) = {}_{a'}\langle l_a|\,{}_{b'}\langle l_b|\,U_y^{aa'}(\theta_a)\,U_y^{bb'}(\theta_b)\,U_z^{ab}(\varphi)\,|0\rangle_{a'}\,|0\rangle_{b'}, \tag{5.22}$$

[8] $i = \{1 \equiv x, 2 \equiv y, 3 \equiv z\}$.
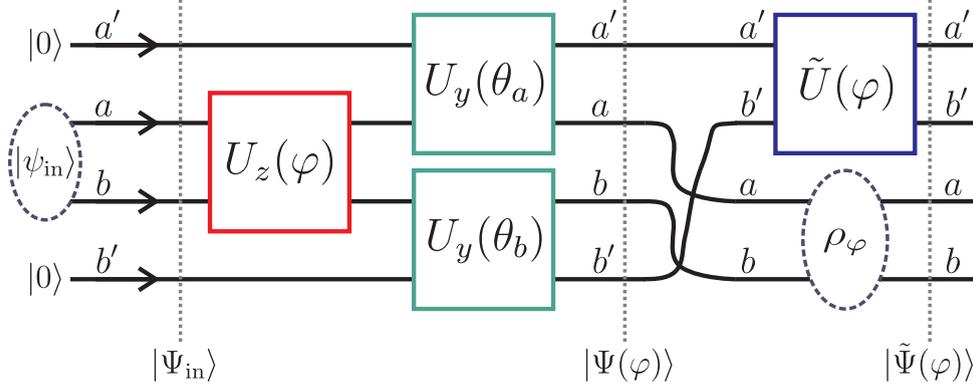
FIGURE 5.4: **Jordan-Schwinger representation of the lossy interferometer** depicted in Fig. 5.1. From left to right, the unitary transformations in the figure correspond to: $U_z^{ab}(\varphi)$ – phase accumulation; $U_y^{aa'}(\theta_a)$ and $U_y^{bb'}(\theta_b)$ – fictitious beam-splitters introduced in arms a and b to mimic losses; $\tilde{U}^{a'b'}(\varphi) = e^{-i\hat{h}_{a'b'}\delta\varphi}$ – local ($\varphi = \varphi_0 + \delta\varphi$) unitary rotation of the excluded environmental modes that must be optimised over all $\hat{h}_{a'b'}$ to establish the optimal purification, $|\tilde{\Psi}(\varphi)\rangle$, of the output state. $|\psi_{\text{in}}\rangle$ and ρ_φ represent respectively the interferometer input and output states of Fig. 5.1, whereas $|\Psi_{\text{in}}\rangle$ and $|\Psi(\varphi)\rangle$ denote their corresponding purifications that also include the environmental modes a' and b'.
Note that the colours of above transformations match the adequate optical elements in Fig. 5.1.

where $\theta_{a/b} = 2\arccos\sqrt{\eta_{a/b}}$ and $|n\rangle_m$ generally represents an n-photon Fock state in mode m. Notice that Eq. (5.22) represents exactly the "circuit" of Fig. 5.4 with the environmental modes a' and b' being accordingly prepared in the vacuum state $|0\rangle$ and eventually projected onto the l_a- and l_b-photon Fock states.

Crucially, the JS representation of Fig. 5.4 allows us to straightforwardly vary the output state purification, i.e. $|\Psi(\varphi)\rangle$ containing the environmental modes, by introducing a general unitary transformation $\tilde{U}^{a'b'}(\varphi)$ acting only on modes a' and b', so that $|\tilde{\Psi}(\varphi)\rangle = \tilde{U}^{a'b'}(\varphi)|\Psi(\varphi)\rangle$. Furthermore, due to the locality of the QFI, when estimating around a given φ_0 with $\varphi = \varphi_0 + \delta\varphi$, all the adequate "environment-rotations" may be parametrised as $\tilde{U}^{a'b'}(\varphi) = e^{-i\hat{h}_{a'b'}\delta\varphi}$, where the Hermitian generator, $\hat{h}_{a'b'}$, plays exactly the role of $\hat{h}_{\text{E}}$ introduced in Sec. 3.2.2.2, yet being now generalised to an infinite-dimensional Hilbert space. Previously, $\hat{h}_{\text{E}}$ yielded a Hermitian matrix h defined in some basis of the environmental subspace and the optimisation over purifications (see Sec. 3.2.2.2) corresponded to a search through all such matrices. Now, as unfortunately we may build *any* Hermitian generator, $\hat{h}_{a'b'}$, from the modal bosonic operators: $\hat{a}'$, $\hat{a}'^\dagger$, $\hat{b}'$, $\hat{b}'^\dagger$; that may be chosen to be of arbitrary order, the minimisation of the QFI over all purifications cannot be in principle explicitly performed.

Nevertheless, as we are only interested in establishing a bound $F_{\text{Q}}[\rho_\varphi^N] \le F_{\text{Q}}[|\tilde{\Psi}_\varphi\rangle]$ in the asymptotic N limit, motivated by the work of [Escher *et al.*, 2011] and the symmetry of the problem, we constrain the generator to have the form $\hat{h}_{a'b'} = -\frac{1}{2}(\gamma_a\hat{n}_a - \gamma_b\hat{n}_b)$, where by $\hat{n}_m = \hat{m}^\dagger\hat{m}$ we denote the photon-number operator of a given mode m. Intuitively (similarly to $e^{-i\hat{h}_{\text{E}}\delta\varphi}$ in Fig. 4.2), we may thus interpret $\tilde{U}^{a'b'}(\varphi)$ as an *erasure operation* which "unwinds" the phase encoded in the environmental modes. Such an interpretation becomes even more evident in the equal-losses scenario, when we may assume the optimal input $|\psi_{\text{in}}^N\rangle$ to possess the modal (a/b) interchange symmetry and set $\gamma = \gamma_a = \gamma_b$, so that $\tilde{U}^{a'b'}(\varphi) = U_z^{a'b'}(-\gamma\varphi)$ is exactly the reverse phase accumulation. Then, because of the relation $U_z^{a'b'}(-\varphi)U_y^{aa'}(\theta_a)U_y^{bb'}(\theta_b)U_z^{ab}(\varphi) = U_z^{ab}(\varphi)U_y^{aa'}(\theta_a)U_y^{bb'}(\theta_b)$, we may also interpret $\tilde{U}^{a'b'}(\varphi)$ as an operation that—by varying γ from 0 to 1—commutes the losses with the phase delay, so that when $\gamma = 1$ they

occur before the phase delay not allowing the environmental modes to possess any information about[9] φ. Although in the most general scenario the optimal purification has also to account for the asymmetry of the loss model and $\gamma_a \neq \gamma_b$, we are always able to write the QFI of the generated purification, and thus the required bound, as:

$$F_Q\left[\rho_\varphi^N\right] \leq \min_{\gamma_a,\gamma_b} F_Q\left[\left|\tilde{\Psi}(\varphi)\right\rangle\right] = \min_{\gamma_a,\gamma_b}\left\{\Delta^2 \hat{h}_{a'b'}^{\text{in}}\big|_{\text{in}} + \Delta^2 \hat{J}_z^{ab}\big|_{\text{in}} - 2\,\text{cov}\left(\hat{h}_{a'b'}^{\text{in}}, \hat{J}_z^{ab}\right)\big|_{\text{in}}\right\} = \tag{5.23}$$

$$= \min_{\gamma_a,\gamma_b}\left\{\begin{array}{c}(1-\tilde{\gamma}_a)^2\,\Delta^2 \hat{n}_a\big|_{\text{in}} + \eta_a\tilde{\gamma}_a\langle\hat{n}_a\rangle_{\text{in}} + (1-\tilde{\gamma}_b)^2\,\Delta^2 \hat{n}_b\big|_{\text{in}} + \eta_b\tilde{\gamma}_b\langle\hat{n}_b\rangle_{\text{in}} + \\ -2\,(1-\tilde{\gamma}_a)\,(1-\tilde{\gamma}_b)\,\text{cov}(\hat{n}_a,\hat{n}_b)\big|_{\text{in}}\end{array}\right\}.$$

The first expression in Eq. (5.23) follows by writing explicitly the QFI (3.57) of the purification $\left|\tilde{\Psi}(\varphi)\right\rangle$ after conveniently defining[10] $\hat{h}_{a'b'}^{\text{in}} = U_z^{ab}(-\varphi_0)U_y^{aa'}(-\theta_a)U_y^{bb'}(-\theta_b)\hat{h}_{a'b'}U_y^{aa'}(\theta_a)U_y^{bb'}(\theta_b)U_z^{ab}(\varphi_0)$ as the generator of $\tilde{U}^{a'b'}(\varphi)$ evolved back through the "circuit" of Fig. 5.4 in the Heisenberg picture, so that it now acts on the purified input state $|\Psi_{\text{in}}\rangle = |\psi_{\text{in}}^N\rangle_{ab}|0\rangle_{a'}|0\rangle_{b'}$, with respect to which all the quantities in Eq. (5.23) are evaluated, i.e. $\bullet|_{\text{in}} \equiv \langle\Psi_{\text{in}}|\bullet|\Psi_{\text{in}}\rangle$. Notice that such a formula may be intuitively explained, as the parameter may be treated at the level of $|\Psi_{\text{in}}\rangle$ as if it was encoded by two non-commuting generators ($\hat{J}_z^{ab}$ and $\hat{h}_{a'b'}$ evaluated for the initial stage of the "circuit"), so that the QFI then corresponds to the *sum of their variances minus twice their covariance*[11] [Demkowicz-Dobrzański *et al.*, 2015]. In order to derive the second expression in Eq. (5.23) with $\tilde{\gamma}_{a/b} = \gamma_{a/b}(1-\eta_{a/b})$, one must commute $\hat{h}_{a'b'}$ through the "circuit" of Fig. 5.4, what is possible by utilising the Baker-Campbell-Hausdorff formula and the adequate bosonic operators commutation relations[12]. Finally, Eq. (5.23) may be explicitly minimised w.r.t. $\gamma_{a/b}$, and for the special case of the input state (5.1) consisting of a *definite* number of photons, N, it may be also upper-limited by [Escher *et al.*, 2011]:

$$F_Q\left[\rho_\varphi^N\right] \leq \left(\frac{2N}{\sqrt{1+\frac{1-\eta_a}{\eta_a}N} + \sqrt{1+\frac{1-\eta_b}{\eta_b}N}}\right)^2 \leq \frac{4\,N}{\left(\sqrt{\frac{1-\eta_a}{\eta_a}} + \sqrt{\frac{1-\eta_b}{\eta_b}}\right)^2}, \tag{5.24}$$

so that the required SQL-like bound $\mathcal{F}_{as}^{bound}$ of Eq. (4.14) is recovered, which indeed reproduces the ultimate quantum limit on precision (5.12).

5.3   Bayesian approach – global phase estimation

Finally, we analyse the lossy interferometer of Fig. 5.1 within the complementary *Bayesian approach* to phase estimation. Following the recipe of Sec. 3.2.3, we assume *no prior knowledge* about the estimated phase and utilise the *average cost* (3.74) that reads:

$$\langle\mathcal{C}_H\rangle = \text{Tr}\left\{\left\langle\rho_\varphi^N\right\rangle_{C_H}\Xi^N\right\} = 2 - \boldsymbol{\alpha}^T\mathbf{A}\,\boldsymbol{\alpha} \tag{5.25}$$

with ρ_φ^N now representing the general output state of the interferometer (5.2). The second expression in (5.25) constitutes an analogue of Eq. (3.85) obtained previously for the noiseless MZ interferometer,

---

[9]Such an interpretation may be slightly misleading, as it suggests to always naively set $\gamma = 1$, in order not to let any "information" about the parameter leak out into the environmental modes. Such choice, however, is actually *not* optimal even in the $N \to \infty$ limit [Escher *et al.*, 2011].

[10]Due to locality, we may without loss of generality set $\varphi_0 = 0$ to simplify calculations, what implies $U_z^{ab}(\varphi_0) = \tilde{U}^{a'b'}(\varphi_0) = \mathbb{I}$.

[11]As defined in Eq. (2.20) before, the *covariance* of any two observables $\hat{A}$ and $\hat{B}$ reads: $\text{cov}(\hat{A},\hat{B}) = \frac{1}{2}\left\langle\left\{\hat{A},\hat{B}\right\}\right\rangle - \langle\hat{A}\rangle\langle\hat{B}\rangle$.

[12]We do not include the derivation here, as it is equivalent to the one presented in the Sup. Mat. of [Escher *et al.*, 2011].

where again the matrix $\mathbf{A}$ possesses only non-zero terms on its first off-diagonals that this time read:

$$A_{n,n-1} = A_{n-1,n} = \sum_{l_a,l_b=0}^{n,N-n} \sqrt{b_n^{(l_a,l_b)} b_{n-1}^{(l_a,l_b)}} \tag{5.26}$$

with $b_n^{(l_a,l_b)}$ being the binomial coefficients defined in Eq. (5.3). Before analysing the problem of optimisation of the average cost over the input state (5.1) real coefficients: $\boldsymbol{\alpha} = \{\alpha_0,\ldots,\alpha_N\}$, we discuss the derivation of Eq. (5.25) in more detail.

As in the case of the noiseless interferometer analysis of Sec. 3.2.4.2, in order to obtain the second expression in (5.25), we must firstly minimise the average cost over all *covariant POVMs*, i.e. their seed elements: $\Xi^N \geq 0$. Yet, the direct-sum structure of the lossy interferometer output state (5.2) implies that without loss of optimality we may also assume: $\Xi_{\text{opt}}^N = \bigoplus_{N'=0}^N \Xi_{\text{opt}}^{N'}$. Physically, the block-diagonal form of Ξ_{opt}^N indicates that the optimal covariant measurement requires a non-demolition photon-number measurement to be performed before carrying out any phase measurements, so that the orthogonal subspaces, labelled by the number of surviving photons N', may be firstly distinguished. One may correctly expect that, subsequently after learning N', it is optimal to just perform the measurement derived in Sec. 3.2.4.2 for the noiseless scenario, i.e. $\Xi_{\text{opt}}^{N'} = |e^{N'}\rangle\langle e^{N'}|$ with $|e^{N'}\rangle = \sum_{n=0}^{N'} |n, N'-n\rangle$. However, due to the losses potentially present in *both* arms, one cannot compensate any more at the measurement stage for the complex phases of the input state coefficients, i.e. ϕ_n in each $\alpha_n = |\alpha_n| e^{i\phi_n}$, as this would also require the exact knowledge of how many photons were lost in each of the arms. Note that evaluating the "lossy" equivalent of Eq. (3.82) for the output state (5.2), ρ_φ^N, we most generally obtain:

$$\begin{aligned}
\langle \mathcal{C}_{\text{H}} \rangle &= 2 \left(1 - \text{Re} \left\{ \sum_{N'=0}^N \sum_{l_a=0}^{N-N'} \sqrt{b_{n+l_a}^{(l_a,N-N'-l_a)} b_{n-1+l_a}^{(l_a,N-N'-l_a)}} \sum_{n=0}^{N'} \alpha_{n+l_a}^* \alpha_{n-1+l_a} \Xi_{n,n-1}^{N'} \right\} \right) \\
&\geq 2 \left(1 - \sum_{N'=0}^N \sum_{l_a=0}^{N-N'} \sqrt{b_{n+l_a}^{(l_a,N-N'-l_a)} b_{n-1+l_a}^{(l_a,N-N'-l_a)}} \sum_{n=0}^{N'} |\alpha_{n+l_a}| |\alpha_{n-1+l_a}| \left| \Xi_{n,n-1}^{N'} \right| \right) \\
&\geq 2 \left(1 - \sum_{n,m=0}^N A_{n,n-1} |\alpha_n| |\alpha_{n-1}| \right). \tag{5.27}
\end{aligned}$$

For the first inequality above to be saturated—as l_a is in principle not measurable—we must just impose $\forall_{n=0}^N : \alpha_n = |\alpha_n|$, so that after indeed choosing $\Xi^N = \Xi_{\text{opt}}^N$ we obtain the third expression coinciding with Eq. (5.25). On the other hand, for the case of *single-arm* losses such requirement is unnecessary, as by measuring N' we also then learn $l_a = N - N'$ and may thus always set $|e^{N'}\rangle = \sum_{n=0}^{N'} e^{i\phi_{n+l_a}} |n, N'-n\rangle$ in Ξ_{opt}^N, in order to saturate the first inequality independently of the input state complex phases.

Nevertheless, after taking without loss of optimality all α_n in Eq. (5.1) to be real, we may identify similarly to Sec. 3.2.4.2 the *minimal* average cost (5.25) for the lossy interferometer as $2 - \lambda_{\text{max}}$, where λ_{max} is again the maximal eigenvalue of the matrix $\mathbf{A}$ and the corresponding eigenvector $\boldsymbol{\alpha}$ provides the *optimal* input state coefficients. Recall (see Sec. 3.1.3.2) that the minimal average cost quantifies the maximal achievable precision and in the $N \to \infty$ limit may be interpreted as the $\overline{\text{MSE}}$ (3.26), $\langle \Delta^2 \tilde{\varphi} \rangle$, due to the convergence of the cost function (3.40) to the squared distance, $(\tilde{\varphi} - \varphi)^2$, as $\tilde{\varphi} \to \varphi$ with N.

On the other hand, similarly to the case of the frequentist approach of Sec. 5.2 and, in particular, the investigation of the QFI bound (5.9) [Demkowicz-Dobrzański *et al.*, 2009], one should also verify if the bosonic input states (5.1) are really optimal from the Bayesian perspective, as more general inputs consisting of *distinguishable photons* could in principle yield better precision. Yet, we explicitly demonstrate in App. J that by generalising the above analysis to input states which allow for individual targeting of each of the constituent photons, the average cost (5.25) may only increase, so that it is indeed

sufficient to restrict only to the inputs (5.1). Furthermore, we show that the corresponding covariant POVMs that apply to such distinguishable-photons–strategies encompass the adaptive measurement strategies [Wiseman *et al.*, 2009; Wiseman and Killip, 1997, 1998], and hence—similarly to the local approach of Sec. 5.2—the measurement adaptivity is theoretically not necessary to attain the ultimate precision determined within the Bayesian approach, although helpful from practical perspective.

Classical input states

As in the case of the local approach, we derive the minimal average cost (3.74) attained by the *classical strategy* employing a coherent state, $|\alpha\rangle$, split on the input beam-splitter of with transmittance τ_{in} shown in Fig. 5.1. This corresponds to evaluating (5.25) for the state $\rho_\varphi^{|\alpha\rangle}$ (5.7), what, however, requires a straightforward generalisation of the optimal covariant POVM to $\Xi_{\mathrm{opt}} = \bigoplus_{N'=0}^\infty \Xi_{\mathrm{opt}}^{N'}$, so that its block-diagonal form matches Eq. (5.7) for all the N'-photon sectors up to $N' \to \infty$. After setting $|\alpha|^2 = N$ and assuming a general τ_{in}, we obtain:

$$\langle \mathcal{C}_{\mathrm{H}} \rangle_{\mathrm{cl}} = 2 - \frac{2\,\mathcal{B}[N\eta_a\tau_{\mathrm{in}}]\,\mathcal{B}[N\eta_b(1-\tau_{\mathrm{in}})]}{N\sqrt{\eta_a\tau_{\mathrm{in}}\,\eta_b\,(1-\tau_{\mathrm{in}})}} \stackrel{N\to\infty}{=} \left(\frac{1}{\tau_{\mathrm{in}}\eta_a} + \frac{1}{(1-\tau_{\mathrm{in}})\eta_b} \right) \frac{1}{N}. \tag{5.28}$$

where $\mathcal{B}(x) = e^{-x} \sum_{n=0}^\infty \frac{x^n}{n!}\sqrt{n}$ is the Bell polynomial of order $1/2$. Although for finite N Eq. (5.28) may only be numerically minimised over τ_{in}, in the strong-beam regime of $N \to \infty$ it takes a more appealing form shown in the second expression above, which similarly to the local approach indicates that it is optimal to intuitively set $\tau_{\mathrm{in}} = 1/(1 + \sqrt{\eta_a/\eta_b})$ yielding the highest *visibility* of the interferometer. Moreover, as indicated in Fig. 5.2(**b**) (*short-dashed blue lines*), asymptotically the same precision is then achieved as within the frequentist approach, so that the minimal average cost (5.28) attains the local classical bound (5.10), i.e.

$$\langle \mathcal{C}_{\mathrm{H}} \rangle_{\mathrm{cl}} \stackrel{N\to\infty}{=} \frac{1}{4} \left(\frac{1}{\sqrt{\eta_a}} + \frac{1}{\sqrt{\eta_b}} \right)^2 \frac{1}{N}. \tag{5.29}$$

Let us note that the convergence of Eqs. (5.10) and (5.29) should have been expected, as we are dealing with uncorrelated particles within the classical strategy, so that the relation (3.34), which connects frequentist and Bayesian precision measures at the classical level, directly applies.

5.3.1 Numerical solution for moderate N

In the noiseless case discussed in Sec. 3.2.4.2, despite the deteriorating assumption of the lack of prior knowledge, the minimal average cost (3.87) has allowed to prove the asymptotic HL-like scaling of precision, which is attained by the BW states (3.86) employed as inputs. In Fig. 5.2(**b**), we present the change of the coefficient values defined by $\boldsymbol{\alpha}$ minimising Eq. (5.25), as the losses are gradually introduced in the interferometer. In contrast to the frequentist approach depicted in Fig. 5.2(**a**)—in which the optimal $|\alpha_n|^2$-distribution rapidly changes abandoning NOON states (3.78) for $\eta < 1$, so that other α_n coefficients must be discontinuously introduced as η further decreases [Knysh *et al.*, 2011]—the optimal solution in the Bayesian case is "smoothly" modified with the growth of losses. Nevertheless, qualitatively a similar effect is observed in both cases, that as the amount of losses increases higher weights must be associated with intermediate α_n-coefficients at the expense of the marginal ones, in order to increase robustness of the input and preserve the quantum superposition even after some photons are lost. On the other hand, a growing bias towards coefficients with higher n must be introduced due to the asymmetry of the single-arm–losses model depicted in Fig. 5.2. In the regime of very high losses, $\eta \to 0$, the $|\alpha_n|^2$-distributions derived within the complementary frequentist and Bayesian approaches attain a

common one, suggesting that the role of the prior knowledge is severely reduced when nearly all the photons are lost in the setup.

5.3.2 Asymptotic SQL-like bound on precision

Unfortunately, we can only numerically determine the minimal average cost (5.25) (corresponding to $2 - \lambda_{\max}$ with $\lambda_{\max}$ being the maximal eigenvalue of the matrix $\mathbf{A}$), so that the asymptotic precision scaling with N cannot be again verified analytically. However, similarly to the case of the local approach in which we have utilised the QFI-upper–bounding techniques, we prove the asymptotic SQL-like scaling of precision within the Bayesian approach by constructing a lower bound on the minimal average cost that scales classically, as $1/N$, with the number of photons, whenever any losses are present.

We achieve this by constructing a valid upper bound on $\lambda_{\max}$ that applies if either $\eta_a < 1$ or $\eta_b < 1$. Recall that $\mathbf{A}$ is a symmetric real matrix, and thus for any arbitrary normalised real vector $\boldsymbol{v}$, $\boldsymbol{v}^T \mathbf{A} \boldsymbol{v} \le \lambda_{\max}$. Let $\boldsymbol{\alpha}$ be the eigenvector corresponding to $\lambda_{\max}$, so that $\boldsymbol{\alpha}^T \mathbf{A} \boldsymbol{\alpha} = \lambda_{\max}$. The fact that all matrix elements of $\mathbf{A}$ are non-negative implies that also all $\alpha_n \ge 0$. Let us now define a matrix $\mathbf{A}'$ such that all non-zero entries of $\mathbf{A}$ are replaced by its largest element $A_N^\uparrow = \max_n \{A_{n,n-1}\}$, which is contained within the only non-zero, off-diagonal entries (5.26). In case of the *single-arm losses*, $A_N^\uparrow$ corresponds either to the first or last term, i.e. $A_{N,N-1}$ for ($\eta_a < 1$, $\eta_b = 1$) and $A_{1,0}$ for ($\eta_a = 1$, $\eta_b < 1$), whereas for *equal losses* ($\eta_a = \eta_b < 1$) to the middle term $A_{\lceil \frac{N}{2} \rceil, \lceil \frac{N}{2} \rceil - 1}$. For *unequal-losses* scenario $A_N^\uparrow$ is one of the entries in between, which unfortunately may only be determined numerically. We have labelled $A_N^\uparrow$ with the subscript N, in order to emphasise that the maximal entry of $\mathbf{A}$ defined by Eq. (5.26) explicitly depends on the photon number, with the *only* exception being the noiseless scenario in which $\forall_n : A_{n,n-1} = 1$ in accordance with Eq. (3.85).

Since $\alpha_n \ge 0$ and $A'_{n,m} \ge A_{n,m} \ge 0$, we can write:

$$\lambda_{\max} = \boldsymbol{\alpha}^T \mathbf{A} \boldsymbol{\alpha} \le \boldsymbol{\alpha}^T \mathbf{A}' \boldsymbol{\alpha} \le \lambda'_{\max}, \tag{5.30}$$

where $\lambda'_{\max}$ is now the maximal eigenvalue of $\mathbf{A}'$. Crucially, $\lambda'_{\max}$ can be easily found analytically after noticing that $\mathbf{A}' = A_N^\uparrow \bar{\mathbf{A}}$, where $\bar{\mathbf{A}}$ is the matrix occurring in the noiseless scenario and defined in Eq. (3.85). As multiplication by a constant just rescales equally all the eigenvalues, so that in particular $\lambda'_{\max} = A_N^\uparrow \bar{\lambda}_{\max} = 2 A_N^\uparrow \cos [\pi/(N+2)]$ with $\bar{\lambda}_{\max}$ corresponding to the maximal eigenvalue in the noiseless setting, we obtain the desired lower bound on the minimal average cost (5.25) that reads:

$$\langle \mathcal{C}_{\mathrm{H}} \rangle \ge 2 \left[1 - A_N^\uparrow \cos \left(\frac{\pi}{N+2} \right) \right]. \tag{5.31}$$

As $A_N^\uparrow$ is analytically defined *only* for single-arm and equal losses, we can explicitly determine the analytical form of Eq. (5.31) for $N \to \infty$ only in these two cases. Nevertheless, we numerically verify that Eq. (5.31) actually coincides in the asymptotic N limit with the bound (5.12) derived within the local approach independently of the $\eta_{a/b}$ chosen, so that most generally:

$$\langle \mathcal{C}_{\mathrm{H}} \rangle \ge \frac{1}{4} \left(\sqrt{\frac{1 - \eta_a}{\eta_a}} + \sqrt{\frac{1 - \eta_b}{\eta_b}} \right)^2 \frac{1}{N}. \tag{5.32}$$

In Fig. 5.3(**b**), we study the minimal average cost (5.25) attainable in the *single-arm* losses scenario ($\eta_a = \eta$, $\eta_b = 1$) after setting $\eta = 0.7$. The results show that the above asymptotic bound (5.32) (*solid grey line*) is indeed saturated within the global approach by the optimal input states (*solid black line*) with α_n-coefficients distributed according to Fig. 5.2(**b**). Although the results are numerical, one should note that

within the Bayesian approach—as the optimisation corresponds just to determining the largest eigenvalue of the **A** matrix—much higher photon numbers ($N \lesssim 1000$) are reachable allowing to numerically verify the convergence of the minimal average cost to the limit dictated by Eq. (5.32) for the whole range of $\eta_{a/b}$ up to a convincing accuracy. For comparison, as in case of the local approach analysed in Fig. 5.3(**a**), we show also the cost, $\langle \mathcal{C}_{\mathrm{H}} \rangle_{\mathrm{cl}}$, attained by the coherent-state–based *classical strategy* (*short-dashed blue line*), which is determined for large N by Eq. (5.29). Moreover, we plot $\langle \mathcal{C}_{\mathrm{H}} \rangle_{\mathrm{BW}}$ (3.87) (*long-dashed black line*) achieved by the BW states (3.86) in the *noiseless* setting, which should be really treated as a benchmark defining the HL within the Bayesian framework.

Let us emphasise that due to the presence of losses the frequentist and Bayesian asymptotic precision bounds (5.12) and (5.32) coincide, what importantly contrasts the noiseless case in which a discrepancy-factor of π^2 has been observed in between Eqs. (3.79) and (3.87). In the case of classical inputs, i.e. the coherent input states (5.5), such behaviour could have been expected (see Eqs. (5.10) and (5.29)), as similarly to Sec. 3.1.4 the interfering photons are then uncorrelated, so that the relation (3.34) may be again invoked to prove the convergence of the CRB (3.7) and the average Bayesian cost (3.36) at the classical level of probabilities. On the other hand, such explanation fails for the quantum-enhanced strategies, as the optimal input states must then benefit from the inter-particle entanglement even in the presence of losses and the constituent photons cannot be thus treated as independent "statistical objects" also in the asymptotic N limit. Yet, the convergence of the asymptotic bounds, combined with the fact that they are known to be saturable for $N \to \infty$ without need of conducting procedure repetitions ($\nu = 1$ in Fig. 3.6) [Demkowicz-Dobrzański *et al.*, 2015; Knysh *et al.*, 2014], suggests that the optimal input states may be well-approximated for sufficiently large N by inputs exhibiting only short-range correlations of the particles. As a result the input states may then be effectively thought of as if they consisted of entangled groups of particles (photons) that do not possess any correlations in between the groupings, so that by setting $N \to \infty$ the regime of infinite sampling (infinite number of uncorrelated groups) is naturally attained in a single shot. Such an intuition has recently been shown to be indeed correct by efficiently approximating the optimal input states in the presence of losses with use of the so-called matrix-product states [Jarzyna and Demkowicz-Dobrzański, 2013] that do not exhibit long range inter-particle correlations.

Chapter 6

Conclusions and outlook

In this thesis, we have analysed in detail the effects of uncorrelated noise in quantum metrological scenarios, and most importantly presented tools that allow to bound the maximal attainable precision in the presence of decoherence. We have shown that generic types of such noise, even infinitesimally small in magnitude, force the precision to scale at the Standard Quantum Limit (i.e. limit the Mean Squared Error to follow the $1/N$ scaling) when the number of particles employed, N, becomes infinitely large, and thus constrain the ultimate quantum enhancement to a constant factor. Furthermore, we have proposed a generalisation of these asymptotic methods to the finite-N regime, which importantly is always efficiently computable due to its semi-definite program form.

Yet, as the bounds obtained are typically saturated in the asymptotic N limit by employing non-complex quantum states and measurements, our results crucially prove that, in order to improve large-scale experiments involving many particles, the primary priority of an experimentalist should be the reduction of the noise present, as investing in more exotic quantum states and detection strategies does not then lead to any significant precision improvement. A spectacular application of our precision bounds has been demonstrated in the analysis of gravitational wave-detectors [LIGO Collaboration, 2011, 2013], in which Demkowicz-Dobrzański et al. [2013] have explicitly shown that the currently conducted detection schemes are working already at the theoretical limits predicted by our techniques. On the other hand, our finite-N semi-definite programming methods allow to verify the size of the small-N regime, in which the presence of decoherence may be effectively ignored. As a consequence, they prove for the first time the optimality of the protocols utilised in the pioneering quantum metrology experiments (e.g. [Leibfried et al., 2003; Mitchell et al., 2004]), which assumed the absence of noise.

Moreover, as our results also apply to metrological scenarios in which the single-particle evolution varies with the total number of particles involved (e.g. the optical depth of an atomic ensemble changes with the number of atoms employed in an experiment [Wasilewski et al., 2010]), they may be equivalently utilised to quantify the attainable precision in frequency estimation models—see Sec. 4.4—in which the time duration of each experimental trial serves as an extra parameter that may in principle be freely adjusted. In [Chaves et al., 2013], our methods have been applied to a similar model as the one presented in Sec. 4.4, yet with non-commuting Liouvillian and Hamiltonian parts of the master equation (see Sec. 2.2.2). In particular, the tools proposed in this thesis have allowed to identify for the first time (according to our best knowledge) an example of a Markovian uncorrelated noise for which the SQL-like scaling may still be surpassed—corresponding to a dephasing-type noise directed in a transversal direction to the Hamiltonian encoding the frequency. As the attainability of super-classical precision

scalings has also been demonstrated in Non-Markovian noise models [Chin *et al.*, 2012; Matsuzaki *et al.*, 2011], an interesting question arises, whether our bounding-techniques (which are then also applicable) could suggest further room for improvement in such frequency estimation protocols with memory.

We would also like to point that our SQL-like bounds allow one to avoid some of the controversies characteristic for the idealised noiseless scenarios. When decoherence is *not* present and the probe states with indefinite number of particles are considered (e.g. squeezed states of light), the exact form of the Heisenberg Limit needs to be reconsidered [Berry *et al.*, 2012; Giovannetti and Maccone, 2012; Hall *et al.*, 2012; Hofmann, 2009; Hyllus *et al.*, 2010b], since the direct replacement of N with mean number of particles $\bar{N}$ may make the HL invalid. Moreover, the final claims on the achievable precision scaling may strongly depend on the form of the prior knowledge about the parameter assumed, and lead to apparent contradictions [Anisimov *et al.*, 2010; Giovannetti *et al.*, 2012]. These difficulties do not arise in the presence of uncorrelated noise, as the asymptotic SQL-like bounds are then valid also when N is replaced by $\bar{N}$—as explicitly shown in Sec. 5.2.2. In particular, the bounds obtained are saturated in a single-shot scenario, unlike the decoherence-free case when only after some number of independently repeated experiments one may expect to approach the theoretical limits [Giovannetti and Maccone, 2012; Pezzé and Smerzi, 2009]. This is due to the fact that by employing input states of grouped particles, which possess no correlations in between the groupings, and by letting the groups to be of finite but sufficiently large size, one can attain the ultimate asymptotic SQL-like bound up to any precision [Jarzyna and Demkowicz-Dobrzański, 2014]. It is thus the uncorrelated noise that makes the asymptotic N limit naturally account for the infinite sampling regime, what—by the arguments of Sec. 3.1.3.1—also explains why the frequentist and Bayesian approaches yield same asymptotic measures of accuracy. We have demonstrated such a behaviour explicitly for the lossy interferometer model in Chap. 5, yet a general explanation has been recently provided in [Jarzyna and Demkowicz-Dobrzański, 2014]. As a result, we also expect the *information theoretic* results [Hall and Wiseman, 2012; Nair, 2012], complementary to frequentist and Bayesian methods, to recover bounds in the presence of uncorrelated noise that are compatible with our techniques determined for the local (Fisher-Information–based) measures. Notice that the above argument also suggests the asymptotically optimal form of the input states, which should include ones that do not possess long-range correlations in between the particles. This observation has already been made in [Sørensen and Mølmer, 2001] and indicates that in methods designed to search for the optimal inputs in scenarios with uncorrelated noise one may restrict himself to states with short-range correlations such as for example the matrix product states of low bond dimensions [Jarzyna and Demkowicz-Dobrzański, 2013].

For completeness, in order to also be critical, let us note that the best-performing method we have proposed stems from the Channel Extension presumption—see Sec. 4.3.1.3—assuming presence of ancillary non-evolving particles, what allows then to carry out the semi-definite program reformulation. Thus, although the Channel Extension method turns out to be most agile and effective, it may still provide asymptotic and finite-N bounds that are not saturable in real-life protocols, in which such auxiliary particles are not available. Hence, it is important always to verify the tightness of the bounds predicted by our results, what, however, has very recently been shown to be possible by utilising a more involved approach based on the calculus of variations [Knysh *et al.*, 2014]. What is more, throughout this thesis we have analysed models in which one deals with a *single* latent parameter of interest to be estimated in a metrological scenario, so that a natural future work on our methods would be to generalise them and study their applicability in the *multi-parameter* estimation schemes [Crowley *et al.*, 2014; Genoni *et al.*, 2013; Humphreys *et al.*, 2013; Vidrighin *et al.*, 2014]. However, let us note that in such a case there does not exist a clear equivalent of the Fisher Information (matrix) in the quantum setting [Hayashi, 2005], what questions whether saturable non-trivial precision bounds can even be constructed. However, as in the multi-parameter estimation scenario the non-commutativity of observables—in particular, the

generators of the estimated parameters—starts to play a crucial role, many unexpected and interesting issues may arise when following these lines of research. On the other hand, we have also entirely focussed on the uncorrelated noise models, so that one may wonder what are the effects of collective decoherence, which coherently affects all the constituent particles of a given quantum system. As physically relevant types of such noise may be treated as fluctuations of the estimated parameter itself [Dorner, 2012; Genoni *et al.*, 2011; Knysh *et al.*, 2014; Maciejewski *et al.*, 2014], correlated noise generally yields an extra constant offset-term in the error of an estimator, which importantly does not diminish with an increase in the number of particles N. Thus, one may intuitively expect the uncorrelated noise to vary the scaling of the degradable term to become SQL-like in accordance with our methods, while the other correlated-noise term persists unaffected [Knysh *et al.*, 2014]. An interesting question arises, whether it is possible to develop techniques similar to ours that would predict bounds when considering noise-types of a fixed correlation size that may be varied smoothly from a single particle (uncorrelated noise studied in this thesis) to all particles (correlated noise) [Jeske *et al.*, 2013; Maciejewski, 2014].

Lastly, let us remark that, as noted in various parts of this work (see Notes 2.5 and 4.1), the parameter estimation techniques can be naturally employed to quantify precision with which the time duration of system evolution may be determined. This is a very special metrology-like problem that has been analysed in various ways not directly related to statistical inference techniques of estimation theory [Aharonov *et al.*, 2002], and most notably also with use of the so-called "*speed limits*" that bound the speed with which a quantum system may evolve [Deffner and Lutz, 2013a; Giovannetti *et al.*, 2003a; Margolus and Levitin, 1998]. Furthermore, it has been demonstrated similarly to the metrological scenario that, given a system consisting of N particles, their inter-particle entanglement substantially speeds up the evolution process [Fröwis, 2012; Giovannetti *et al.*, 2003b; Zander *et al.*, 2007], as also predicted by the Quantum Cramér-Rao Bound (4.3) yielding the HL-like scaling, $1/N^2$, of the time-duration estimator for a maximally entangled state. Thus, it is a natural question to ask whether our methods could be adapted to such a particular setting, in order to account for the impact of uncorrelated noise; and how are they related to the recent results which allow to incorporate decoherence-effects into the 'speed-limit' formalism [del Campo *et al.*, 2013; Taddei *et al.*, 2013], or even the Non-Markovian behaviour of the evolution [Deffner and Lutz, 2013b].

Appendix A

Choi criterion for extremality of a quantum channel

Let us consider a quantum channel—a CPTP map (see Sec. 2.2.1)—$\Lambda : \mathcal{B}(\mathcal{H}_{\text{in}}) \to \mathcal{B}(\mathcal{H}_{\text{out}})$ of rank r, so that it admits a set of linearly independent Kraus operators $\{K_i\}_{i=1}^r$. For simplicity, let us also denote the space of all CPTP maps as $\mathcal{CPTP}$ and its subset containing the *extremal* quantum channels as $\partial\mathcal{CPTP}$. We explicitly prove, following [Choi, 1975], the *Choi criterion* 2.4.2 stating that:

Theorem:

Λ *is an extremal map if and only if* $\left\{K_i^\dagger K_j\right\}_{ij}$ *is a linearly independent set of* r^2 *matrices, i.e. the only* $r \times r$ *matrix* $\boldsymbol{\mu}$ *that satisfies* $\sum_{ij} \boldsymbol{\mu}_{ij} K_i^\dagger K_j = 0$ *is the trivial one* $\boldsymbol{\mu} = 0$.

Proof:

$\Lambda \in \partial\mathcal{CPTP} \implies$ the set of matrices $\left\{K_i^\dagger K_j\right\}_{ij}$ is linearly independent.

Suppose there exists $\boldsymbol{\mu} \neq 0$ such that $\epsilon \sum_{ij} \boldsymbol{\mu}_{ij} K_i^\dagger K_j = 0$ for any $\epsilon > 0$. $\boldsymbol{\mu}$ can be assumed to be Hermitian, as by taking the Hermitian conjugate: $\left(\sum_{ij} \boldsymbol{\mu}_{ij} K_i^\dagger K_j\right)^\dagger = \sum_{ij} \boldsymbol{\mu}_{ij}^* K_j^\dagger K_i = \sum_{ij} \boldsymbol{\mu}_{ij}^\dagger K_i^\dagger K_j$, so that equivalently $\epsilon \sum_{ij} \left(\boldsymbol{\mu} \pm \boldsymbol{\mu}^\dagger\right)_{ij} K_i^\dagger K_j = 0$. Let us define two maps $\Lambda_\pm : \mathcal{B}(\mathcal{H}_{\text{in}}) \to \mathcal{B}(\mathcal{H}_{\text{out}})$ such that $\forall_{\varrho \in \mathcal{B}(\mathcal{H}_{\text{in}})} : \Lambda_\pm[\varrho] = \sum_{i,j=1}^r (\mathbb{I} \pm \epsilon\boldsymbol{\mu})_{ij} K_i \varrho K_j^\dagger$. Firstly, $\text{Tr}\{\Lambda_\pm[\varrho]\} = \text{Tr}\{\varrho\} \pm \epsilon \, \text{Tr}\left\{\sum_{i,j=1}^r \boldsymbol{\mu}_{ij} K_j^\dagger K_i \, \varrho\right\} = 1$, what proves that both $\Lambda_\pm$ are TP. On the other hand, let us choose ϵ small enough, so that $\mathbb{I} \pm \epsilon\boldsymbol{\mu} \geq 0$, and construct $\boldsymbol{\nu} = \sqrt{\mathbb{I} \pm \epsilon\boldsymbol{\mu}}$. Then, $\Lambda_\pm[\varrho] = \sum_{i,j=1}^r \left(\boldsymbol{\nu}_\pm^2\right)_{ij} K_i \varrho K_j^\dagger = \sum_{i=1}^r K_i^\pm \varrho K_i^{\pm\dagger}$ with effective Kraus operators $K_i^\pm = \sum_{j=1}^r \boldsymbol{\nu}_{\pm ji} K_j$, what proves that $\Lambda_\pm \in \mathcal{CPTP}$. Now, as the original channel is obtained by composing $\Lambda = \frac{1}{2}\left(\Lambda_+ + \Lambda_-\right)$, it *cannot* be extremal according to the original Def. 2.4.1. ∎

The set of matrices $\left\{K_i^\dagger K_j\right\}_{ij}$ is linearly independent $\implies \Lambda \in \partial\mathcal{CPTP}$.

Suppose $\Lambda \notin \partial\mathcal{CPTP}$, so that there exist two maps $\Lambda_\pm \in \mathcal{CPTP}$ that yield $\Lambda = \frac{1}{2}(\Lambda_+ + \Lambda_-)$ and possess Kraus representations such that $\forall_{\varrho \in \mathcal{B}(\mathcal{H}_{\text{in}})} : \Lambda_\pm[\varrho] = \sum_{i=1}^{r_\pm} K_i^\pm \varrho K_i^{\pm\dagger}$ with $\sum_{i=1}^{r_\pm} K_i^{\pm\dagger} K_i^\pm = \mathbb{I}$. On the other hand, in the CJ matrix (2.24) representation $\Omega_\Lambda = \frac{1}{2}(\Omega_{\Lambda_+} + \Omega_{\Lambda_-})$, so that the support of the CJ matrix of Λ, $\Omega_\Lambda = \sum_{i=1}^r |K_i\rangle\langle K_i|$, must contain both $\Omega_{\Lambda_\pm} = \sum_{i=1}^{r_\pm} |K_i^\pm\rangle\langle K_i^\pm|$. Thus, we may express any $|K_i^\pm\rangle$ as a superposition of $\{|K_i\rangle\}_{i=1}^r$, what means that in the channel-picture all Kraus operators $\{K_i^\pm\}_{i=1}^{r_\pm}$ can be written as linear compositions of $\{K_i\}_{i=1}^r$, e.g. $K_i^+ = \sum_{j=1}^r \boldsymbol{\alpha}_{ij} K_j$ where $\boldsymbol{\alpha}$ is some (non-unitary) $r^+ \times r$ matrix. Then, the TP properties of respectively Λ_+ and Λ imply $\sum_{i,j=1}^r \left(\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}\right)_{ij} K_i^\dagger K_j = \mathbb{I}$ and $\sum_{i=1}^r K_i^\dagger K_i = \mathbb{I}$, so that there exists $\boldsymbol{\mu} = \boldsymbol{\alpha}^\dagger\boldsymbol{\alpha} - \mathbb{I} \neq 0$ for which $\sum_{i,j=1}^r \boldsymbol{\mu}_{ij} K_i^\dagger K_j = 0$. Hence, the set of $\left\{K_i^\dagger K_j\right\}_{ij}$ is linearly dependent. ∎

Appendix B

Criterion for φ-extremality of a quantum channel

Interpreting a family of CPTP maps $\{\Lambda_\varphi\}_\varphi$ parametrised by φ as a curve in the space all quantum channels, we have *geometrically* defined in Def. 2.4.1 the quantum channel Λ_φ to be *φ-extremal* at φ_0, if it *cannot* be decomposed there into a probabilistic mixture of two quantum channels $\Lambda_\pm$, i.e. $\Lambda_{\varphi_0} = p_+\Lambda_+ + p_-\Lambda_-$, that lie along the tangent to the curve at φ_0 or, equivalently, along the direction defined by the derivative $\partial_\varphi \Lambda_\varphi|_{\varphi_0} \equiv \dot{\Lambda}_\varphi|_{\varphi_0}$, so that $\Lambda_\pm = \Lambda_{\varphi_0} \pm \epsilon \dot{\Lambda}_\varphi|_{\varphi_0}$ for some non-zero ϵ. Due to the CJ-isomorphism described in Sec. 2.4.1, the above requirement can be directly translated onto the density matrix representation of quantum channels, so that it is tantamount to non-existence of $\epsilon > 0$ such that

$$\Omega_{\Lambda_{\varphi_0}} \pm \epsilon \dot{\Omega}_{\Lambda_{\varphi_0}} \geq 0, \tag{B.1}$$

where $\Omega_{\Lambda_{\varphi_0}} = \sum_i |K_i\rangle\langle K_i|$ and $\dot{\Omega}_{\Lambda_{\varphi_0}} = \sum_i |\dot{K}_i\rangle\langle K_i| + |K_i\rangle\langle \dot{K}_i|$ are respectively the CJ matrix of Λ_φ and its derivative at φ_0, with $|K_i\rangle = K_i \otimes \mathbb{I}|\mathbb{I}\rangle$ (we drop the explicit φ-dependence of the Kraus operators for simplicity) corresponding to the canonical Kraus representation of Λ_φ introduced in Sec. 2.4.1. We prove that:

Theorem:
Λ_φ *is φ-extremal at φ_0 if and only if $\dot{\Omega}_{\Lambda_\varphi}$ is not contained within the support of Ω_{Λ_φ} at φ_0.*

In other words, there does not exist $\epsilon > 0$ satisfying Eq. (B.1) if and only if there exists a non-zero Hermitian matrix $\boldsymbol{\mu}$ such that

$$\dot{\Omega}_{\Lambda_{\varphi_0}} = \sum_{ij} \boldsymbol{\mu}_{ij} |K_i\rangle\langle K_j|. \tag{B.2}$$

Proof:
We show the equivalence of Eqs. (B.1) and (B.2), what proves also the equivalence of the above theorem and Def. 2.4.1.

Eq. (B.1) $\implies$ Eq. (B.2)
Assuming Eq. (B.1) to hold, we write explicitly its l.h.s., so that Eq. (B.1) states that

$$\langle\psi|\left[\sum_i |K_i\rangle\langle K_i| \pm \epsilon\left(\sum_i |\dot{K}_i\rangle\langle K_i| + |K_i\rangle\langle \dot{K}_i|\right)\right]|\psi\rangle \geq 0 \tag{B.3}$$

for any (even not normalised) $|\psi\rangle$. In order to prove Eq. (B.2), it is enough to show that all $|\dot{K}_i\rangle$ can be written as linear combinations of $|K_i\rangle$. If this was not the case for one of them, e.g. $|\dot{K}_{\bar{i}}\rangle$, its decomposition would additionally require a vector $|L_{\bar{i}}\rangle$ which is not contained within the support of $\Omega_{\Lambda_{\varphi_0}}$, i.e. orthogonal to all $|K_i\rangle$. Then, taking $|\psi\rangle = \sqrt{\lambda}\,|K_{\bar{i}}\rangle + \mathrm{e}^{\mathrm{i}\phi}\sqrt{1-\lambda}\,|L_{\bar{i}}\rangle$, Eq. (B.3) reads:

$$\lambda\,\langle K_{\bar{i}}|K_{\bar{i}}\rangle \left(\langle K_{\bar{i}}|K_{\bar{i}}\rangle \pm \epsilon\,\frac{\partial\langle K_{\bar{i}}|K_{\bar{i}}\rangle}{\partial\varphi}\bigg|_{\varphi_0}\right) \pm 2\epsilon\sqrt{\lambda(1-\lambda)}\,\mathrm{Re}\big\{\mathrm{e}^{\mathrm{i}\phi}\langle L_{\bar{i}}|\dot{K}_{\bar{i}}\rangle\big\} \quad =$$
$$\overset{\lambda\lessgtr1}{=} \quad \pm 2\epsilon\sqrt{\lambda}\,\mathrm{Re}\big\{\mathrm{e}^{\mathrm{i}\phi}\langle L_{\bar{i}}|\dot{K}_{\bar{i}}\rangle\big\} \;+\; O(\lambda) \;\geq\; 0. \tag{B.4}$$

Thus, for any non-zero ϵ and $\langle L_{\hat{i}}|\dot{K}_{\hat{i}}\rangle$, we may always set λ small enough, so that there exists ϕ for which the l.h.s. of Eq. (B.4) is negative. This leads to a contradiction, hence (B.2) must hold.

Eq. (B.2) $\implies$ Eq. (B.1)
Let us assume the decomposition (B.2) to be valid, so that the condition (B.1) now reads

$$\sum_i |K_i\rangle\langle K_i| \pm \epsilon \sum_{ij} \boldsymbol{\mu}_{ij}|K_j\rangle\langle K_i| \;\geq\; 0 \tag{B.5}$$

and we must show that it holds for some $\epsilon>0$. Yet, we may always define matrices $\boldsymbol{\nu}^\pm = \mathbb{I}\pm\epsilon\boldsymbol{\mu}$ which are positive semi-definite for small enough ϵ, so that by taking their square root we can construct $|\tilde{K}_i^\pm\rangle = \sum_j \big[\sqrt{\boldsymbol{\nu}^\pm}\big]_{ji}|K_j\rangle$. Now, we recover the l.h.s. of Eq. (B.5) by evaluating adequately one of $\sum_i |\tilde{K}_i^\pm\rangle\langle \tilde{K}_i^\pm|$ that are clearly positive semi-definite, what completes the proof. $\blacksquare$

As aside, one should note that in general—by writing explicitly $|K_i\rangle = K_i\otimes\mathbb{I}\,|\mathbb{I}\rangle$ with $|\mathbb{I}\rangle = \sum_{i=1}^{\dim\mathcal{H}_{\mathrm{in}}} |i\rangle_{\mathrm{S}}|i\rangle_{\mathrm{A}}$ and φ-dependence dropped again for simplicity—$\mathrm{Tr}_{\mathrm{S}}\big\{\dot{\Omega}_{\Lambda_\varphi}\big\} = \partial_\varphi\big(\sum_i K_i^\dagger K_i\big) = 0$, and thus by tracing out the $\mathcal{H}_{\mathrm{in}}$ subspace of Eq. (B.2) one obtains:

$$0 = \sum_{ij} \boldsymbol{\mu}_{ij} K_j^\dagger K_i, \tag{B.6}$$

what is the negation of *Choi criterion* for channel extremality (Crit. 2.4.2) discussed in Sec. 2.4.2 and App. A, and thus the condition for the *non-extremality* of the channel Λ_φ. Hence, as Eq. (B.2) implies Eq. (B.6), an intuitive statement is confirmed that: if a channel is φ-non-extremal, then it cannot be extremal; or equivalently: *if a channel is extremal, then it must also be φ-extremal*.

Appendix C

Equivalence of the purification-based QFI definitions (3.63) and (3.64)

For a given state $\varrho_\varphi = \sum_i \lambda_i(\varphi) |e_i(\varphi)\rangle\langle e_i(\varphi)|$ supported by the Hilbert space $\mathcal{H}_\mathrm{S}$ and a parameter true value φ_0, we unambiguously choose the system purification $|\Psi_{\varphi_0}\rangle = \sum_i |\xi_i(\varphi_0)\rangle_\mathrm{S} |i\rangle_\mathrm{E}$ defined in the minimal extended Hilbert space $\mathcal{H}_\mathrm{S} \times \mathcal{H}_\mathrm{E}$ with all $|\xi_i(\varphi_0)\rangle = \lambda_i(\varphi_0)|e_i(\varphi_0)\rangle$ and $\langle i|j\rangle = \delta_{ij}$, so that correctly $\varrho_{\varphi_0} = \mathrm{Tr}_\mathrm{E}\{|\Psi_{\varphi_0}\rangle\langle\Psi_{\varphi_0}|\}$. Then, all the relevant purifications, $|\tilde{\Psi}_\varphi\rangle$, which differ in their corresponding QFIs (3.57) and lead to a change in the r.h.s. of Eqs. (3.63) and (3.64), may be constructed at φ_0 by a unitary rotation of the environment part of $|\Psi_{\varphi_0}\rangle$, i.e. by applying $u_\varphi^\mathrm{E} = \mathrm{e}^{-\mathrm{i}\hat{h}_\mathrm{E}(\varphi-\varphi_0)}$ with any Hermitian $\hat{h}_\mathrm{E}$ that generates $|\tilde{\Psi}_\varphi\rangle = u_\varphi^\mathrm{E}|\Psi_\varphi\rangle$ and locally shifts only the first derivative: $|\tilde{\Psi}_{\varphi_0}\rangle = |\Psi_{\varphi_0}\rangle$, $|\dot{\tilde{\Psi}}_{\varphi_0}\rangle = |\dot{\Psi}_{\varphi_0}\rangle - \mathrm{i}\hat{h}_\mathrm{E}|\Psi_{\varphi_0}\rangle$. In [Fujiwara and Imai, 2008], it has been shown that the purification-based definition (3.64) is minimised and correctly reproduces the QFI *if and only if* a purification is chosen that satisfies the condition $|\dot{\tilde{\Psi}}_{\varphi_0}^\mathrm{opt}\rangle = \frac{1}{2}L_\mathrm{S} \otimes \mathbb{I}^\mathrm{E}|\tilde{\Psi}_{\varphi_0}^\mathrm{opt}\rangle$[1]. Firstly, let us note that consistently the second term in the other purification-based QFI definition (3.63) always vanishes for $|\tilde{\Psi}_{\varphi_0}^\mathrm{opt}\rangle$, as

$$\langle\tilde{\Psi}_{\varphi_0}^\mathrm{opt}|\dot{\tilde{\Psi}}_{\varphi_0}^\mathrm{opt}\rangle = \langle\tilde{\Psi}_{\varphi_0}^\mathrm{opt}|\frac{1}{2}L_\mathrm{S}\otimes\mathbb{I}^\mathrm{E}|\tilde{\Psi}_{\varphi_0}^\mathrm{opt}\rangle = \frac{1}{2}\mathrm{Tr}\{\varrho_{\varphi_0}L_\mathrm{S}\} = \frac{1}{4}\mathrm{Tr}\{\varrho_{\varphi_0}L_\mathrm{S} + L_\mathrm{S}\varrho_{\varphi_0}\} = \frac{1}{2}\mathrm{Tr}\{\dot{\varrho}_{\varphi_0}\} = 0, \qquad \text{(C.1)}$$

what follows from the SLD definition (3.56), so that Eqs. (3.63) and (3.64) are indeed compatible. However, in order to complete the proof and show their equivalence, we demonstrate that $|\tilde{\Psi}_\varphi^\mathrm{opt}\rangle$ is also the purification that minimises Eq. (3.63).

Rewriting the r.h.s. of the definition (3.63) at φ_0 for purifications $|\tilde{\Psi}_\varphi\rangle$, we obtain

$$4\min_{\tilde{\Psi}_{\varphi_0}}\left\{\langle\dot{\tilde{\Psi}}_{\varphi_0}|\dot{\tilde{\Psi}}_{\varphi_0}\rangle - |\langle\tilde{\Psi}_{\varphi_0}|\dot{\tilde{\Psi}}_{\varphi_0}\rangle|^2\right\} = \qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(C.2)}$$

$$= 4\min_{\hat{h}_\mathrm{E}}\left\{\langle\dot{\Psi}_{\varphi_0}|\dot{\Psi}_{\varphi_0}\rangle + 2\,\mathrm{Im}\{\langle\dot{\Psi}_{\varphi_0}|\hat{h}_\mathrm{E}|\Psi_{\varphi_0}\rangle\} + \langle\Psi_{\varphi_0}|\hat{h}_\mathrm{E}^2|\Psi_{\varphi_0}\rangle - |\langle\Psi_{\varphi_0}|\dot{\Psi}_{\varphi_0}\rangle - \mathrm{i}\langle\Psi_{\varphi_0}|\hat{h}_\mathrm{E}|\Psi_{\varphi_0}\rangle|^2\right\}.$$

However, as also $|\dot{\Psi}_{\varphi_0}\rangle = \sum_i |\dot{\xi}_i(\varphi_0)\rangle_\mathrm{S}|i\rangle_\mathrm{E}$, we may define the purification derivative with help of a non-Hermitian operator $D = \sum_i |\dot{\xi}_i(\varphi_0)\rangle\langle\xi_i(\varphi_0)|$ for which $|\dot{\Psi}_{\varphi_0}\rangle = \frac{1}{2}D\otimes\mathbb{I}^\mathrm{E}|\Psi_{\varphi_0}\rangle$. Notice that D acts only on the system subspace and constitutes a logarithmic derivative, as $\dot{\varrho}_{\varphi_0} = \mathrm{Tr}_\mathrm{E}\{|\Psi_{\varphi_0}\rangle\langle\dot{\Psi}_{\varphi_0}| + |\dot{\Psi}_{\varphi_0}\rangle\langle\Psi_{\varphi_0}|\} =$

---

[1] For simplicity, we shorten the notation for the SLD (3.56) of the state ϱ_φ at φ_0, so that $L_\mathrm{S} \equiv L_\mathrm{S}[\varrho_{\varphi_0}]$.

$\frac{1}{2}\left(\varrho_{\varphi_0}D^\dagger + D\varrho_{\varphi_0}\right)$. Hence, we can rewrite Eq. (C.2) further as

$$4\min_{\hat{h}_{\mathrm{E}}}\left\{\left\langle\frac{1}{4}D^\dagger D + \frac{\mathrm{i}}{2}\left(D - D^\dagger\right)\otimes\hat{h}_{\mathrm{E}} + \hat{h}_{\mathrm{E}}^2\right\rangle - \left|\left\langle\frac{1}{2}D - \mathrm{i}\hat{h}_{\mathrm{E}}\right\rangle\right|^2\right\},\tag{C.3}$$

where $\langle\dots\rangle \equiv \langle\Psi_{\varphi_0}|\dots|\Psi_{\varphi_0}\rangle$. Defining the difference of the derivative D from the SLD (3.56) as $\Delta = D - L_{\mathrm{S}}$, which satisfies $\varrho_{\varphi_0}\Delta^\dagger + \Delta\varrho_{\varphi_0} = 0$, and benefiting from: $\langle L_{\mathrm{S}}\rangle = 0$ that follows from Eq. (C.1), $\langle L_{\mathrm{S}}^2\rangle = \mathrm{Tr}\{\varrho_{\varphi_0}L_{\mathrm{S}}^2\} = F_{\mathrm{Q}}[\varrho_\varphi]|_{\varphi_0}$, and $\langle\Delta^\dagger L_{\mathrm{S}} + L_{\mathrm{S}}\Delta\rangle = \mathrm{Tr}\{\varrho_{\varphi_0}\left(\Delta^\dagger L_{\mathrm{S}} + L_{\mathrm{S}}\Delta\right)\} = \mathrm{Tr}\{\left(\varrho_{\varphi_0}\Delta^\dagger + \Delta\varrho_{\varphi_0}\right)L_{\mathrm{S}}\} = 0$, we arrive at

$$4\min_{\hat{h}_{\mathrm{E}}}\left\{\left\langle\frac{1}{4}\left(\Delta^\dagger + L_{\mathrm{S}}\right)\left(\Delta + L_{\mathrm{S}}\right) + \frac{\mathrm{i}}{2}\left(\Delta - \Delta^\dagger\right)\otimes\hat{h}_{\mathrm{E}} + \hat{h}_{\mathrm{E}}^2\right\rangle - \left|\left\langle\frac{1}{2}\Delta - \mathrm{i}\hat{h}_{\mathrm{E}}\right\rangle\right|^2\right\} =$$

$$= \min_{\hat{h}_{\mathrm{E}}}\left\{\left\langle L_{\mathrm{S}}^2\right\rangle + \left\langle\Delta^\dagger\Delta\right\rangle + \left\langle\Delta^\dagger L_{\mathrm{S}} + L_{\mathrm{S}}\Delta\right\rangle + \left\langle 2\mathrm{i}\left(\Delta - \Delta^\dagger\right)\otimes\hat{h}_{\mathrm{E}} + 4\hat{h}_{\mathrm{E}}^2\right\rangle - \left|\left\langle\Delta - 2\mathrm{i}\hat{h}_{\mathrm{E}}\right\rangle\right|^2\right\} =$$

$$= F_{\mathrm{Q}}[\varrho_\varphi]|_{\varphi_0} + \min_{\hat{h}_{\mathrm{E}}}\left\{\left\langle\Delta^\dagger\Delta\right\rangle + \left\langle 2\mathrm{i}\left(\Delta - \Delta^\dagger\right)\otimes\hat{h}_{\mathrm{E}} + 4\hat{h}_{\mathrm{E}}^2\right\rangle - \left|\left\langle\Delta - 2\mathrm{i}\hat{h}_{\mathrm{E}}\right\rangle\right|^2\right\}$$

$$= F_{\mathrm{Q}}[\varrho_\varphi]|_{\varphi_0} + 4\min_{\hat{h}_{\mathrm{E}}}\left\langle\left(\frac{1}{2}\Delta^\dagger + \mathrm{i}\hat{h}_{\mathrm{E}}\right)\left[\mathbb{I}^{\mathrm{SE}} - |\Psi_{\varphi_0}\rangle\langle\Psi_{\varphi_0}|\right]\left(\frac{1}{2}\Delta - \mathrm{i}\hat{h}_{\mathrm{E}}\right)\right\rangle.\tag{C.4}$$

The second term in Eq. (C.4), which should still be minimised over $\hat{h}_{\mathrm{E}}$, is importantly non-negative, as it represents a projection of a non-negative operator, $\mathbb{I}^{\mathrm{SE}} - |\Psi_{\varphi_0}\rangle\langle\Psi_{\varphi_0}|$, onto an (unnormalised) state $\left(\frac{1}{2}\Delta - \mathrm{i}\hat{h}_{\mathrm{E}}\right)|\Psi_{\varphi_0}\rangle$. Importantly, it can be always made zero by choosing $\hat{h}_{\mathrm{E}}$ such that $\left(\frac{1}{2}\Delta - \mathrm{i}\hat{h}_{\mathrm{E}}\right)|\Psi_{\varphi_0}\rangle = 0$, what leads exactly to the previously claimed optimal purification $|\tilde{\Psi}_{\varphi_0}\rangle = |\tilde{\Psi}_{\varphi_0}^{\mathrm{opt}}\rangle$, as

$$\left(\frac{1}{2}\Delta - \mathrm{i}\hat{h}_{\mathrm{E}}\right)|\Psi_{\varphi_0}\rangle = |\dot{\Psi}_{\varphi_0}\rangle - \left(\frac{1}{2}L_{\mathrm{S}} + \mathrm{i}\hat{h}_{\mathrm{E}}\right)|\Psi_{\varphi_0}\rangle = 0\tag{C.5}$$

$$\implies \quad |\dot{\Psi}_{\varphi_0}\rangle - \mathrm{i}\hat{h}_{\mathrm{E}}|\Psi_{\varphi_0}\rangle = \frac{1}{2}L_{\mathrm{S}}\otimes\mathbb{I}^{\mathrm{E}}|\Psi_{\varphi_0}\rangle \quad \iff \quad |\dot{\tilde{\Psi}}_{\varphi_0}\rangle = \frac{1}{2}L_{\mathrm{S}}\otimes\mathbb{I}^{\mathrm{E}}|\tilde{\Psi}_{\varphi_0}\rangle,$$

and completes an alternative to [Escher *et al.*, 2011] proof of the purification-based QFI definition (3.63). ∎

Appendix D

Optimality of covariant POVMs

We follow the *covariant POVM* analysis of [Holevo, 1982], in order to assess the optimality of such a measurement class when employed in case of a single repetition ($\nu = 1$) of the phase estimation scheme of Fig. 3.6, for which the output state of the system most generally reads $\rho_\varphi^N = \mathcal{U}_\varphi^{\otimes N}[\rho_0^N]$ with $\rho_0^N = \mathcal{D}[\rho_{\text{in}}^N]$. Let us write the Bayesian, circularly symmetric average cost (3.72) for a particular POVM M_X acting on all the N particles—which satisfies $\forall_x : M_x \geq 0$, $\oint \mathrm{d}x \, M_x = \mathbb{I}$—and an estimator $\tilde{\varphi}(x) \equiv \tilde{\varphi}_{\nu=1}(x)$:

$$
\begin{aligned}
\langle \mathcal{C}_{\text{H}}(\tilde{\varphi}) \rangle &= \int \frac{\mathrm{d}\varphi}{2\pi} \oint \mathrm{d}x \, \text{Tr}\{\rho_\varphi^N M_x\} \, C_{\text{H}}(\tilde{\varphi}(x) - \varphi) \\
&= \int \frac{\mathrm{d}\varphi}{2\pi} \oint \mathrm{d}x \, \text{Tr}\{\rho_0^N \, U_\varphi^{\dagger \otimes N} M_x \, U_\varphi^{\otimes N}\} \, C_{\text{H}}(\tilde{\varphi}(x) - \varphi) \, .
\end{aligned} \tag{D.1}
$$

Interchanging the order of the integrals and freely shifting $\varphi \to \varphi + \tilde{\varphi}(x)$, what does not affect the φ-integral limits due to invariance of the Haar measure, i.e. $\forall_\delta : \int^{\varphi+\delta} \frac{\mathrm{d}\varphi}{2\pi} \dots = \int^\varphi \frac{\mathrm{d}\varphi}{2\pi} \dots$, we may rewrite Eq. (D.1) as

$$
\begin{aligned}
\langle \mathcal{C}_{\text{H}}(\tilde{\varphi}) \rangle &= \oint \mathrm{d}x \int \frac{\mathrm{d}\varphi}{2\pi} \, \text{Tr}\{\rho_\varphi^N \, U_{\tilde{\varphi}(x)}^{\dagger \otimes N} M_{x_i} U_{\tilde{\varphi}(x)}^{\otimes N}\} \, C_{\text{H}}(\varphi) \\
&= \int \frac{\mathrm{d}\varphi}{2\pi} \, \text{Tr}\{\rho_\varphi^N \, \Xi^N\} \, C_{\text{H}}(\varphi)
\end{aligned} \tag{D.2}
$$

with the *seed element* $\Xi^N = \oint \mathrm{d}x \, U_{\tilde{\varphi}(x)}^{\dagger \otimes N} M_x U_{\tilde{\varphi}(x)}^{\otimes N}$, which thus $\Xi^N \geq 0$ corresponding to a mixture of positive semi-definite operators. Without loss of generality, we may introduce another parameter $\tilde{\varphi}$ playing now the role of the estimator that is integrated also with respect to the invariant Haar measure, so that $\forall_\delta : \int^{\tilde{\varphi}+\delta} \frac{\mathrm{d}\tilde{\varphi}}{2\pi} \dots = \int^{\tilde{\varphi}} \frac{\mathrm{d}\tilde{\varphi}}{2\pi} \dots$ and $\int \frac{\mathrm{d}\tilde{\varphi}}{2\pi} = 1$. After shifting again $\varphi \to \varphi - \tilde{\varphi}$, Eq. (D.2) may finally be written as

$$
\langle \mathcal{C}_{\text{H}}(\tilde{\varphi}) \rangle = \int \frac{\mathrm{d}\varphi}{2\pi} \int \frac{\mathrm{d}\tilde{\varphi}}{2\pi} \, \text{Tr}\{\rho_\varphi^N M_{\tilde{\varphi}}\} \, C_{\text{H}}(\tilde{\varphi} - \varphi) \, , \tag{D.3}
$$

where $M_{\tilde{\varphi}} = U_{\tilde{\varphi}}^{\otimes N} \Xi^N U_{\tilde{\varphi}}^{\dagger \otimes N}$ has the interpretation of a *continuously parametrised covariant POVM*. Consistently, the elements of $M_{\tilde{\varphi}}$ specified by $\tilde{\varphi}$ are positive semi-definite and the overall POVM is complete, as $\int \frac{\mathrm{d}\tilde{\varphi}}{2\pi} M_{\tilde{\varphi}} = \mathbb{I}$, what may be verified by: interchanging the order of the integrals, freely shifting the parameter $\tilde{\varphi}$ to eliminate the $\tilde{\varphi}(x)$-dependence, and acknowledging that $\oint \mathrm{d}x \, M_x = \mathbb{I}$. Crucially, Eq. (D.3) proves that *for any POVM M_X assumed to calculate the average cost (D.1), there always exists a covariant POVM $M_{\tilde{\varphi}}$ that also achieves $\langle \mathcal{C}_H(\tilde{\varphi}) \rangle$.* ∎

Appendix E

RLD bound applies *to and only to* φ-non-extremal quantum channels

In what follows we adopt the notation of Sec. 2.4.1 and App. B, so that for a given quantum channel Λ_φ: $\Omega_{\Lambda_\varphi} = \sum_i |K_i\rangle\!\langle K_i|$ and $\dot{\Omega}_{\Lambda_\varphi} = \sum_i |\dot{K}_i\rangle\!\langle K_i| + |K_i\rangle\!\langle \dot{K}_i|$ respectively represent the corresponding Choi-Jamiołkowski (CJ) matrix and its derivative w.r.t. φ, where $|K_i\rangle = K_i \otimes \mathbb{I} |\mathbb{I}\rangle$ are defined with use of the canonical Kraus operators, which φ-dependence is dropped for simplicity.

In [Hayashi, 2011], it has been proved that the RLD bound (4.9) applies to a given channel Λ_φ for a given parameter value φ_0 if and only if $\left(\dot{\Omega}_{\Lambda_\varphi}\right)^2$ is contained within the support of Ω_{Λ_φ} at φ_0. Here, we show that $(\dots)^2$ is unnecessary in the above statement, so that—recalling the definition of channel φ-non-extremality presented in App. B—it may be rewritten as:

Theorem:
The RLD bound (4.9) on the extended-channel QFI (4.7) applies to a given channel Λ_φ at φ_0 if and only if Λ_φ is φ-non-extremal there.

Proof:
The condition of Hayashi [2011] for the RLD bound (4.9) to be applicable to the channel Λ_φ at φ_0 can be formally written as

$$P_\perp \left(\dot{\Omega}_{\Lambda_{\varphi_0}}\right)^2 P_\perp = 0\,, \tag{E.1}$$

where we denote by $P_\parallel$ and $P_\perp$ projections onto the support and null-space of $\Omega_{\Lambda_{\varphi_0}}$ respectively, so that $\forall_i\colon P_\parallel |K_i\rangle = |K_i\rangle$ whereas $\forall_i\colon P_\perp |K_i\rangle = 0$. On the other hand, we have shown in App. B that the definition of the channel Λ_φ to be φ-non-extremal at φ_0 is equivalent to the statement that $P_\parallel \dot{\Omega}_{\Lambda_{\varphi_0}} P_\parallel = \dot{\Omega}_{\Lambda_{\varphi_0}}$, or in other words—see Eq. (B.2)—, that there exists a non-zero Hermitian matrix $\boldsymbol{\mu}_{ij}$ such that

$$\dot{\Omega}_{\Lambda_{\varphi_0}} = \sum_{ij} \boldsymbol{\mu}_{ij} |K_i\rangle\langle K_j|. \tag{E.2}$$

Eq. (E.2) implies Eq. (E.1), as by substitution

$$P_\perp \left(\sum_{ij} \boldsymbol{\mu}_{ij} |K_i\rangle\langle K_j|\right)^2 P_\perp = \sum_{ij} \left(\sum_p \boldsymbol{\mu}_{ip} \langle K_p|K_p\rangle \boldsymbol{\mu}_{pj}\right) P_\perp |K_i\rangle\langle K_j| P_\perp = 0\,. \tag{E.3}$$

In order to prove the other direction, we split the derivatives of each vector $|K_i\rangle$ into components contained within the support and the null-space of $\Omega_{\Lambda_{\varphi_0}}$, i.e. $|\dot{K}_i\rangle = \sum_j \boldsymbol{\nu}_{ij} |K_j\rangle + |L_i^\perp\rangle$ with $\forall_{i,j} : \langle L_i^\perp | K_j \rangle = 0$. Hence, after substituting for $\dot{\Omega}_{\Lambda_{\varphi_0}}$ the Eq. (E.1) then simplifies to

$$\left(\sum_i |L_i^\perp\rangle\langle K_i| \right) \left(\sum_j |K_j\rangle\langle L_j^\perp| \right) = 0, \tag{E.4}$$

and since $A^\dagger A = 0$ implies $A^\dagger = A = 0$ and $\{|K_i\rangle\}_i$ are orthogonal, we conclude that all $|L_i^\perp\rangle = 0$. Thus, Eq. (E.1) implies that $|\dot{K}_i\rangle = \sum_j \boldsymbol{\nu}_{ij} |K_j\rangle$, which due to the local ambiguity of Kraus representations (4.5) is equivalent to $|\dot{\tilde{K}}_i\rangle = \sum_j (\boldsymbol{\nu}_{ij} - \mathrm{i}\, \mathsf{h}_{ij}) |\tilde{K}_j\rangle$ for any Hermitian matrix h. Therefore, without loss of generality, we may set $\mathsf{h} = -\mathrm{i}\, \boldsymbol{\nu}^{\mathrm{AH}}$ after splitting $\boldsymbol{\nu}$ into its Hermitian and anti-Hermitian parts $\boldsymbol{\nu} = \boldsymbol{\nu}^{\mathrm{H}} + \boldsymbol{\nu}^{\mathrm{AH}}$, so that $|\dot{\tilde{K}}_i\rangle = \sum_j \boldsymbol{\nu}_{ij}^{\mathrm{H}} |\tilde{K}_j\rangle$ with $\boldsymbol{\nu}^{\mathrm{H}} \neq 0$ for any non-trivial channel[1]. Writing the derivative of the CJ matrix in the shifted Kraus representation basis, $\{|\tilde{K}_i\rangle\}_i$, and bearing in mind that at φ_0 all $|\tilde{K}_i\rangle = |K_i\rangle$, we obtain

$$\dot{\Omega}_{\Lambda_{\varphi_0}} = \sum_i |\dot{\tilde{K}}_i\rangle\langle \tilde{K}_i| + |\tilde{K}_i\rangle\langle \dot{\tilde{K}}_i| = 2 \sum_{ij} \boldsymbol{\nu}_{ji}^{\mathrm{H}} |\tilde{K}_i\rangle\langle \tilde{K}_j| = 2 \sum_{ij} \boldsymbol{\nu}_{ji}^{\mathrm{H}} |K_i\rangle\langle K_j| \tag{E.5}$$

and satisfy the condition (E.2) with $\boldsymbol{\mu}^T = 2\boldsymbol{\nu}^{\mathrm{H}}$. ∎

---

[1] Otherwise, there would exist a Kraus representation in which $\forall_i : |\dot{\tilde{K}}_i\rangle = 0$, so that the channel QFI (4.6) vanishes independently of the input!

Appendix F

Optimal purifications that yield extended-channel QFIs and QS/CE bounds

We state below the form of the optimal generators h that shift adequately—see Eq. (4.5)—the first derivatives of the canonical Kraus operators listed in Tab. 2.1 for each of the noisy-phase–estimation models of Fig. 2.2. For each of the noise-types, we provide two versions of optimal $\mathsf{h}_{\mathrm{opt}}$, first of which minimises Eq. (4.8) and thus yields the correct form of the *extended-channel QFI*, $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$, or equivalently, the optimal purification $|\tilde{\Psi}_\varphi^{\mathrm{ext}}\rangle$ depicted in Fig. 4.2(**c**). The second version of $\mathsf{h}_{\mathrm{opt}}$ corresponds in each case to the optimal generator minimising Eq. (4.24), and thus specifying the *CE bound*, $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$. Yet, as for all the channels considered, apart from the spontaneous emission map, such a generator leads to a Kraus representation satisfying the condition $\alpha_{\tilde{K}_{\mathrm{opt}}} \propto \mathbb{I}$ in Eq. (4.20), $\mathsf{h}_{\mathrm{opt}}$ minimises also Eq. (4.24) and thus yields also the coinciding *QS bound*, i.e. $\mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} = \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$. We adopt the standard notation, in which $\hat{\sigma}_0$ represents a 2×2 identity matrix and $\{\hat{\sigma}_i\}_{i=1}^{3}$ are the Pauli operators.

Dephasing:
Optimal h minimising the expression (4.8) for the extended-channel QFI, $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$:

$$\mathsf{h}_{\mathrm{opt}} = -\frac{\sqrt{1-\eta^2}}{2}\,\hat{\sigma}_1\,. \tag{F.1}$$

Optimal h minimising the expressions (4.21)/(4.24) for the QS/CE bounds, $\mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} = \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$:

$$\mathsf{h}_{\mathrm{opt}} = -\frac{1}{2\sqrt{1-\eta^2}}\,\hat{\sigma}_1\,. \tag{F.2}$$

Depolarisation:

Optimal h minimising the expression (4.8) for the extended-channel QFI, $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$:

$$\mathsf{h}_{\mathrm{opt}} = -\frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & \xi \\ 0 & \begin{bmatrix} \hat{\sigma}_2 \end{bmatrix} & 0 \\ 0 & & & 0 \\ \xi & 0 & 0 & 0 \end{pmatrix} \quad \text{with} \quad \xi = \frac{\sqrt{(1+3\eta)(1-\eta)}}{1+\eta}. \tag{F.3}$$

Optimal h minimising the expressions (4.21)/(4.24) for the QS/CE bounds, $\mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} = \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$:

$$\mathsf{h}_{\mathrm{opt}} = -\frac{c}{2}\begin{pmatrix} 0 & 0 & 0 & \xi \\ 0 & \begin{bmatrix} \hat{\sigma}_2 \end{bmatrix} & 0 \\ 0 & & & 0 \\ \xi & 0 & 0 & 0 \end{pmatrix} \quad \text{with} \quad \xi = \frac{\sqrt{(1+3\eta)(1-\eta)}}{1+\eta} \quad \text{and} \quad c = \frac{1+\eta}{(1+2\eta)(1-\eta)}. \tag{F.4}$$

Loss:

Optimal h minimising the expression (4.8) for the extended-channel QFI, $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$:

$$\mathsf{h}_{\mathrm{opt}} = -\frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & \begin{bmatrix} \hat{\sigma}_3 \end{bmatrix} \\ 0 & & \end{pmatrix}. \tag{F.5}$$

Optimal h minimising the expressions (4.21)/(4.24) for the QS/CE bounds, $\mathcal{F}_{\mathrm{as}}^{\mathrm{QS}} = \mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$:

$$\mathsf{h}_{\mathrm{opt}} = -\frac{1}{2(1-\eta)}\begin{pmatrix} 0 & 0 & 0 \\ 0 & \begin{bmatrix} \hat{\sigma}_3 \end{bmatrix} \\ 0 & & \end{pmatrix}. \tag{F.6}$$

Spontaneous emission (amplitude damping):

Optimal h minimising the expression (4.8) for the extended-channel QFI, $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}]$:

$$\mathsf{h}_{\mathrm{opt}} = \frac{1}{2}\begin{pmatrix} \xi & 0 \\ 0 & 1 \end{pmatrix} \quad \text{with} \quad \xi = \frac{1+\sqrt{\eta}}{1-\sqrt{\eta}}. \tag{F.7}$$

Optimal h minimising the expression (4.24) for the CE bound, $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$:

$$\mathsf{h}_{\mathrm{opt}} = \frac{1}{2(1-\eta)}\left(\eta\,\hat{\sigma}_0 - \hat{\sigma}_3\right) \tag{F.8}$$

Appendix G

CE bound applies to all φ-non-extremal maps and is tighter than the RLD bound

Firstly, let us note that in App. E, while proving that any channel that admits the RLD bound (4.9) must be φ-non-extremal, we chose a Kraus representation generated by $\mathsf{h} = -\mathrm{i}\,\boldsymbol{\nu}^{\mathrm{AH}}$ that actually satisfies the necessary $\beta_{\tilde{K}} = 0$ *condition* (4.23) of the CE method. Recalling that in App. E the matrix $\boldsymbol{\nu}$ was specified so that for all i: $\big|\dot{K}_i\big\rangle = \sum_j \boldsymbol{\nu}_{ij}\,|K_j\rangle$, we obtain exactly Eq. (4.23) by taking the partial trace over the subspace $\mathcal{H}_{\mathrm{S}}$ of both sides of the identity:

$$\sum_{ij} \mathsf{h}_{ij}\,|K_j\rangle\langle K_i| = -\mathrm{i}\sum_{ij}\frac{1}{2}\big(\boldsymbol{\nu}_{ij} - \boldsymbol{\nu}_{ij}^{\dagger}\big)\,|K_j\rangle\langle K_i| = \frac{\mathrm{i}}{2}\sum_i |K_i\rangle\big\langle\dot{K}_i\big| - \big|\dot{K}_i\big\rangle\langle K_i|, \tag{G.1}$$

what thus proves that the *CE method applies to all φ-non-extremal maps*, which are really the ones that admit the RLD bound—see App. E. This is consistent, as the CE method, being applicable to all the quantum simulable channels described in Sec. 4.3.1.2, must also trivially apply to all the classically simulable ones that are also exactly the φ-non-extremal maps.

Furthermore, we show that the CE bound (4.24) is at least as tight as the RLD bound (4.9). We prove this by substituting the representation (E.5) of the CJ matrix derivative into the definition of $\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$ in Eq. (4.9), so that

$$\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}] = 4\left\|\mathrm{Tr}_{\mathcal{H}_{\mathrm{S}}}\bigg\{\sum_{ij}\boldsymbol{\nu}_{ji}^{\mathrm{H}}\big|\tilde{K}_i\big\rangle\sum_{pq}\boldsymbol{\nu}_{pq}^{\mathrm{H}}\big\langle\tilde{K}_q\big|\bigg\}\right\| = 4\left\|\sum_i \dot{\tilde{K}}_i^{\dagger}\dot{\tilde{K}}_i\right\|, \tag{G.2}$$

where we have used the fact that $\big\langle\tilde{K}_j\big|\Omega_\varphi^{-1}\big|\tilde{K}_p\big\rangle = \delta_{jp}$. Hence, $\mathcal{F}^{\mathrm{RLD}}[\Lambda_\varphi \otimes \mathcal{I}]$ is an example of the CE bound (4.24) with a possibly sub-optimal Kraus operators chosen that satisfy $\forall_i : \big|\dot{\tilde{K}}_i\big\rangle = \sum_j \boldsymbol{\nu}_{ij}^{\mathrm{H}}\big|\tilde{K}_j\big\rangle$ and the $\beta_{\tilde{K}} = 0$ condition (4.23), as shown in the paragraph above. ∎

Appendix H

Optimal local QS of a channel

A given channel Λ_φ of rank r in order to be locally *quantum simulable* at φ_0 must fulfil the condition (see Sec. 4.3.1.2):

$$\Lambda_\varphi[\varrho] = \mathrm{Tr}_{\mathrm{E}_\Phi \mathrm{E}_\sigma}\big\{ U(\varrho \otimes |\xi_\varphi\rangle\langle\xi_\varphi|) U^\dagger \big\} + O(\delta\varphi^2) = \sum_{i=1}^{r' \geq r} \bar{K}_i(\varphi)\, \varrho\, \bar{K}_i(\varphi)^\dagger + O(\delta\varphi^2), \qquad \text{(H.1)}$$

where $\varphi = \varphi_0 + \delta\varphi$ and $\bar{K}_i(\varphi) = \langle i| U |\xi_\varphi\rangle$ and $\{|i\rangle\}_{i=1}^{r'}$ form any basis in the r' dimensional $\mathcal{H}_{\mathrm{E}_\Phi} \times \mathcal{H}_{\mathrm{E}_\sigma}$ space containing $|\xi_\varphi\rangle$. Hence, Λ_φ must admit at φ_0 a Kraus representation $\{\tilde{K}_i\}_{i=1}^{r'}$—with possibly linearly dependent Kraus operators, as for generality we assume $r' \geq r$—that coincides with the one of Eq. (H.1) up to $O(\delta\varphi^2)$, i.e. satisfies $\tilde{K}_i = \bar{K}_i$ and $\dot{\tilde{K}}_i = \dot{\bar{K}}_i$ for all i.

We construct a valid decomposition of $|\dot{\xi}_{\varphi_0}\rangle$ into its (normalised) components parallel and perpendicular to $|\xi_{\varphi_0}\rangle$: $|\dot{\xi}_{\varphi_0}\rangle = \mathrm{i}\, a |\xi_{\varphi_0}\rangle - \mathrm{i}\, b |\xi_{\varphi_0}^\perp\rangle$, where we can choose $a, b \in \mathbb{R}$ because of $\frac{\partial\langle\xi_\varphi|\xi_\varphi\rangle}{\partial\varphi} = 0$ and the irrelevance of the global phase. Then, the asymptotic bound $\mathcal{F}_{\mathrm{as}}^{\mathrm{bound}}$ of Eq. (4.14) determined by the local QS (H.1) at φ_0 simply reads $F_{\mathrm{Q}}[|\xi_\varphi\rangle]|_{\varphi_0} = 4b^2$ and the required Kraus operators $\{\tilde{K}_i\}_{i=1}^{r'}$ of Λ_φ must fulfil conditions $\tilde{K}_i = \langle i| U |\xi_{\varphi_0}\rangle$ and $\dot{\tilde{K}}_i = \langle i| U |\dot{\xi}_{\varphi_0}\rangle = \mathrm{i}\, a \tilde{K}_i - \mathrm{i}\, b \langle i| U |\xi_{\varphi_0}^\perp\rangle$. Hence, for the local QS of channel Λ_φ to be valid, b must be finite and we must always be able to construct

$$U = \begin{bmatrix} \tilde{K}_1 & \frac{a}{b}\tilde{K}_1 + \frac{\mathrm{i}}{b}\dot{\tilde{K}}_1 & \bullet & \dots & \bullet \\ \tilde{K}_2 & \frac{a}{b}\tilde{K}_2 + \frac{\mathrm{i}}{b}\dot{\tilde{K}}_2 & \bullet & \dots & \bullet \\ \tilde{K}_3 & \frac{a}{b}\tilde{K}_3 + \frac{\mathrm{i}}{b}\dot{\tilde{K}}_3 & \vdots & \ddots & \vdots \\ \vdots & \vdots & \bullet & \dots & \bullet \end{bmatrix} \qquad \text{(H.2)}$$

with first two columns adequately fixed to give for all i the correct $\langle i| U |\xi_{\varphi_0}\rangle$ and $\langle i| U |\xi_{\varphi_0}^\perp\rangle$ respectively. Yet, owing to the locality of the condition (H.1), all entries marked with $\bullet$ in Eq. (H.2) can be chosen freely, in order to satisfy the unitarity condition: $U^\dagger U = U U^\dagger = \mathbb{I}$, which nevertheless constraints the Kraus operators to simultaneously fulfil $\mathrm{i} \sum_{i=1}^{r'} \dot{\tilde{K}}_i^\dagger \tilde{K}_i = a\,\mathbb{I}$ and $\sum_{i=1}^{r'} \dot{\tilde{K}}_i^\dagger \dot{\tilde{K}}_i = (b^2 + a^2)\,\mathbb{I}$. However, as without loss of generality we may shift their phase at φ_0 via $\tilde{K}_i \to \mathrm{e}^{-\mathrm{i} a\varphi_0} \tilde{K}_i$, the above conditions can be made independent of a, so that: $\mathrm{i} \sum_{i=1}^{r'} \dot{\tilde{K}}_i^\dagger \tilde{K}_i = 0$ and $\sum_{i=1}^{r'} \dot{\tilde{K}}_i^\dagger \dot{\tilde{K}}_i = b^2 \mathbb{I}$. On the other hand, such constraints do not require $r' > r$, as rewriting for example the first one as $\mathrm{i} \sum_{i=1}^{r'} \langle \dot{\xi}_{\varphi_0}| U |i\rangle\langle i| U |\xi_{\varphi_0}\rangle = 0$, one can always resolve the identity with some basis vectors $\sum_{i=1}^{r'} |i\rangle\langle i| = \sum_{i=1}^{r} |e_i\rangle\langle e_i|$ and define linearly independent Kraus operators $\{K_i = \langle e_i| U |\xi_{\varphi_0}\rangle\}_{i=1}^{r}$ also fulfilling the necessary requirements.

Finally, realising that $b^2 = \frac{1}{4} F_Q[|\xi_\varphi\rangle]|_{\varphi_0}$, we may conclude that Λ_φ is locally *quantum simulable* at φ_0, if—as stated in the main text in Eq. (4.20)—it admits there a Kraus representation satisfying conditions:

$$\mathrm{i} \sum_{i=1}^{r} \dot{K}_i(\varphi_0)^\dagger K_i(\varphi_0) = 0 \qquad \text{and} \qquad \sum_{i=1}^{r} \dot{K}_i(\varphi_0)^\dagger \dot{K}_i(\varphi_0) = \frac{1}{4} F_Q[|\xi_\varphi\rangle]|_{\varphi_0} \, \mathbb{I}, \tag{H.3}$$

where $\{K_i(\varphi)\}_i$ should be constructable from any valid linearly independent Kraus operators via the unitary transformation (4.5) generated by some Hermitian $r \times r$ matrix h. ∎

Appendix I

Finite-N CE method as an SDP

The finite-N CE bound has been defined in Eq. (4.25) as

$$\mathcal{F}_N^{\mathrm{CE}} = 4 \min_{\mathsf{h}} \left\{ \left\| \alpha_{\tilde{K}} \right\| + (N-1) \left\| \beta_{\tilde{K}} \right\|^2 \right\}, \tag{I.1}$$

where $\|\cdot\|$ denotes the operator norm, $\alpha_{\tilde{K}} = \sum_i \dot{\tilde{K}}_i^\dagger \dot{\tilde{K}}_i$ and $\beta_{\tilde{K}} = \mathrm{i} \sum_i \dot{\tilde{K}}_i^\dagger \tilde{K}_i$. Given a channel Λ_φ mapping states between d_{in}- and d_{out}-dimensional Hilbert spaces and the set of its linearly independent Kraus operators $\{K_i\}_{i=1}^r$—corresponding to $d_{\mathrm{out}} \times d_{\mathrm{in}}$ matrices—in order to compute $\mathcal{F}_N^{\mathrm{CE}}$ we must minimise Eq. (I.1) over locally equivalent Kraus representations (4.5) of Λ_φ generated by all Hermitian, $r \times r$ matrices h.

Adopting a concise notation in which $\mathbf{K}$ is a column vector containing the starting Kraus operators K_i as its elements, we can associate all locally equivalent Kraus representations $\tilde{\mathbf{K}}$ in Eq. (I.1) with those generated by any h via $\tilde{\mathbf{K}} = \mathbf{K}$ and $\dot{\tilde{\mathbf{K}}} = \dot{\mathbf{K}} - \mathrm{i}\mathsf{h}\mathbf{K}$. By constructing matrices—$\mathbb{I}_d$ represents a $d \times d$ identity matrix—:

$$\mathbf{A} = \begin{bmatrix} \sqrt{\lambda_a} \mathbb{I}_{d_{\mathrm{in}}} & \dot{\tilde{\mathbf{K}}}^\dagger \\ \dot{\tilde{\mathbf{K}}} & \sqrt{\lambda_a} \mathbb{I}_{r \cdot d_{\mathrm{out}}} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \sqrt{\lambda_b} \mathbb{I}_{d_{\mathrm{in}}} & \left(\mathrm{i}\dot{\tilde{\mathbf{K}}}^\dagger \tilde{\mathbf{K}} \right)^\dagger \\ \mathrm{i}\dot{\tilde{\mathbf{K}}}^\dagger \tilde{\mathbf{K}} & \sqrt{\lambda_b} \mathbb{I}_{d_{\mathrm{in}}} \end{bmatrix}, \tag{I.2}$$

which positive semi-definiteness conditions correspond respectively to

$$\alpha_{\tilde{K}} = \dot{\tilde{\mathbf{K}}}^\dagger \dot{\tilde{\mathbf{K}}} \leq \lambda_a \mathbb{I}_{d_{\mathrm{in}}} \qquad \beta_{\tilde{K}}^\dagger \beta_{\tilde{K}} = \tilde{\mathbf{K}}^\dagger \dot{\tilde{\mathbf{K}}} \dot{\tilde{\mathbf{K}}}^\dagger \tilde{\mathbf{K}} \leq \lambda_b \mathbb{I}_{d_{\mathrm{in}}}, \tag{I.3}$$

we may rewrite Eq. (I.1) into the desired SDP form as

$$\begin{aligned} \mathcal{F}_N^{\mathrm{CE}} &= 4 \min_{\mathsf{h}} \left\{ \lambda_a + (N-1)\lambda_b \right\}, \\ &\text{s.t.} \quad \mathbf{A} \geq 0, \mathbf{B} \geq 0. \end{aligned} \tag{I.4}$$

For the purpose of this work we have implemented all the required SDPs using the CVX package for Matlab [Grant and Boyd, 2012], which efficiently evaluates Eq. (I.4) given the set of Kraus operators and their derivatives of a generic channel Λ_φ.

Lastly, one should note that by slightly modifying the program in Eq. (I.4) we are able to also efficiently evaluate: the *extended-channel QFI* (4.8), as $\mathcal{F}[\Lambda_\varphi \otimes \mathcal{I}] = \mathcal{F}_{N=1}^{\mathrm{CE}}$; and the *CE bound* (4.24), $\mathcal{F}_{\mathrm{as}}^{\mathrm{CE}}$, by setting $N=1$ and imposing in Eq. (I.4) also the $\beta_{\tilde{K}} = 0$ condition (4.23) that is importantly linear in $\tilde{\mathbf{K}}$ and $\dot{\tilde{\mathbf{K}}}$.

Appendix J

Lossy interferometry with distinguishable photons and adaptive measurements

Distinguishability of photons

If the photons travelling through the lossy interferometer of Fig. 5.1 are *distinguishable*, e.g. they are prepared in different time-bins, we must generalise the description of the input state (5.1) in accordance with Sec. 2.1.4 to:

$$|\psi_{\text{in}}^N\rangle = \sum_{\boldsymbol{n}=\boldsymbol{0}^N}^{\boldsymbol{1}^N} \alpha_{\boldsymbol{n}} |\boldsymbol{n}\rangle, \tag{J.1}$$

where the sum runs over all N-bit sequences $\boldsymbol{n}$ with $|\boldsymbol{n}\rangle = |n_1\rangle \otimes \ldots |n_N\rangle$, and $|n_i\rangle = |1\rangle(|0\rangle) \equiv |a\rangle(|b\rangle)$ denotes the photon in the i-th time-bin, propagating in the $a(b)$ arm of the interferometer respectively.

Taking photonic losses into account, we additionally need to track the time-slots in which photons were lost. Therefore, we define a binary string $\boldsymbol{l}_a = l_{a,1} l_{a,2} \ldots l_{a,N}$ with 1's representing the time-bins in which a given photon was lost in arm a and similarly $\boldsymbol{l}_b$ for the arm b. Formally, using the bitwise subtraction, we can thus also define $\boldsymbol{N}' = \boldsymbol{1} - \boldsymbol{l}_a - \boldsymbol{l}_b$, where 1's in the binary string $\boldsymbol{N}'$ denote the time-bins in which photons were successfully transmitted. As a result—introducing the notation in which for any binary sequence $\boldsymbol{x}$ we denote by $x = |\boldsymbol{x}| = \sum_{i=1}^N x_i$ the number of appearing 1's—we may identify $N' = |\boldsymbol{N}'|$ as exactly the overall number of surviving photons introduced in Sec. 5.1.

Following the same argumentation as in the indistinguishable photons case, we may assume, due to the lack of global-phase reference [Bartlett *et al.*, 2007; Jarzyna and Demkowicz-Dobrzański, 2012; Mølmer, 1997], the general *seed element* of a covariant POVM to still have a block-diagonal structure, but now w.r.t. different patterns of surviving photons, i.e. $\Xi^N = \bigoplus_{\boldsymbol{N}'=\boldsymbol{0}^N}^{\boldsymbol{1}^N} \Xi^{\boldsymbol{N}'}$. Nevertheless, let us emphasise that—by the reasoning of App. D—for any general measurement $M_{\boldsymbol{i}}$ on the N photons, possessing now potentially 2^N outcomes: $\boldsymbol{i} = i_1 i_2 \ldots i_N$, a covariant POVMs constructed with help of the above Ξ^N can always be found that achieves the same precision of estimation. Moreover, we may write each block—parametrised by a particular pattern—of the seed element in a basis: $\Xi^{\boldsymbol{N}'} = \sum_{\boldsymbol{n}',\boldsymbol{m}'=\boldsymbol{0}^{N'}}^{\boldsymbol{1}^{N'}} \Xi_{\boldsymbol{n}',\boldsymbol{m}'}^{\boldsymbol{N}'} |\boldsymbol{n}'\rangle\langle\boldsymbol{m}'|$, in which $\boldsymbol{n}'$ stands for a string with N' bits placed at positions corresponding to 1's in $\boldsymbol{N}'$ with complementary positions left empty (neither 0 nor 1).

Adapting Eq. (5.27) to the distinguishable-photons case and, for completeness, assuming a general form of a circularly symmetric cost function (3.38), we obtain (after conveniently setting $\mathbf{0} \equiv \mathbf{0}^N$, $\mathbf{1} \equiv \mathbf{1}^N$):

$$\langle \mathcal{C} \rangle = c_0 + \sum_{\substack{\boldsymbol{n},\boldsymbol{m}=\mathbf{0} \\ \boldsymbol{n} \neq \boldsymbol{m}}}^{\mathbf{1}} \sum_{\boldsymbol{l}_a=\mathbf{0}}^{\min(\boldsymbol{n},\boldsymbol{m})} \sum_{\boldsymbol{l}_b=\mathbf{0}}^{\mathbf{1}-\max(\boldsymbol{n},\boldsymbol{m})} \frac{c_{|n-m|}}{2} \sqrt{B_n^{(l_a,l_b)} B_m^{(l_a,l_b)}} \, \alpha_{\boldsymbol{n}}^* \alpha_{\boldsymbol{m}} \, \Xi_{\boldsymbol{n} \smallsetminus (\boldsymbol{l}_a+\boldsymbol{l}_b),\boldsymbol{m} \smallsetminus (\boldsymbol{l}_a+\boldsymbol{l}_b)}^{\mathbf{1}-(\boldsymbol{l}_a+\boldsymbol{l}_b)} \,, \tag{J.2}$$

which for the cost function (3.40), C_H, utilised in the main text must be substituted $c_0 = -c_1 = 2$ and $\forall_{k>1}: c_k = 0$ (see Sec. 3.1.3.3). The min, max in Eq. (J.2) should be understood as bitwise operations and $B_n^{(l_a,l_b)}$ are the non-combinatorial constituents of the binomial coefficients (5.3), so that $b_n^{(l_a,l_b)} = \binom{n}{l_a}\binom{N-n}{l_b} B_n^{(l_a,l_b)}$. For compactness, we have introduced above the notation, in which $\boldsymbol{x} \smallsetminus \boldsymbol{y}$ represents a binary string $\boldsymbol{x}$ with empty entries at positions corresponding to 1's in $\boldsymbol{y}$.

We now split each of the sums over $\boldsymbol{l}_{a/b}$ in Eq. (J.2) into a sum over $l_{a/b}$, i.e. number of 1's in $\boldsymbol{l}_{a/b}$ respectively, and a sum over all permutations of 1's within $\boldsymbol{l}_{a/b}$. We proceed analogously for summations over $\boldsymbol{n}$ and $\boldsymbol{m}$ obtaining

$$\langle \mathcal{C} \rangle = c_0 + \sum_{\substack{n,m=0 \\ n \neq m}}^{N} \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} \frac{c_{|n-m|}}{2} \sqrt{B_n^{(l_a,l_b)} B_m^{(l_a,l_b)}} \times \tag{J.3}$$

$$\times \sum_{\substack{\boldsymbol{n}=\mathbf{0} \\ |\boldsymbol{n}|=n}}^{\mathbf{1}} \sum_{\substack{\boldsymbol{m}=\mathbf{0} \\ |\boldsymbol{m}|=m}}^{\mathbf{1}} \alpha_{\boldsymbol{n}}^* \alpha_{\boldsymbol{m}} \sum_{\substack{\boldsymbol{l}_a=\mathbf{0} \\ |\boldsymbol{l}_a|=l_a}}^{\min(\boldsymbol{n},\boldsymbol{m})} \sum_{\substack{\boldsymbol{l}_b=\mathbf{0} \\ |\boldsymbol{l}_b|=l_b}}^{\mathbf{1}-\max(\boldsymbol{n},\boldsymbol{m})} \Xi_{\boldsymbol{n} \smallsetminus (\boldsymbol{l}_a+\boldsymbol{l}_b),\boldsymbol{m} \smallsetminus (\boldsymbol{l}_a+\boldsymbol{l}_b)}^{\mathbf{1}-(\boldsymbol{l}_a+\boldsymbol{l}_b)} \,.$$

In order to proceed further let us for the moment return to the lossless scenario ($\eta_a = \eta_b = 1$), in which the above formula simplifies to:

$$\langle \mathcal{C} \rangle = c_0 + \sum_{\substack{n,m=0 \\ n \neq m}}^{N} \frac{c_{|n-m|}}{2} \sum_{\substack{\boldsymbol{n}=\mathbf{0} \\ |\boldsymbol{n}|=n}}^{\mathbf{1}} \sum_{\substack{\boldsymbol{m}=\mathbf{0} \\ |\boldsymbol{m}|=m}}^{\mathbf{1}} \alpha_{\boldsymbol{n}}^* \alpha_{\boldsymbol{m}} \Xi_{\boldsymbol{n},\boldsymbol{m}}^{\mathbf{1}} \,. \tag{J.4}$$

Recall (see Sec. 3.2.3) that $\Xi^{\mathbf{1}}$ needs to be a positive semi-definite operator, and by the completeness constraint has to also fulfil: $\Xi_{\boldsymbol{m},\boldsymbol{n}}^{\mathbf{1}} \underset{n=m}{=} \delta_{\boldsymbol{m},\boldsymbol{n}}$. Since the diagonal blocks of $\Xi^{\mathbf{1}}$ (corresponding to $n=m$) must thus correspond to identity matrices, this implies that none of the off-diagonal blocks of $\Xi^{\mathbf{1}}$ (corresponding to $n \neq m$) can have a singular value larger than 1. Such fact can be proven as follows.

Let us assume that for certain block $[n,m]$ with $n \neq m$, the largest singular value $\lambda > 1$ and let $|\boldsymbol{v}_n\rangle$, $|\boldsymbol{w}_m\rangle$ ($|\boldsymbol{v}_n| = n$, $|\boldsymbol{w}_m| = m$) be the normalised left and right singular vectors corresponding to the singular value λ. Defining $|\boldsymbol{z}\rangle = |\boldsymbol{v}_n\rangle - |\boldsymbol{w}_m\rangle$, we calculate

$$\langle \boldsymbol{z}|\Xi^{\mathbf{1}}|\boldsymbol{z}\rangle = \langle \boldsymbol{v}_n|\Xi^{\mathbf{1}}|\boldsymbol{v}_n\rangle + \langle \boldsymbol{w}_m|\Xi^{\mathbf{1}}|\boldsymbol{w}_m\rangle - 2\,\mathrm{Re}\big\{\langle \boldsymbol{v}_n|\Xi^{\mathbf{1}}|\boldsymbol{w}_m\rangle\big\} = 2(1-\lambda) < 0, \tag{J.5}$$

what contradicts the positivity semi-definiteness of $\Xi^{\mathbf{1}}$ and proves that λ can never be greater than 1. ∎

Crucially, as all the singular values of any $[n,m]$ block of $\Xi^{\mathbf{1}}$ cannot exceed one, the following inequality must hold: $\sum_{\substack{\boldsymbol{n}=\mathbf{0} \\ |\boldsymbol{n}|=n}}^{\mathbf{1}} \sum_{\substack{\boldsymbol{m}=\mathbf{0} \\ |\boldsymbol{m}|=m}}^{\mathbf{1}} \alpha_{\boldsymbol{n}}^* \alpha_{\boldsymbol{m}} \Xi_{\boldsymbol{n},\boldsymbol{m}}^{\mathbf{1}} \leq \alpha_n^* \alpha_m$, where $\alpha_n = \sqrt{\sum_{\substack{\boldsymbol{n}=\mathbf{0} \\ |\boldsymbol{n}|=n}}^{\mathbf{1}} |\alpha_{\boldsymbol{n}}|^2}$ may then be interpreted as the symmetric coefficients built for a given set of $\alpha_{\boldsymbol{n}}$ that parametrise an input state (5.1), $|\psi_{\mathrm{in}}^N\rangle$, consisting of indistinguishable photons. Hence, we may always lower-bound the average cost (J.4) via:

$$\langle \mathcal{C} \rangle \geq c_0 + \sum_{\substack{n,m=0 \\ n \neq m}}^{N} \frac{c_{|n-m|}}{2} \alpha_n^* \alpha_m, \tag{J.6}$$

where the r.h.s. represents the general average cost optimised over measurements for the input $|\psi_{\text{in}}^N\rangle$, i.e. the generalised version of the minimal $\langle\mathcal{C}_H\rangle$ (3.83) obtained for the lossless interferometer assuming the cost function (3.40). In other words, Eq. (J.6) proves that in the absence of losses, for any given input state consisting of distinguishable photons, there always exists a state $|\psi_{\text{in}}^N\rangle$ that performs at least as good, so one may always restrict to inputs (5.1) without sacrificing optimality.

Returning to Eq. (J.3), we realise that a similar argumentation can be applied after acknowledging the positive semi-definiteness of the operator: $\Xi_{\boldsymbol{m},\boldsymbol{n}}^{(l_a,l_b)} = \sum_{\substack{\boldsymbol{l_a}=\boldsymbol{0}\\|\boldsymbol{l_a}|=l_a}}^{\min(\boldsymbol{n},\boldsymbol{m})} \sum_{\substack{\boldsymbol{l_b}=\boldsymbol{0}\\|\boldsymbol{l_b}|=l_b}}^{\boldsymbol{1}-\max(\boldsymbol{n},\boldsymbol{m})} \Xi_{\boldsymbol{m}\smallsetminus(\boldsymbol{l_a}+\boldsymbol{l_b}),\boldsymbol{n}\smallsetminus(\boldsymbol{l_a}+\boldsymbol{l_b})}^{\boldsymbol{1}-(\boldsymbol{l_a}+\boldsymbol{l_b})}$. We notice that the completeness constraint again implies a block structure of $\Xi^{(l_a,l_b)}$ w.r.t. $[\boldsymbol{n},\boldsymbol{m}]$, with the diagonal blocks $[\boldsymbol{n},\boldsymbol{n}]$ corresponding now to $\sum_{\substack{\boldsymbol{l_a}=\boldsymbol{0}\\|\boldsymbol{l_a}|=l_a}}^{\boldsymbol{n}} \sum_{\substack{\boldsymbol{l_b}=\boldsymbol{0}\\|\boldsymbol{l_b}|=l_b}}^{\boldsymbol{1}-\boldsymbol{n}} 1 = \binom{n}{l_a}\binom{N-n}{l_b}$ times the identity matrix.

Consequently, the maximum singular value of any non-diagonal block, $[\boldsymbol{n},\boldsymbol{m}]$ with $n\neq m$, of $\Xi^{(l_a,l_b)}$ is upper-limited by $\binom{\min(n,m)}{l_a}\binom{N-\max(n,m)}{l_b}$. As a result, we obtain the following lower bound on the general average cost (J.3):

$$\langle\mathcal{C}\rangle \geq c_0 + \sum_{\substack{n,m=0\\n\neq m}}^{N} \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} \frac{c_{|n-m|}}{2}\sqrt{B_n^{(l_a,l_b)}B_m^{(l_a,l_b)}}\binom{\min(n,m)}{l_a}\binom{N-\max(n,m)}{l_b}\alpha_n^\star\alpha_m$$

$$\geq c_0 + \sum_{\substack{n,m=0\\n\neq m}}^{N} \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} \frac{c_{|n-m|}}{2}\sqrt{b_n^{(l_a,l_b)}b_m^{(l_a,l_b)}}\,|\alpha_n||\alpha_m|, \tag{J.7}$$

which is just the generalisation of the minimal $\langle\mathcal{C}_H\rangle$ (5.27) evaluated for $|\psi_{\text{in}}^N\rangle$ and the lossy interferometer of Fig. 5.1 to all valid cost functions (3.38)[1]. Thus, the inequality in Eq. (J.7) completes the proof that the optimal phase estimation in the presence of losses is indeed achievable within the Bayesian approach when considering only the input states (5.1) consisting of indistinguishable photons. ∎

Adaptive measurement schemes

Let us describe the general structure of *adaptive measurement schemes* performed on N distinguishable photons, e.g. arriving in consecutive time-bins. Let $\{M_{i_1}^{(1)}\}_{i_1}$ be a POVM performed on the first photon. Then, depending on the measurement result i_1 a POVM $\{M_{i_2}^{(2)}(i_1)\}_{i_2}$ is performed on the second one. In general, a POVM performed on the k-th photon $\{M_{i_k}^{(k)}(i_1,\ldots,i_{k-1})\}_{i_k}$ depends on all previous measurement outcomes. Thus, the adaptive measurement mathematically corresponds to an overall POVM:

$$M_{\boldsymbol{i}} = M_{i_1}^{(1)} \otimes M_{i_2}^{(2)}(i_1) \otimes M_{i_3}^{(3)}(i_1,i_2) \otimes \cdots \otimes M_{i_N}^{(N)}(i_1,\ldots,i_{N-1}), \tag{J.8}$$

which, on the other hand, can be treated as an instance of a *global*—acting on all the photons—POVM with potentially 2^N measurement outcomes indexed by $\boldsymbol{i} = i_1 i_2 \ldots i_N$. Such an observation proves that the optimisation of a given quantum estimation problem over *all global* POVMs, $M_{\boldsymbol{i}}$, naturally covers also the case of adaptive measurements described in [Wiseman *et al.*, 2009; Wiseman and Killip, 1997, 1998]. Therefore, as we have proved above that the ultimate Bayesian bounds on precision, derived for the lossy interferometry scheme after restricting to bosonic input states, apply equivalently to the scenarios employing distinguishable photons and their most general global measurements, these global bounds on precision must also apply to *all* adaptive measurement strategies. ∎

---

[1]Such a fact becomes completely evident after exchanging the order of the sums in Eq. (J.7), so that $\sum_{\substack{n,m=0\\n\neq m}}^{N} \sum_{l_a=0}^{\min(n,m)} \sum_{l_b=0}^{N-\max(n,m)} \equiv \sum_{l_a=0}^{N} \sum_{l_b=0}^{N-l_a} \sum_{\substack{n,m=l_a\\n\neq m}}^{N-l_b}$, and re-parametrising l_b to $N' = N - l_a - l_b$.

Bibliography

Acín, A., D. Bruß, M. Lewenstein, and A. Sanpera (2001), "*Classification of mixed three-qubit states,*" Phys. Rev. Lett. **87**, 040401.

Addis, C., B. Bylicka, D. Chruściński, and S. Maniscalco (2014), "*Comparative study of non-Markovianity measures in exactly solvable one- and two-qubit models,*" Phys. Rev. A **90**, 052103.

Adesso, G., F. Dell'Anno, S. D. Siena, F. Illuminati, and L. A. M. Souza (2009), "*Optimal estimation of losses at the ultimate quantum limit with non-Gaussian states,*" Phys. Rev. A **79**, 040305(R).

Aharonov, Y., S. Massar, and S. Popescu (2002), "*Measuring energy, estimating Hamiltonians, and the time-energy uncertainty relation,*" Phys. Rev. A **66**, 052107.

Altland, A., and B. D. Simons (2010), *Condensed Matter Field Theory*, 2nd ed. (Cambridge University Press).

Amari, S.-I., O. E. Barndorff-Nielsen, R. Kass, S. Lauritzen, and C. Rao (1987), *Differential Geometry in Statistical Inference* (JSTOR).

Amari, S.-I., and H. Nagaoka (2007), *Methods of Information Geometry*, Vol. 191 (American Mathematical Society).

Anandan, J., and Y. Aharonov (1990), "*Geometry of quantum evolution,*" Phys. Rev. Lett. **65**, 1697.

Andersson, E., J. D. Cresser, and M. J. W. Hall (2007), "*Finding the Kraus decomposition from a master equation and vice versa,*" J. Mod. Opt. **54**, 1695.

Andrè, A., and M. Lukin (2002), "*Atom correlations and spin squeezing near the Heisenberg limit: Finite-size effect and decoherence,*" Phys. Rev. A **65**, 053819.

Andrè, A., A. S. Sørensen, and M. D. Lukin (2004), "*Stability of atomic clocks based on entangled atoms,*" Phys. Rev. Lett. **92**, 230801.

Anisimov, P. M., G. M. Raterman, A. Chiruvelli, W. N. Plick, S. D. Huver, H. Lee, and J. P. Dowling (2010), "*Quantum metrology with two-mode squeezed vacuum: Parity detection beats the Heisenberg limit,*" Phys. Rev. Lett. **104**, 103602.

Appel, J., P. J. Windpassinger, D. Oblak, U. B. Hoff, N. Kjærgaard, and E. S. Polzik (2009), "*Mesoscopic atomic entanglement for precision measurements beyond the standard quantum limit,*" Proc. Natl. Acad. Sci. U.S.A. **106**, 10960.

Arndt, C. (2001), *Information Measures: Information and its Description in Science and Engineering* (Springer).

Banaszek, K., R. Demkowicz-Dobrzański, and I. A. Walmsley (2009), "*Quantum states made to measure,*" Nature Photon. **3**, 673.

Barlow, R. (2013), *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (Wiley).

Barndorff-Nielsen, O. E., and R. D. Gill (2000), "*Fisher information in quantum statistics,*" J. Phys. A: Math. Gen. **30**, 4481.

Bartlett, S. D., T. Rudolph, and R. W. Spekkens (2007), "*Reference frames, superselection rules, and quantum information,*" Rev. Mod. Phys. **79**, 555.

Benatti, F., R. Floreanini, and K. Titimbo (2014), "*Entanglement of identical particles,*" Open Syst. Inf. Dyn. **21**, 1440003.

Bengtsson, I., and K. Życzkowski (2006), *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press).

Bennett, C. H., D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters (1999a), "*Quantum nonlocality without entanglement,*" Phys. Rev. A **59**, 1070.

Bennett, C. H., D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal (1999b), "*Unextendible product bases and bound entanglement,*" Phys. Rev. Lett. **82**, 5385.

Berry, D. W., M. J. W. Hall, M. Zwierz, and H. M. Wiseman (2012), "*Optimal Heisenberg-style bounds for the average performance of arbitrary phase estimates,*" Phys. Rev. A **86**, 053813.

Berry, D. W., B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman (2009), "*How to perform the most accurate possible phase measurements,*" Phys. Rev. A **80**, 052114.

Berry, D. W., and H. M. Wiseman (2000), "*Optimal states and almost optimal adaptive measurements for quantum interferometry,*" Phys. Rev. Lett. **85**, 5098.

Biedenharn, L. C., and J. D. Louck (1981), *Angular Momentum in Quantum Physics* (Addison Wesley).

Bollinger, J. J., W. M. Itano, D. J. Wineland, and D. J. Heinzen (1996), "*Optimal frequency measurements with maximally correlated states,*" Phys. Rev. A **54**, R4649.

Bondurant, R. S., and J. H. Shapiro (1984), "*Squeezed states in phase-sensing interferometers,*" Phys. Rev. D **30**, 2548.

Borregaard, J., and A. S. Sørensen (2013), "*Near-Heisenberg-limited atomic clocks in the presence of decoherence,*" Phys. Rev. Lett. **111**, 090801.

Branciard, C. (2013), "*Error-tradeoff and error-disturbance relations for incompatible quantum measurements,*" Proc. Natl. Acad. Sci. U.S.A. **110** (17), 6742.

Braunstein, S. L., and C. M. Caves (1994), "*Statistical distance and the geometry of quantum states,*" Phys. Rev. Lett. **72**, 3439.

Braunstein, S. L., C. M. Caves, and G. Milburn (1996), "*Generalized uncertainty relations: Theory, examples, and Lorentz invariance,*" Ann. Phys. **247**, 135.

Breuer, H.-P., and F. Petruccione (2002), *The Theory of Open Quantum Systems* (Oxford University Press).

Brunner, N., D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner (2014), "*Bell nonlocality,*" Rev. Mod. Phys. **86**, 419.

Bruß, D., and G. Leuchs, Eds. (2007), *Lectures on quantum information* (Wiley-VCH).

Busch, P., P. Lahti, and R. F. Werner (2013), "*Proof of Heisenberg's error-disturbance relation,*" Phys. Rev. Lett. **111**, 160405.

Bužek, V., R. Derka, and S. Massar (1999), "*Optimal quantum clocks,*" Phys. Rev. Lett. **82**, 2207.

del Campo, A., I. L. Egusquiza, M. B. Plenio, and S. F. Huelga (2013), "*Quantum speed limits in open system dynamics*," Phys. Rev. Lett. **110**, 050403.

Campos, R. A., B. E. A. Saleh, and M. C. Teich (1989), "*Quantum-mechanical lossless beam splitter: SU(2) symmetry and photon statistics*," Phys. Rev. A **40**, 1371.

Caves, C. M. (1981), "*Quantum-mechanical noise in an interferometer*," Phys. Rev. D **23**, 1693.

Chaves, R., J. B. Brask, M. Markiewicz, J. Kołodyński, and A. Acín (2013), "*Noisy metrology beyond the standard quantum limit*," Phys. Rev. Lett. **111**, 120401.

Chin, A. W., S. F. Huelga, and M. B. Plenio (2012), "*Quantum metrology in non-Markovian environments*," Phys. Rev. Lett. **109**, 233601.

Chiribella, G., G. M. D'Ariano, and M. F. Sacchi (2005), "*Optimal estimation of group transformations using entanglement*," Phys. Rev. A **72**, 042338.

Chiruvelli, A., and H. Lee (2011), "*Parity measurements in quantum optical metrology*," J. Mod. Opt. **58** (11), 945.

Choi, M.-D. (1975), "*Completely positive linear maps on complex matrices*," Linear Algebr. Appl. **10**, 285.

Chruściński, D., and A. Kossakowski (2010), "*Non-Markovian quantum dynamics: Local versus nonlocal*," Phys. Rev. Lett. **104**, 070406.

Crowley, P. J. D., A. Datta, M. Barbieri, and I. A. Walmsley (2014), "*Tradeoff in simultaneous quantum-limited phase and loss estimation in interferometry*," Phys. Rev. A **89**, 023845.

Czekaj, Ł., A. Przysiężna, M. Horodecki, and P. Horodecki (2014), "*Quantum metrology: Heisenberg limit with bound entanglement*," arXiv e-print, 1403.5867.

Czichos, H., T. Saito, and L. E. Smith (2011), *Springer Handbook of Metrology and Testing* (Springer).

D'Ariano, G. M., P. Lo Presti, and M. G. A. Paris (2001), "*Using entanglement improves the precision of quantum measurements*," Phys. Rev. Lett. **87**, 270404.

D'Ariano, G. M., M. F. Sacchi, and J. Kahn (2005), "*Minimax discrimination of two Pauli channels*," Phys. Rev. A **72**, 052302.

Deffner, S., and E. Lutz (2013a), "*Energy-time uncertainty relation for driven quantum systems*," J. Phys. A: Math. Theor. **46**, 335302.

Deffner, S., and E. Lutz (2013b), "*Quantum speed limit for non-Markovian dynamics*," Phys. Rev. Lett. **111**, 010402.

Demkowicz-Dobrzański, R. (2011), "*Optimal phase estimation with arbitrary a priori knowledge*," Phys. Rev. A **83**, 061802(R).

Demkowicz-Dobrzański, R., K. Banaszek, and R. Schnabel (2013), "*Fundamental quantum interferometry bound for the squeezed-light-enhanced gravitational wave detector GEO 600*," Phys. Rev. A **88**, 041802(R).

Demkowicz-Dobrzański, R., U. Dorner, B. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley (2009), "*Quantum phase estimation with lossy interferometers*," Phys. Rev. A **80**, 013825.

Demkowicz-Dobrzański, R., M. Jarzyna, and J. Kołodyński (2015), "*Quantum limits in optical interferometry*," in *Progress in Optics*, Vol. 60, edited by E. Wolf (Elsevier), *in press*, arXiv:1405.7703 [quant-ph] .

Demkowicz-Dobrzański, R., J. Kołodyński, and M. Guţă (2012), "*The elusive Heisenberg limit in quantum-enhanced metrology*," Nat. Commun. **3**, 1063.

Demkowicz-Dobrzański, R., and L. Maccone (2014), "*Using entanglement against noise in quantum metrology*," Phys. Rev. Lett. **113**, 250801.

Derka, R., V. Bužek, and A. K. Ekert (1998), "*Universal algorithm for optimal estimation of quantum states from finite ensembles via realizable generalized measurement*," Phys. Rev. Lett. **80**, 1571.

Devanathan, V. (1999), *Angular Momentum Techniques in Quantum Mechanics* (Springer).

Dicke, R. (1954), "*Coherence in spontaneous radiation processes*," Phys. Rev. **93** (1), 99.

Dorner, U. (2012), "*Quantum frequency estimation with trapped ions and atoms*," New J. Phys. **14**, 043011.

Dorner, U., R. Demkowicz-Dobrzański, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley (2009), "*Optimal quantum phase estimation*," Phys. Rev. Lett. **102**, 040403.

Dowling, J. P. (1998), "*Correlated input-port, matter-wave interferometer: Quantum-noise limits to the atom-laser gyroscope*," Phys. Rev. A **57**, 4736.

Dür, W., M. Skotiniotis, F. Fröwis, and B. Kraus (2014), "*Improved quantum metrology using quantum error correction*," Phys. Rev. Lett. **112**, 080801.

Einstein, A., B. Podolsky, and N. Rosen (1935), "*Can quantum-mechanical description of physical reality be considered complete?*" Phys. Rev. **47**, 777.

Englert, B.-G. (2013), "*On quantum theory*," Eur. Phys. J. D **67**, 238.

Escher, B. M., L. Davidovich, N. Zagury, and R. L. de Matos Filho (2012), "*Quantum metrological limits via a variational approach*," Phys. Rev. Lett. **109**, 190404.

Escher, B. M., R. L. de Matos Filho, and L. Davidovich (2011), "*General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology*," Nature Phys. **7**, 406.

Everitt, B., and A. Skrondal (2010), *The Cambridge Dictionary of Statistics* (Cambridge University Press).

Frey, M., D. Collins, and K. Gerlach (2011), "*Probing the qudit depolarizing channel*," J. Phys. A: Math. Theor. **44**, 205306.

Fröwis, F. (2012), "*Kind of entanglement that speeds up quantum evolution*," Phys. Rev. A **85**, 052127.

Fujiwara, A. (1994), "*Multi-parameter pure state estimation based on the right-logarithmic derivative*," Math. Eng. Tech. Rep. **94**, 9.

Fujiwara, A. (2001), "*Quantum channel identification problem*," Phys. Rev. A **63**, 042304.

Fujiwara, A. (2004), "*Estimation of a generalized amplitude-damping channel*," Phys. Rev. A **70**, 012317.

Fujiwara, A., and H. Imai (2003), "*Quantum parameter estimation of a generalized Pauli channel*," J. Phys. A: Math. Gen. **36**, 8093.

Fujiwara, A., and H. Imai (2008), "*A fibre bundle over manifolds of quantum channels and its application to quantum statistics*," J. Phys. A: Math. Theor. **41**, 255304.

Gardiner, C., and P. Zoller (2000), *Quantum Noise* (Springer).

Genoni, M. G., S. Olivares, D. Brivio, S. Cialdi, D. Cipriani, A. Santamato, S. Vezzoli, and M. G. Paris (2012), "*Optical interferometry in the presence of large phase diffusion*," Phys. Rev. A **85**, 043817.

Genoni, M. G., S. Olivares, and M. G. Paris (2011), "*Optical phase estimation in the presence of phase diffusion*," Phys. Rev. Lett. **106**, 153603.

Genoni, M. G., M. G. A. Paris, G. Adesso, H. Nha, P. L. Knight, and M. S. Kim (2013), "*Optimal estimation of joint parameters in phase space*," Phys. Rev. A **87**, 012107.

Gerry, C. C., and J. Mimih (2010), "*The parity operator in quantum optical metrology*," Contemp. Phys. **51**, 497.

Gill, R. D., and M. I. Guţă (2013), "*On asymptotic quantum statistical inference*," IMS Collections **9**, 105.

Gill, R. D., and B. Y. Levit (1995), "*Applications of the van Trees inequality: a Bayesian Cramér-Rao bound*," Bernoulli **1(1/2)**, 59.

Giovannetti, V., S. Lloyd, and L. Maccone (2001), "*Quantum-enhanced positioning and clock synchronization*," Nature **412**, 417.

Giovannetti, V., S. Lloyd, and L. Maccone (2003a), "*Quantum limits to dynamical evolution*," Phys. Rev. A **67**, 052109.

Giovannetti, V., S. Lloyd, and L. Maccone (2003b), "*The role of entanglement in dynamical evolution*," Europhys. Lett. **62**, 615.

Giovannetti, V., S. Lloyd, and L. Maccone (2004), "*Quantum-enhanced measurements: Beating the Standard Quantum Limit*," Science **306**, 1330.

Giovannetti, V., S. Lloyd, and L. Maccone (2006), "*Quantum metrology*," Phys. Rev. Lett. **96**, 010401.

Giovannetti, V., S. Lloyd, and L. Maccone (2011), "*Advances in quantum metrology*," Nature Photon. **5**, 222.

Giovannetti, V., S. Lloyd, and L. Maccone (2012), "*Quantum measurement bounds beyond the uncertainty relations*," Phys. Rev. Lett. **108**, 260405.

Giovannetti, V., and L. Maccone (2012), "*Sub-Heisenberg estimation strategies are ineffective*," Phys. Rev. Lett. **108**, 210404.

Gorini, V., A. Kossakowski, and E. C. G. Sudarshan (1976), "*Completely positive dynamical semigroups of N-level systems*," J. Math. Phys. **17**, 821.

Grant, M., and S. Boyd (2012), "*CVX: Matlab software for disciplined convex programming,*" .

Greenberger, D. M., M. Horne, and A. Zeilinger (1989), "*Going beyond Bells theorem*," in *Bells Theorem, Quantum Theory, and Conceptions of the Universe*, Fundamental Theories of Physics, Vol. 37, edited by M. Kafatos (Springer Netherlands) pp. 69–72.

Griffiths, D. (2013), *Introduction to Quantum Mechanics* (Pearson Education, Limited).

Guţă, M., and A. Jenčová (2007), "*Local asymptotic normality in quantum statistics*," Commun. Math. Phys. **276**, 341.

Guţă, M., and J. Kahn (2006), "*Local asymptotic normality for qubit states*," Phys. Rev. A **73**, 052108.

Hall, B. C. (2013), *Quantum Theory for Mathematicians* (Springer).

Hall, M. J. W., D. W. Berry, M. Zwierz, and H. M. Wiseman (2012), "*Universality of the Heisenberg limit for estimates of random phase shifts*," Phys. Rev. A **85**, 041802.

Hall, M. J. W., and H. M. Wiseman (2012), "*Heisenberg-style bounds for arbitrary estimates of shift parameters including prior information*," New J. Phys. **14**, 033040.

Hammerer, K., A. S. Sørensen, and E. S. Polzik (2010), "*Quantum interface between light and atomic ensembles,*" Rev. Mod. Phys. **82**, 1041.

Hariharan, P. (2003), *Optical Interferometry* (Elsevier, Amsterdam).

Haroche, S. (2013), "*Nobel lecture: Controlling photons in a box and exploring the quantum to classical boundary,*" Rev. Mod. Phys. **85**, 1083.

Hayashi, M., Ed. (2005), *Asymptotic Theory of Quantum Statistical Inference* (World Scientific).

Hayashi, M. (2011), "*Comparison between the Cramér-Rao and the Mini-max approaches in quantum channel estimation,*" Comm. Math. Phys. **304**, 689.

Helstrom, C. W. (1976), *Quantum Detection and Estimation Theory* (Academic Press).

Higgins, B. L., D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde (2009), "*Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements,*" New J. Phys. **11**, 073023.

Higgins, B. L., D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde (2007), "*Entanglement-free heisenberg-limited phase estimation,*" Nature **450**, 393.

Hofmann, H. F. (2009), "*All path-symmetric pure states achieve their maximal phase sensitivity in conventional two-path interferometry,*" Phys. Rev. A **79**, 033822.

Holevo, A. S. (1982), *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland).

Holland, M. J., and K. Burnett (1993), "*Interferometric detection of optical phase shifts at the Heisenberg limit,*" Phys. Rev. Lett. **71**, 1355.

Horodecki, R., P. Horodecki, M. Horodecki, and K. Horodecki (2009), "*Quantum entanglement,*" Rev. Mod. Phys. **81**, 865.

Hotta, M., T. Karasawa, and M. Ozawa (2005), "*Ancilla-assisted enhancement of channel estimation for low-noise parameters,*" Phys. Rev. A **72**, 052334.

Hotta, M., T. Karasawa, and M. Ozawa (2006), "*N-body-extended channel estimation for low-noise parameters,*" J. Phys. A: Math. Gen. **39**, 14465.

Hradil, Z., R. Myška, J. Peřina, M. Zawisky, Y. Hasegawa, and H. Rauch (1996), "*Quantum phase in interferometry,*" Phys. Rev. Lett. **76**, 4295.

Huelga, S. F., C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac (1997), "*Improvement of frequency standards with quantum entanglement,*" Phys. Rev. Lett. **79**, 3865.

Humphreys, P. C., M. Barbieri, A. Datta, and I. A. Walmsley (2013), "*Quantum enhanced multiple phase estimation,*" Phys. Rev. Lett. **111**, 070403.

Hyllus, P., O. Gühne, and A. Smerzi (2010a), "*Not all pure entangled states are useful for sub-shot-noise interferometry,*" Phys. Rev. A **82**, 012337.

Hyllus, P., W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi (2012), "*Fisher information and multiparticle entanglement,*" Phys. Rev. A **85**, 022321.

Hyllus, P., L. Pezzé, and A. Smerzi (2010b), "*Entanglement and sensitivity in precision measurements with states of a fluctuating number of particles,*" Phys. Rev. Lett. **105**, 120501.

Jamiołkowski, A. (1972), "*Linear transformations which preserve trace and positive semidefiniteness of operators,*" Rep. Math. Phys **3**, 275.

Jarzyna, M., and R. Demkowicz-Dobrzański (2012), "*Quantum interferometry with and without an external phase reference*," Phys. Rev. A **85**, 011801(R).

Jarzyna, M., and R. Demkowicz-Dobrzański (2013), "*Matrix product states for quantum metrology*," Phys. Rev. Lett. **110**, 240405.

Jarzyna, M., and R. Demkowicz-Dobrzański (2014), "*True precision limits in quantum metrology*," arXiv e-print, 1407.4805.

Jensen, J. L. W. V. (1906), "*Sur les fonctions convexes et les inégalités entre les valeurs moyennes*," Acta Mathematica **30**, 175.

Jeske, J., J. H. Cole, and S. Huelga (2013), "*Quantum metrology in the presence of spatially correlated noise: Restoring Heisenberg scaling*," arXiv e-print , 1307.6301.

Ji, Z., G. Wang, R. Duan, Y. Feng, and M. Ying (2008), "*Parameter estimation of quantum channels*," IEEE Trans. Inf. Theory **54**, 5172.

Jordan, P. (1935), "*Der Zusammenhang der symmetrischen und linearen Gruppen und das Mehrkörperproblem*," Z. Phys. **94** (7-8), 531.

Kacprowicz, M., R. Demkowicz-Dobrzański, W. Wasilewski, and K. Banaszek (2010), "*Experimental quantum enhanced phase-estimation in the presence of loss*," Nature Photon. **4**, 357.

Kahn, J., and M. Guţă (2009), "*Local asymptotic normality for finite dimensional quantum systems*," Commun. Math. Phys. **289**, 597.

Kay, S. M. (1993), *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall).

Keyl, M., and R. F. Werner (2007), "*Channels and maps*," in *Lectures on Quantum Information*, edited by D. Bruß and G. Leuchs, Chap. 5 (Wiley-VCH) pp. 73–86.

Killoran, N., M. Cramer, and M. B. Plenio (2014), "*Extracting entanglement from identical particles*," Phys. Rev. Lett. **112**, 150501.

Knysh, S. I., E. H. Chen, and G. A. Durkin (2014), "*True limits to precision via unique quantum probe*," arXiv e-print, 1402.0495.

Knysh, S. I., V. N. Smelyanskiy, and G. A. Durkin (2011), "*Scaling laws for precision in quantum interferometry and the bifurcation landscape of the optimal state*," Phys. Rev. A **83**, 021804.

Kok, P., H. Lee, and J. P. Dowling (2002), "*Creation of large-photon-number path entanglement conditioned on photodetection*," Phys. Rev. A **65**, 052104.

Kołodyński, J., and R. Demkowicz-Dobrzański (2010), "*Phase estimation without a priori phase knowledge in the presence of loss*," Phys. Rev. A **82**, 053804.

Kołodyński, J., and R. Demkowicz-Dobrzański (2013), "*Efficient tools for quantum metrology with uncorrelated noise*," New J. Phys. **15**, 073043.

Koschorreck, M., M. Napolitano, B. Dubost, and M. W. Mitchell (2010), "*Sub-projection-noise sensitivity in broadband atomic magnetometry*," Phys. Rev. Lett. **104**, 093602.

Kraus, K. (1983), *States, Effects, and Operations* (Springer-Verlag).

Lee, H., P. Kok, and J. P. Dowling (2002), "*A quantum Rosetta Stone for interferometry*," J. Mod. Opt. **49**, 2325.

Lehmann, E. L., and G. Casella (1998), *Theory of Point Estimation*, Vol. 31 (Springer).

Leibfried, D., M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland (2004), "*Toward Heisenberg-limited spectroscopy with multiparticle entangled states*," Science **304**, 1476.

Leibfried, D., R. Blatt, C. Monroe, and D. Wineland (2003), "*Quantum dynamics of single trapped ions*," Rev. Mod. Phys. **75**, 281.

Leroux, I. D., M. H. Schleier-Smith, and V. Vuletić (2010), "*Implementation of cavity squeezing of a collective atomic spin*," Phys. Rev. Lett. **104**, 073602.

LIGO Collaboration, (2011), "*A gravitational wave observatory operating beyond the quantum shot-noise limit*," Nature Phys. **7**, 962.

LIGO Collaboration, (2013), "*Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light*," Nature Photon. **7**, 613.

Lindblad, G. (1976), "*On the generators of quantum dynamical semigroups*," Commun. in Math. Phys. **48**, 119.

Louchet-Chauvet, A., J. Appel, J. J. Renema, D. Oblak, N. Kjaergaard, and E. S. Polzik (2010), "*Entanglement-assisted atomic clock beyond the projection noise limit*," New J. Phys. **12**, 065032.

Ma, J., X. Wang, C. Sun, and F. Nori (2011), "*Quantum spin squeezing*," Phys. Rep. **509**, 89.

Maccone, L. (2013), "*Intuitive reason for the usefulness of entanglement in quantum metrology*," Phys. Rev. A **88**, 042109.

Maccone, L., and V. Giovannetti (2011), "*Quantum metrology: Beauty and the noisy beast*," Nature Phys. **7**, 376.

Macieszczak, K. (2014), "*Upper bounds on the quantum Fisher information in the presence of general dephasing*," arXiv e-print, 1403.0955.

Macieszczak, K., M. Fraas, and R. Demkowicz-Dobrzański (2014), "*Bayesian quantum frequency estimation in presence of collective dephasing*," New J. Phys. **16** (11), 113002.

Mandelstam, L., and I. Tamm (1945), "*The uncertainty relation between energy and time in nonrelativistic quantum mechanics*," J. Phys. (USSR) **9** (249), 1.

Margolus, N., and L. B. Levitin (1998), "*The maximum speed of dynamical evolution*," Physica D (Amsterdam) **120**, 188.

Matsumoto, K. (2010), "*On metric of quantum channel spaces*," arXiv e-print, 1006.0300.

Matsuzaki, Y., S. C. Benjamin, and J. Fitzsimons (2011), "*Magnetic field sensing beyond the standard quantum limit under the effect of decoherence*," Phys. Rev. A **84**, 012103.

Mitchell, M. W., J. S. Lundeen, and A. M. Steinberg (2004), "*Super-resolving phase measurements with a multi-photon entangled state*," Nature **429**, 161.

Mølmer, K. (1997), "*Optical coherence: A convenient fiction*," Phys. Rev. A **55**, 3195.

Monras, A., and M. G. A. Paris (2007), "*Optimal quantum estimation of loss in bosonic channels*," Phys. Rev. Lett. **98**, 160401.

Nagaoka, H. (1989), "*A new approach to Cramér-Rao bounds for quantum state estimation*," IEICE Tech. Rep. **IT 89–42**, 9.

Nagaoka, H. (2005), "*On Fisher information of quantum statistical models*," in *Asymptotic Theory of Quantum Statistical Inference*, edited by M. Hayashi, Chap. 9 (World Scientific) pp. 113–125.

Nagata, T., R. Okamoto, J. L. O'Brien, K. Sasaki, and S. Takeuchi (2007), "*Beating the standard quantum limit with four entangled photons,*" Science **316**, 726.

Nair, R. (2011), "*Discriminating quantum-optical beam-splitter channels with number-diagonal signal states: Applications to quantum reading and target detection,*" Phys. Rev. A **84**, 032312.

Nair, R. (2012), "*Fundamental limits on the accuracy of optical phase estimation from rate-distortion theory,*" arXiv e-print, 1204.3761.

Napolitano, M., M. Koschorreck, B. Dubost, N. Behbood, R. Sewell, and M. W. Mitchell (2011), "*Interaction-based quantum metrology showing scaling beyond the Heisenberg limit,*" Nature **471** (7339), 486.

von Neumann, J. (1932), *Mathematische Grundlagen der Quantenmechanic* (Springer, Berlin).

Nielsen, M. A., and I. L. Chuang (2000), *Quantum Computing and Quantum Information* (Cambridge University Press).

Okamoto, R., H. F. Hofmann, T. Nagata, J. L. O'Brien, K. Sasaki, and S. Takeuchi (2008), "*Beating the standard quantum limit: phase super-sensitivity of N-photon interferometers,*" New J. Phys. **10**, 073033.

Ospelkaus, C., U. Warring, Y. Colombe, K. Brown, J. Amini, D. Leibfried, and D. Wineland (2011), "*Microwave quantum logic gates for trapped ions,*" Nature **476**, 181.

Ozawa, M. (2004), "*Uncertainty relations for noise and disturbance in generalized quantum measurements,*" Ann. Phys. **311** (2), 350.

Pegg, D. T., and S. M. Barnett (1988), "*Unitary phase operator in quantum mechanics,*" Europhys. Lett. **6**, 483.

Petz, D. (1996), "*Monotone metrics on matrix spaces,*" Linear Algebr. Appl. **244**, 81.

Petz, D. (2002), "*Covariance and Fisher information in quantum mechanics,*" J. Phys. A: Math. Gen. **35**, 929.

Petz, D., and C. Sudar (1999), "*Extending the Fisher metric to density matrices,*" in *Geometry in Present Day Science*, edited by O. E. Barndorff-Nielsen and E. B. V. Jensen (World Scientific) pp. 21–34.

Pezzé, L., and A. Smerzi (2009), "*Entanglement, nonlinear dynamics, and the Heisenberg limit,*" Phys. Rev. Lett. **102**, 100401.

Phillips, A. C. (2003), *Introduction to Quantum Mechanics*, edited by A. C. P. D. J. Sandiford, F. Mandl, The Manchester Physics Series (Wiley).

Pirandola, S. (2011), "*Quantum reading of a classical digital memory,*" Phys. Rev. Lett. **106**, 090504.

Pirandola, S., C. Lupo, V. Giovannetti, S. Mancini, and S. L. Braunstein (2011), "*Quantum reading capacity,*" New J. Phys. **13**, 113012.

Reed, M., and B. Simon (1981), *Functional Analysis*, Methods of Modern Mathematical Physics, Vol. I (Elsevier Science).

Resch, K. J., K. L. Pregnell, R. Prevedel, A. Gilchrist, G. J. Pryde, J. L. O'Brien, and A. G. White (2007), "*Time-reversal and super-resolving phase measurements,*" Phys. Rev. Lett. **98**, 223601.

Rivas, Á., S. F. Huelga, and M. B. Plenio (2014), "*Quantum non-Markovianity: Characterization, quantification and detection,*" Rep. Prog. Phys. **77**, 094001.

Robertson, H. (1934), "*An indeterminacy relation for several observables and its classical interpretation*," Phys. Rev. **46** (9), 794.

Roos, C., M. Chwalla, K. Kim, M. Riebe, and R. Blatt (2006), "*Designer atoms for quantum metrology*," Nature **443**, 316.

Sacchi, M. F. (2005), "*Optimal discrimination of quantum operations*," Phys. Rev. A **71**, 062340.

Sanders, B. C., and G. J. Milburn (1995), "*Optimal quantum measurements for phase estimation*," Phys. Rev. Lett. **75**, 2944.

Schmidt, P. O., T. Rosenband, C. Langer, W. Itano, J. Bergquist, and D. Wineland (2005), "*Spectroscopy using quantum logic*," Science **309**, 749.

Schrödinger, E. (1930), "*Zum Heisenbergschen Unschärfeprinzip*," Phys.-Math. Klasse **13**, 296.

Schrödinger, E. (1935), "*The present situation in quantum mechanics: A translation of Schrödinger's "cat paradox" paper*," Naturwissenschaften **23**, 807.

Schwabl, F. (2008), *Advanced Quantum Mechanics* (Springer).

Schwinger, J. (1965), "*On angular momentum*," in *Quantum theory of angular momentum*, edited by L. C. Biedenharn and H. van Dam (Academic Press, New York).

Scully, M. O., and S. Zubairy (1997), *Quantum Optics* (Cambridge University Press).

Sedlák, M., and M. Ziman (2009), "*Unambiguous comparison of unitary channels*," Phys. Rev. A **79**, 012303.

Seshadreesan, K. P., S. Kim, J. P. Dowling, and H. Lee (2013), "*Phase estimation at the quantum Cramér-Rao bound via parity detection*," Phys. Rev. A **87**, 043833.

Sewell, R., M. Koschorreck, M. Napolitano, B. Dubost, N. Behbood, and M. Mitchell (2012), "*Magnetic sensitivity beyond the projection noise limit by spin squeezing*," Phys. Rev. Lett. **109**, 253605.

Shaji, A., and C. M. Caves (2007), "*Qubit metrology and decoherence*," Phys. Rev. A **76**, 032111.

Shi, Y. (2003), "*Quantum entanglement of identical particles*," Phys. Rev. A **67**, 024301.

Simon, C., and J. Kempe (2002), "*Robustness of multiparty entanglement*," Phys. Rev. A **65**, 052327.

Sørensen, A. S., and K. Mølmer (2001), "*Entanglement and extreme spin squeezing*," Phys. Rev. Lett. **86**, 4431.

Spagnolo, N., C. Vitelli, V. G. Lucivero, V. Giovannetti, L. Maccone, and F. Sciarrino (2012), "*Phase estimation via quantum interferometry for noisy detectors*," Phys. Rev. Lett. **108**, 233602.

Steinberg, A. M. (2014), "*Quantum measurements: a modern view for quantum optics experimentalists*," in *101st Les Houches Summer School on "Quantum Optics and Nanophotonics"* (Oxford University Press) arXiv:1406.5535 [quant-ph] .

Stockton, J. K., J. Geremia, A. C. Doherty, and H. Mabuchi (2003), "*Characterizing the entanglement of symmetric many-particle spin-1/2 systems*," Phys. Rev. A **67**, 022112.

Streltsov, A. (2015), *Quantum Correlations Beyond Entanglement and Their Role in Quantum Information Theory*, SpringerBriefs in Physics (Springer) arXiv:1411.3208 [quant-ph] .

Summy, G. S., and D. T. Pegg (1990), "*Phase optimized quantum states of light*," Opt. Commun. **77**, 75.

Taddei, M. M., B. M. Escher, L. Davidovich, and R. L. de Matos Filho (2013), "*Quantum speed limit for physical processes,*" Phys. Rev. Lett. **110**, 050402.

Tóth, G. (2012), "*Multipartite entanglement and high-precision metrology,*" Phys. Rev. A **85**, 022322.

Tóth, G., and D. Petz (2013), "*Extremal properties of the variance and the quantum Fisher information,*" Phys. Rev. A **87**, 032324.

van Trees, H. L. (1968), *Detection, Estimation and Modulation Theory*, Vol. I (Wiley).

Tsang, M. (2012), "*Ziv-Zakai error bounds for quantum parameter estimation,*" Phys. Rev. Lett. **108**, 230401.

Ulam-Orgikh, D., and M. Kitagawa (2001), "*Spin squeezing and decoherence limit in Ramsey spectroscopy,*" Phys. Rev. A **64**, 052106.

van der Vaart, A. W. (1998), *Asymptotic Statistics* (Cambridge University Press).

Vaccaro, J. A., F. Anselmi, and H. M. Wiseman (2003), "*Entanglement of identical particles and reference phase uncertainty,*" Int. J. Quantum Inform. **1**, 427.

Vidrighin, M. D., G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. Kim, A. Datta, M. Barbieri, and I. A. Walmsley (2014), "*Joint estimation of phase and phase diffusion for quantum metrology,*" Nat. Commun. **5**, 3532.

Vitelli, C., N. Spagnolo, L. Toffoli, F. Sciarrino, and F. De Martini (2010), "*Enhanced resolution of lossy interferometry by coherent amplification of single photons,*" Phys. Rev. Lett. **105**, 113602.

Wasilewski, W., K. Jensen, H. Krauter, J. J. Renema, M. V. Balabas, and E. S. Polzik (2010), "*Quantum noise limited and entanglement-assisted magnetometry,*" Phys. Rev. Lett. **104**, 133601.

Wasserman, L. (2004), *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics (Springer).

Wineland, D. J. (2013), "*Nobel lecture: Superposition, entanglement, and raising Schrdingers cat,*" Rev. Mod. Phys. **85**, 1103.

Wineland, D. J., J. J. Bollinger, W. M. Itano, and D. J. Heinzen (1994), "*Squeezed atomic states and projection noise in spectroscopy,*" Phys. Rev. A **50**, 67.

Wineland, D. J., J. J. Bollinger, W. M. Itano, and F. L. Moore (1992), "*Spin squeezing and reduced quantum noise in spectroscopy,*" Phys. Rev. A **46**, R6797.

Wiseman, H. M., D. W. Berry, S. D. Bartlett, B. L. Higgins, and G. J. Pryde (2009), "*Adaptive measurements in the optical quantum information laboratory,*" IEEE J. Sel. Top. Quant. Electron. **15**, 1661.

Wiseman, H. M., and R. B. Killip (1997), "*Adaptive single-shot phase measurements: A semiclassical approach,*" Phys. Rev. A **56**, 944.

Wiseman, H. M., and R. B. Killip (1998), "*Adaptive single-shot phase measurements: The full quantum theory,*" Phys. Rev. A **57**, 2169.

Wiseman, H. M., and J. A. Vaccaro (2003), "*Entanglement of indistinguishable particles shared between two parties,*" Phys. Rev. Lett. **91**, 097902.

Wolfgramm, F., A. Cerè, F. A. Beduini, A. Predojević, M. Koschorreck, and M. W. Mitchell (2010), "*Squeezed-light optical magnetometry,*" Phys. Rev. Lett. **105**, 053601.

Wootters, W. K. (1981), "*Statistical distance and Hilbert space,*" Phys. Rev. D **23** (2), 357.

Xiang, G.-Y., B. L. Higgins, D. W. Berry, H. M. Wiseman, and G. J. Pryde (2010), "*Entanglement-enhanced measurement of a completely unknown optical phase*," Nature Photon. **5**, 43.

Yu, S. (2013), "*Quantum Fisher information as the convex roof of variance*," arXiv e-print, 1302.5311.

Yurke, B., S. L. McCall, and J. R. Klauder (1986), "*SU(2) and SU(1,1) interferometers*," Phys. Rev. A **33**, 4033.

Zander, C., A. R. Plastino, A. Plastino, and M. Casas (2007), "*Entanglement and the speed of evolution of multi-partite quantum systems*," J. Phys. A: Math. Theor. **40**, 2861.

Ziman, M. (2008), "*Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments*," Phys. Rev. A **77**, 062112.

Zurek, W. H. (2003), "*Decoherence, einselection, and the quantum origins of the classical*," Rev. Mod. Phys. **75**, 715.

Zwierz, M., C. A. Pérez-Delgado, and P. Kok (2010), "*General optimality of the Heisenberg limit for quantum metrology*," Phys. Rev. Lett. **105**, 180402.

Zwierz, M., C. A. Pérez-Delgado, and P. Kok (2012), "*Ultimate limits to quantum metrology and the meaning of the Heisenberg limit*," Phys. Rev. A **85**, 042112.