Spring 2021

# Error-corrected quantum metrology

Sisi Zhou
*Yale University Graduate School of Arts and Sciences*, sisi.zhou94@outlook.com

Abstract

Error-Corrected Quantum Metrology

Sisi Zhou

2021

Quantum metrology, which studies parameter estimation in quantum systems, has many applications in science and technology ranging from frequency spectroscopy to gravitational wave detection. Quantum mechanics imposes a fundamental limit on the estimation precision, called the Heisenberg limit (HL), which bears a quadratic enhancement over the standard quantum limit (SQL) determined by classical statistics. The HL is achievable in ideal quantum devices, but is not always achievable in presence of noise.

Quantum error correction (QEC), as a standard tool in quantum information science to combat the effect of noise, was considered as a candidate to enhance quantum metrology in noisy environment. This thesis studies metrological limits in noisy quantum systems and proposes QEC protocols to achieve these limits. First, we consider Hamiltonian estimation under Markovian noise and obtain a necessary and sufficient condition called the "Hamiltonian-not-in-Lindblad-span" condition to achieve the HL. When it holds, we provide ancilla-assisted QEC protocols achieving the HL; when it fails, the SQL is inevitable even using arbitrary quantum controls, but approximate QEC protocols can achieve the optimal SQL coefficient. We generalize the results to parameter estimation in quantum channels, where we obtain the "Hamiltonian-not-in-Kraus-span" condition and find explicit formulas for asymptotic estimation precision by showing attainability of previously established bounds using QEC protocols. All QEC protocols are optimized via semidefinite programming. Finally, we show that reversely, metrological bounds also restrict the performance of error-correcting codes by deriving a powerful bound in covariant QEC.

# Error-Corrected Quantum Metrology

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Sisi Zhou

Dissertation Director: Liang Jiang

June, 2021

# Contents

# Acknowledgements

First, I would like to express my deepest gratitude towards my advisor Liang Jiang for five years of exceptional mentorship. He introduced me into the amazing world of quantum information science and also taught me how to think like a scientist and pursue my own research interests. He has always been kind, supportive and full of insightful and concrete advice. Without his guidance, the thesis would have never been developed in the first place.

My appreciation also goes to my thesis committee, Steven Girvin, Michel Devoret and Meng Cheng for their time and helpful feedbacks on my research. I also want to acknowledge Haidong Yuan for being an external reader for this thesis. I would like to thank my co-authors Paola Cappellaro, Frederic T. Chong, Rafał Demkowicz-Dobrzański, Steven T. Flammia, Wojciech Górecki, Pranav Gokhale, Kade Head-Marsden, Liang Jiang, Stefan Krastanov, David Layden, Zi-Wen Liu, Ziqi Ma, Peter Maurer, Prineha Narang, Changhun Oh, John Preskill, Yat Wong, Xiaofei Yu, Mengzhen Zhang, Tian-Xing Zheng and Chang-Ling Zou for everything I learned from working with them. I was also fortunate to have fruitful discussions with my other colleagues Victor Albert, Senrui Chen, Arpit Dua, Connor Hann, Linshu Li, Wenlong Ma, Sreraman Muralidharan, Kyungjoo Noh, Alireza Seif, Chiao-Hsuan Wang, Jianming Wen, Qian Xu and Changchun Zhong in my advisor's group.

I am immensely grateful to my parents who always encourage me to pursue my dreams. I would also like to thank my wonderful friends for their company over the years which is an indispensable part of my PhD life.

Most of my research projects in this thesis are funded by ARL-CDQI, ARO, ARO MURI, AFOSR MURI, NSF, DOE, the Alfred P. Sloan Foundation and the Packard Foundation.

# List of Publications

## Publications discussed in this thesis

### Refereed journal articles

[1] Wojciech Górecki*, **Sisi Zhou***, Liang Jiang, and Rafał Demkowicz-Dobrzański. Optimal probes and error-correction schemes in multi-parameter quantum metrology. *Quantum*, 4:288, 2020.

[2] **Sisi Zhou** and Liang Jiang. Optimal approximate quantum error correction for quantum metrology. *Physical Review Research*, 2(1):013235, 2020.

[3] David Layden*, **Sisi Zhou***, Paola Cappellaro, and Liang Jiang. Ancilla-free quantum error correction codes for quantum metrology. *Physical Review Letters*, 122(4):040502, 2019.

[4] **Sisi Zhou** and Liang Jiang. Modern description of Rayleigh's criterion. *Physical Review A*, 99(1):013808, 2019.

[5] **Sisi Zhou**, Mengzhen Zhang, John Preskill, and Liang Jiang. Achieving the Heisenberg limit in quantum metrology using quantum error correction. *Nature Communications*, 9:78, 2018.

### Preprints

[6] **Sisi Zhou**, Zi-Wen Liu, and Liang Jiang. New perspectives on covariant quantum error correction. *arXiv:2005.11918*, 2020.

[7] **Sisi Zhou** and Liang Jiang. Asymptotic theory of quantum channel estimation. *arXiv:2003.10559*, 2020.

*First authorship is shared.

## Other publications during PhD

### Publications in refereed journals

[8] **Sisi Zhou**, Chang-Ling Zou, and Liang Jiang. Saturating the quantum Cramér–Rao bound using LOCC. *Quantum Science and Technology*, 5(2):025005, 2020.

[9] Stefan Krastanov, **Sisi Zhou**, Steven T Flammia, and Liang Jiang. Stochastic estimation of dynamical variables. *Quantum Science and Technology*, 4(3):035003, 2019.

[10] **Sisi Zhou** and Liang Jiang. Modern description of Rayleigh's criterion. *Physical Review A*, 99(1):013808, 2019.

### Preprints

[11] Ziqi Ma, Pranav Gokhale, Tian-Xing Zheng, **Sisi Zhou**, Xiaofei Yu, Liang Jiang, Peter Maurer, and Frederic T. Chong. Adaptive circuit learning for quantum metrology. *arXiv:2010.08702*, 2020.

[12] Stefan Krastanov, Kade Head-Marsden, **Sisi Zhou**, Steven T. Flammia, Liang Jiang, and Prineha Narang. Unboxing quantum black box models: learning non-markovian dynamics. *arXiv:2009.03902*, 2020.

[13] Changhun Oh, **Sisi Zhou**, Yat Wong, and Liang Jiang. Quantum limits of superresolution in noisy environment. *arXiv:2008.11339*, 2020.

[14] **Sisi Zhou** and Liang Jiang. An exact correspondence between the quantum Fisher information and the Bures metric. *arXiv:1910.08473*, 2019.

# List of Notations

| Symbol | Variants | Description |
|---|---|---|
| $\mathbb{R}, \mathbb{C}$ | | Real and complex fields |
| $\mathbb{H}$ | $\mathbb{H}_d$ | Hermitian matrix (in $\mathbb{C}^{d\times d}$) |
| $\mathbb{1}$ | $\mathbb{1}_d$ | Identity operator (in $\mathbb{C}^{d\times d}$) |
| $\otimes$ | $(\cdot)^{\otimes N}$ | Tensor product ($N$-fold tensor product) |
| $\oplus$ | | Direct sum of vector spaces |
| $A \geq B$ | | $A - B$ is positive semidefinite |
| $A \perp B$ | | $\mathrm{Tr}(A^\dagger B) = 0$ |
| $A^T$ $(A^*)$ | | Matrix transpose (Complex conjugate) of $A$ |
| $\mathrm{Re}(A)$ $(\mathrm{Im}(A))$ | | Real (Imaginary) part of $A$ |
| $\mathcal{H}$ | $\mathcal{H}_S$, $\mathcal{H}_A$ | Hilbert spaces |
| $\mathscr{S}(\cdot)$ | | Density operators acting on $(\cdot)$ |
| $\mathscr{L}(\cdot)$ | | Bounded linear operators acting on $(\cdot)$ |
| $\mathrm{Tr}(\cdot)$ | $\mathrm{Tr}_A(\cdot)$ | Trace (partical trace) |
| $\|\cdot\|$ | $\|\cdot\|_p$ | Operator norm (Schatten $p$-norm) for matrices |
| $\|\cdot\|_\infty$ | $\|\cdot\|_p$ | Maximum norm ($p$-norm) for vectors |
| $F(\cdot)$ | | (Quantum) Fisher information |
| $\mathfrak{F}(\cdot)$ | | Quantum Fisher information for channels |
| $\dot{\star}$ | | $\frac{\partial \star}{\partial \theta}$, derivative w.r.t. $\theta$ |
| $f(n) = O(g(n))$ | $f = o(g(n))$ | $\limsup_{n\to\infty} \lvert f(n)\rvert / g(n) < \infty$ $(= 0)$ |
| $f(n) = \Theta(g(n))$ | | $0 < \liminf_{n\to\infty} f(n)/g(n) \leq \limsup_{n\to\infty} \lvert f(n)\rvert / g(n) < \infty$ |

Table 1: Overview of notation conventions

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Quantum metrology

Quantum mechanics is a fundamental theory in physics describing the physical properties of microscopic systems at the atomic and subatomic scale. The Heisenberg's uncertainty principle [Heisenberg, 1949] is one of the best known theorems in quantum mechanics, which asserts a fundamental limit to the estimation accuracies of a pair of non-commuting physical observables, e.g. position and momentum operators. Namely, the more precisely the position of some particle is determined, the less precisely its momentum can be predicted, and vice versa. Such a feature of quantum physics is counterintuitive, which highlights a significant difference between quantum measurements and classical measurements, and is one of the motivations behind the study of quantum metrology [Giovannetti et al., 2004; Blatt and Wineland, 2008; Paris, 2009; Giovannetti et al., 2011], the science of measurements and estimation in quantum systems.

Quantum metrology has a broad range of applications in modern science and technology, including optical interferometry [Caves, 1981; Bondurant and Shapiro, 1984; Yurke et al., 1986; Holland and Burnett, 1993; Sanders and Milburn, 1995; Dowling, 1998; Mitchell et al., 2004; Walther et al., 2004; Nagata et al., 2007; Resch et al., 2007] (in particular, gravitational-wave detection [LIGO Collaboration, 2011, 2013]), magnetic and electric field spectroscopy [Wineland et al., 1992, 1994; Bollinger et al., 1996; Leibfried et al., 2004; Roos et al., 2006; Schmidt et al., 2005; Koschorreck et al., 2010; Sewell et al., 2012], atomic

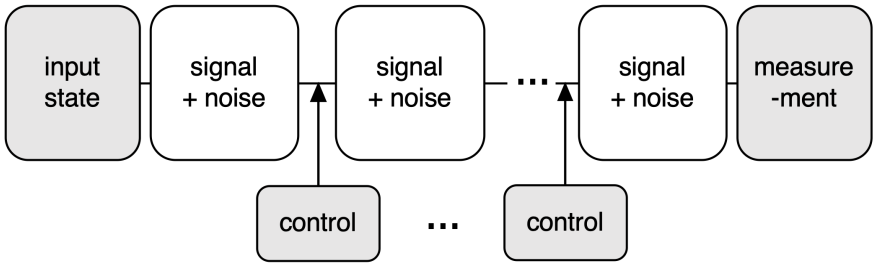


Figure 1.1: Schematic diagram of different components in quantum metrology. In this thesis, we focus on optimization of input states, quantum controls and final measurements, while the system dynamics (signal + noise) is fixed.

clocks [Rosenband et al., 2008; Appel et al., 2009; Leroux et al., 2010; Louchet-Chauvet et al., 2010], etc. In quantum metrology, we usually consider the following generic parameter estimation scenario (see Figure 1.1): The experimentalist first prepares an input quantum state of a physical system, let it evolve under certain quantum dynamics which encodes the relevant unknown parameter(s) and allows quantum controls in the middle, and then measure the output quantum state to extract the information of the parameter(s). The goal is to maximize the estimation precision of the parameter(s) while minimizing the resource, such as the probing time, the number of probing particles or the energy consumption. For example, in optical interferometry, an input laser beam passes through a beam splitter and travels along two different paths where the relevant parameter is encoded in the optical path difference. The final measurement step involves passing through the beam splitter and entering the photodetectors. By squeezing the input vacuum states [Caves, 1981], the shot-noise limited sensitivity in the GEO gravitational wave detector was enhanced by 3.5 dB [LIGO Collaboration, 2011]; in a proof-of-principle experiment in the LIGO gravitational wave detector, the injection of 10 dB of squeezing lowered the shot noise in the interferometer output by approximately 2.15 dB, equivalent to an increase by more than 60% in the power stored in the interferometer arm cavities [LIGO Collaboration, 2013]. This is an example in quantum metrology where the sensitivity of a quantum sensor is improved by choosing a non-classical quantum input state.

The experimental aspect of quantum metrology, usually called quantum sensing [Degen et al., 2017], studies the real-world implementations of quantum sensors on different physi-

cal platforms, such as cold atoms, trapped ions, solid-state spins, superconducting qubits, optical sensors and others. The type of physical platforms usually determines how the desired physical quantity interacts with the system and how initialization and measurement are performed. Building a highly sensitive quantum sensor then largely relies on balancing the trade-off between enhancing the sensitivity of the desired parameter and reducing the initialization and readout noise. These experimental restrictions usually determine the scope of applications of quantum sensors. For example, nitrogen-vacancy centers [Dutt et al., 2007; Neumann et al., 2008; Hanson et al., 2008; Taylor et al., 2008] are electronic spin defects in diamond that combine strong magnetic moment and efficient optical readout of atoms with the high spin densities in the solid state, which find applications microscopy of magnetic fields [Le Sage et al., 2013; Steinert et al., 2013; Fu et al., 2014], but still strong environmental noise in high-density ensembles [Acosta et al., 2009] limits its application in large-scale sensing of homogeneous fields.

Admittedly, for practical detection of physical quantities, it is important to find suitable quantum sensors with high sensitivity and low initialization and readout noise. On the other hand, given a fixed system dynamics with certain types of noise, it still remains a question how to choose the most sensitive initialization, control and readout schemes to maximize the capability of quantum sensors, as shown in Figure 1.1 where the gray blocks are fully controlled by the experimentalists and the white blocks are fixed. Similar to the idea of using squeezed states to enhance the sensitivity of optical interferometers, entanglement in a multi-probe system [Horodecki et al., 2009] was found to provide substantial enhancement in sensitivity [Wineland et al., 1992; Kitagawa and Ueda, 1993; Bollinger et al., 1996; Giovannetti et al., 2001, 2004, 2006, 2011; Pezzè et al., 2018]. According to the central limit theorem, assuming there is no more than classical correlations between each probes, an $N$-probe system reduces the estimation error by a factor of $\sqrt{N}$, which is called the standard quantum limit (SQL). The SQL can be overcome using entangled input states in which case the ultimate estimation limit allowed by quantum mechanics scales as $1/N$, known as the Heisenberg limit (HL). For example, when detecting a homogeneous magnetic field in a noiseless spin system, the HL is achievable using the Greenberger-Horne-Zeilinger (GHZ) state—an equal superposition of all spin up and all spin down states [Yurke et al.,

1986; Holland and Burnett, 1993; Giovannetti et al., 2006].

Instead of focusing on specific physical quantities to be estimated, and real-world quantum sensors on specific physical platforms, one could consider in general parameter estimation on arbitrary quantum states in a Hilbert space with arbitrary signal and noise dynamics. It brings us to the theoretical aspect of quantum metrology, usually called quantum estimation theory [Helstrom, 1976; Holevo, 1982; Hayashi, 2005, 2016], where the subject under study is a quantum state (density operator) $\rho_\theta$ or a quantum channel $\mathcal{E}_\theta$ as a function of unknown parameter(s) $\theta$. Quantum estimation theory is a generalization of classical estimation theory which is a branch of statistics that deals with parameter estimation based on measured empirical data that has random components.

In classical estimation theory [Kay, 1993; Van der Vaart, 2000; Lehmann and Casella, 2006], we estimate a family of probability distributions belonging to a certain parameterized subset $\{p_\theta(x)\}_{\theta\in\Theta}$ by obtaining a large set of independent and identically distributed (i.i.d.) samples $\{x_1, x_2, \ldots, x_N\}$ and infer the value of $\theta$ using an estimator $\hat{\theta}(x)$ as a function from the sample space to the estimate of $\theta$ (see Figure 1.2(a)). Usually, we evaluate the estimation error by the square root of the mean square error (MSE). When $N$ is large enough (or "asymptotically"), the MSE is proportional to $1/N$, matching the scaling $1/\sqrt{N}$ of the estimation error in the SQL, and the optimal coefficient is equal to the inverse of a statistical quantity called the Fisher information (FI). The inverse of the Fisher information naturally forms a lower bound on the MSE, called the Cramér–Rao (CR) bound. In quantum estimation theory, we examine parameter estimation based on i.i.d. samples obtained from quantum measurements on a family of quantum states $\{\rho_\theta\}_{\theta\in\Theta}$ (see Figure 1.2(b)). The estimation process is divided into two parts: the quantum measurement process where we perform a quantum measurement on the system and the data manipulation process where we estimate the true parameter from the observed data. The latter is a problem in classical estimation theory and the choice of quantum measurement is usually the main focus of quantum estimation theory.

Quantum estimation theory was first studied from the middle of 1960s to 1980 by many researchers [Helstrom, 1976; Holevo, 1982], where the lower bounds on the MSE for unbiased estimators were derived which combines the mathematical formulation of quantum

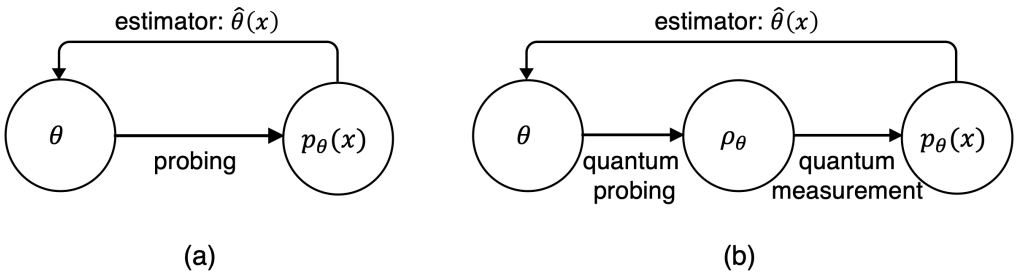


Figure 1.2: (a) Classical estimation theory. (b) Quantum estimation theory.

mechanics and mathematical statistics. This field was then developing rapidly, drawing a lot attention from information theorists, statisticians and physicists. One prominent result was that the quantum CR bound [Helstrom, 1976; Holevo, 1982], defined using the notion of quantum Fisher information (QFI), is attainable asymptotically in single-parameter estimation [Helstrom, 1967; Nagaoka, 1989a; Braunstein and Caves, 1994]. A problem is the optimal measurement attaining the quantum CR bound in general depends on the true value of the parameter. However, when the number of samples are large, it is possible attain the optimal estimation error asymptotically by adaptively choosing the measurement [Nagaoka, 1988, 1989b; Barndorff-Nielsen and Gill, 2000; Gill and Massar, 2000; Fujiwara, 2006]. For multi-parameter estimation, similar quantum lower bounds on the estimation error were derived [Helstrom, 1968; Yuen and Lax, 1973; Holevo, 1982; Nagaoka, 1989a; Matsumoto, 2002], though it remained open how to find the optimal measurement due to the non-commutativity nature of quantum mechanics [Ragy et al., 2016; Demkowicz-Dobrzański et al., 2020], except in a few special estimation models (mean-value estimation for Gaussian states [Holevo, 1982], quantum two-level systems [Nagaoka, 1989a; Hayashi, 2005; Gill and Massar, 2000], etc.). However, when collective measurements (measurements that might be correlated among the i.i.d. copies of parameterized quantum states) are allowed, it is proven that the Holevo bound [Holevo, 1982]—a more informative multi-parameter bound compared to the QFI-based bound—is attainable using quantum local asymptotic normality [Hayashi and Matsumoto, 2008; Kahn and Guţă, 2009; Yamagata et al., 2013; Yang et al., 2019], a useful asymptotic characterization of identically prepared states in terms of Gaussian states.

Notably, though the asymptotic quantum estimation theory is well-established, no uni-

fied treatment exists for the small sample case (the case where the number of samples is small). Possible approaches include the Bayesian approach where the unknown parameter is assumed to be random and the knowledge about the distribution is consecutively updated after each data-collection step, the minimax approach where the maximal error is minimized over all estimators, etc., all suffering from limited scopes of applications. For parameters with group symmetries, however, covariant measurements are always optimal if we adopt use Bayesian approach with an invariant prior and an invariant loss function [Helstrom, 1974; Holevo, 1979; Vidal et al., 1999; Buzek et al., 1999; Chiribella et al., 2005].

It was not until the late 1980s when people started to shift their attention from quantum state estimation to quantum process estimation, and when the intersection between practical quantum sensing and quantum estimation theory becomes clear. Instead of focusing on optimizing the quantum measurement on fixed parametrized quantum states, the importance of choosing optimal input states for parametrized quantum processes was also appreciated. Unitary quantum processes are the most fundamental dynamics in quantum mechanics and it was proven, using quantum estimation theory, that GHZ-type states are optimal, achieving a quadratic enhancement over classical states [Bollinger et al., 1996; Giovannetti et al., 2004]. Similar to GHZ states in many body systems, NOON states in optical interferometry (an equal superposition of $|0, N\rangle$ and $|N, 0\rangle$ Fock states) were also known to achieve the HL [Yurke, 1986; Lee et al., 2002].

It was a remarkable discovery that by exploring entanglement in multi-probe systems, it is possible to achieve a quadratic enhancement in sensing. However, the fact that practical systems are noisy forbids direct applications of the result [Banaszek et al., 2009; Giovannetti et al., 2011; Maccone and Giovannetti, 2011]. Moreover, entangled states such as GHZ states are more fragile compared to product states because both quantum signal and quantum noise accumulate coherently for highly entangled states. It is therefore imperative for us to understand how quantum noise affects the quantum enhancement provided by entangled states and whether quantum controls can overcome it. In the next section, I am going to briefly review the history of studying quantum process estimation under the influence of quantum noise.

## 1.2 Noisy quantum metrology

The study of noise in quantum metrology was initiated 20 years ago due to experimental motivations. For example, to estimate the $Z$-axis magnetic field under dephasing noise, it was shown that for a long enough probing time, the SQL is inevitable and the spin-squeezed states are optimal, providing a constant factor improvement over product states and GHZ states [Huelga et al., 1997; Ulam-Orgikh and Kitagawa, 2001]. Though GHZ states still have the advantage when the probing time is constrained for practical reasons [Wineland et al., 1998; André et al., 2004; Borregaard and Sørensen, 2013; Chaves et al., 2013], it is one of the first lines of evidence that the quantum advantages are conditioned in the presence of noise. Similarly in optical interferometry, the quadratic advantage provided by NOON states is not robust under photon loss [Demkowicz-Dobrzański et al., 2009; Dorner et al., 2009] or phase diffusion [Genoni et al., 2011; Escher et al., 2012] and other types of states such as squeezed states and two-mode states with definite photon numbers were explored as candidates to achieve the optimal estimation precision.

It is of theoretical interest to explore the limit of quantum process estimation as well. Using the language of quantum channels, definitions of the QFI for quantum channels were first studied by [Fujiwara and Imai, 2008] and [Matsumoto, 2010]. The channel QFI is usually defined to be the maximum QFI of the output state over all possible input states and calculating the channel QFI is equivalent to optimizing the parameter estimation over all possible inputs. The method in [Fujiwara and Imai, 2008] is usually called the channel-extension method, where the QFI is computed by first purifying the probe states to a larger Hilbert space and then minimizing the QFI over the freedom in purifications. Given many copies of channels, the upper bounds of the QFI were derived using the channel-extension method for typical noise channels such as depolarization, dephasing, spontaneous emission and photon loss [Escher et al., 2011; Demkowicz-Dobrzański et al., 2012; Kołodyński and Demkowicz-Dobrzański, 2013; Demkowicz-Dobrzański and Maccone, 2014].

Another method to derive upper bounds on the QFI is the channel-simulation method which works for programmable quantum channels [Ji et al., 2008; Demkowicz-Dobrzański et al., 2012], i.e. channels that can be simulated by parameter-independent quantum chan-

nels and parameterized quantum states. The upper bounds are directly related to the QFI of the parameterized quantum states. The channel-simulation method is as general as the channel-extension method, but is in particular useful for characterizing the QFI of teleportation-simulable channels [Pirandola et al., 2017; Pirandola and Lupo, 2017; Pirandola et al., 2018; Takeoka and Wilde, 2016].

Importantly, the above methods demonstrated that an infinitesimal amount of generic noise is enough to destroy the HL with respect to the number of probes and force the scaling to be SQL-like, even with the assistance of arbitrary quantum controls [Escher et al., 2011; Demkowicz-Dobrzański et al., 2012; Kołodyński and Demkowicz-Dobrzański, 2013; Demkowicz-Dobrzański and Maccone, 2014]. Thus, the optimal quantum enhancement is then limited to at most a constant factor improvement over classical strategies for most types of noise. Apart from quantum channels, a similar phenomenon occurs in open quantum systems [Sekatski et al., 2017] where the HL with respect to the probing time is destroyed under generic noise and the scaling of the QFI is at most proportional to the probing time.

Two important open questions remain in the study of noisy quantum metrology. First, although it was established that for generic noise the HL is not achievable, there are still possibilities for quantum systems to achieve the HL under some special types of noise and we don't know if there is such a protocol, except for unitary dynamics. Second, although achieving the HL is the ultimate dream in quantum metrology, the constant-factor improvement in the SQL is still practically relevant. Notably, when the noise rate in the system is low, the constant-factor improvement could be significant. Therefore, it is intriguing to understand how to achieve this significant enhancement in the SQL case. I will show in this thesis that quantum error correction is an effective tool to solve these two open questions, namely, achieving the HL for special types of noise and achieving the optimal SQL coefficient for generic noise. In the next section, I am going briefly review on the concept of quantum error correction and then discuss its application in quantum metrology in particular.

## 1.3 Quantum error correction

Quantum information science is an extensive research field that combining information science and quantum physics which includes both theoretical issues in information-theoretic and computational models as well as experimental topics in quantum physics. The revolution in quantum information science began in the 1990s following the stunning discovery of Peter Shor's quantum factoring algorithm [Shor, 1999]. Since then, there have been rapid developments in many topics, including quantum computation and quantum communication [Nielsen and Chuang, 2010; Wilde, 2013]. Immediate skepticism was also raised along with the theoretical development because quantum systems were known to be fragile to environmental noise (one famous example being Schrödinger's cat) and it is a critical issue how to overcome the unavoidable interactions with environment in the manufacturing of quantum computers. Classically, error correction is a successful technique to combat noise where a message is encoded with redundant information so that a limited number of errors can be corrected to recover the original message [MacWilliams and Sloane, 1977; Huffman and Pless, 2010]. Compared to classical systems, there are two apparent difficulties for implementing quantum error correction (QEC): quantum information cannot be reliably copied (the no-cloning theorem) and quantum measurement usually destroys the information stored in quantum states.

Therefore, it came as a surprise and caused much excitement when powerful QEC methods were discovered to overcome these difficulties. The first QEC codes [Shor, 1995; Steane, 1996b] were discovered around 1995 which can correct arbitrary single-qubit error, followed within a few years by general code constructions and frameworks of QEC [Calderbank and Shor, 1996; Steane, 1996a; Bennett et al., 1996; Knill and Laflamme, 1997] and an important concept called stabilizer codes [Gottesman, 1996; Calderbank et al., 1997] which permits many new codes to be discovered. Quantum fault-tolerance theorems [Knill and Laflamme, 1996; Aharonov and Ben-Or, 1999] were also proven, showing that arbitrarily good quantum computation can be achieved even with faulty quantum gates, provided only that the error rate per gate is below a certain threshold. Besides its broad applications in quantum computing [Gottesman, 2009; Lidar and Brun, 2013], QEC is also an important concept in

quantum communication where the classical or quantum capacity of noisy quantum channels are defined to be the maximum rate of reliable transmission of classical or quantum bits of information through the channel using the optimal encoder and decoder [Bennett and Shor, 2004; Nielsen and Chuang, 2010; Wilde, 2013]. The code optimization is extremely difficult though—for example, the famous Holevo-Schumacher-Westmoreland theorem [Holevo, 1998; Schumacher and Westmoreland, 1997] which links the classical capacity to entropic quantities was proven using the random coding method without explicit code construction.

It is natural to ask whether QEC is useful in quantum metrology. An apparent answer would be "yes": If QEC can combat the effect of noise in quantum computation and communication, it must work in quantum metrology as well. However, in quantum metrology, instead of studying fixed quantum states or channels, we study a family of parametrized states or channels. Therefore, a useful QEC code in metrology must satisfy the following two conditions simultaneously: correct the noise and protect the signal. The problem was first raised in [Preskill, 2000], which considers the possibility of using QEC to improve the robustness of clock synchronization. The answer was negative as the signal would inevitably be corrected along with the noise. Motivated by rapid development of quantum sensors [Waldherr et al., 2014; Hirose and Cappellaro, 2016], the problem was revisited a decade later by several groups independently [Ozeri, 2013; Dür et al., 2014; Arrad et al., 2014; Kessler et al., 2014] where it was shown that QEC can improve the sensitivity for probes sensing a signal in $x$-direction under dephasing noise. The protocol was later experimentally demonstrated in nitrogen-vacancy centers which have highly suppressed bit-flip noise and dominant dephasing noise [Unden et al., 2016].

The result was the beginning of the trip towards disclosing the full potential of QEC in quantum metrology [Herrera-Martí et al., 2015; Lu et al., 2015; Reiter et al., 2017; Matsuzaki and Benjamin, 2017; Sekatski et al., 2017; Demkowicz-Dobrzański et al., 2017; Zhou et al., 2018; Layden and Cappellaro, 2018; Layden et al., 2019; Górecki et al., 2020; Tan et al., 2019; Kapourniotis and Datta, 2019; Zhuang et al., 2020; Layden et al., 2020; Zhou and Jiang, 2020b,a; Chen et al., 2020; Rojkov et al., 2021]. Recall from Section 1.2, powerful bounds that characterize the effect of noise in quantum metrology were discovered, though the tightness of these bounds were never proven. Encouraged by the example of sensing a

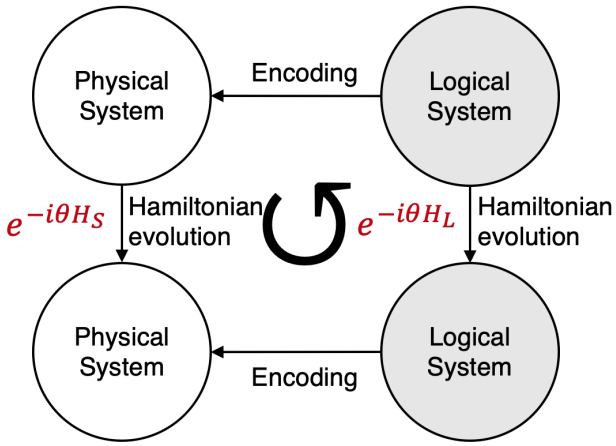


Figure 1.3: Covariant QEC. A covariant code is defined using an encoding channel such that the diagram commutes for arbitrary $\theta$, i.e. $e^{-iH_S\theta}$ acting after an encoding channel is equivalent to $e^{-iH_L\theta}$ acting before it. $H_S$ is the transversal system Hamiltonian and $H_L$ is the logical Hamiltonian.

signal in $x$-direction under dephasing noise, we are now ready to explore the role of QEC in general quantum process estimation. In this thesis, I will present our results on error-corrected quantum metrology. In particular, I will show that using QEC, it is possible to achieve the HL in noisy systems whenever allowed in principle, and to achieve the optimal SQL coefficient in generic noisy systems, solving the two open questions raised at the end of Section 1.2.

QEC is useful in quantum metrology; reversely, quantum metrology also provides powerful bounds in QEC. The key idea of QEC is to encode the logical state into a large system and correct certain types of noise using redundancy, so the structure of the noise must also place restrictions on the QEC codes. This feature was beautifully captured by the Eastin–Knill theorem [Eastin and Knill, 2009] (see also [Bravyi and König, 2013; Pastawski and Yoshida, 2015; Jochym-O'Connor et al., 2018; Wang et al., 2020]), which states that if we divide a system into several subsystems, for any QEC code that corrects local errors in single subsystems, transversal gates (gates that do not couple different blocks) are not universal. Similarly, if we consider covariant codes (see Figure 1.3) whose logical Hamiltonian is simulated by a transversal physical Hamiltonian [Hayden et al., 2017; Faist et al., 2020], it cannot perfectly correct local errors. Covariant QEC is closely connected to many topics in quantum physics including fault-tolerant computing [Eastin and Knill, 2009], quantum clocks [Woods and Alhambra, 2020], quantum gravity [Harlow and Ooguri, 2019, 2018] and

condensed matter physics [Brandão et al., 2019]. The QEC accuracy for covariant codes as a function of noise and number of subsystems was recently lower bounded using complementary channel techniques in terms of erasure errors [Faist et al., 2020; Bény and Oreshkov, 2010; Hayden et al., 2008; Bény et al., 2018], producing a quantification of the Eastin–Knill theorem. Surprisingly, it can also be bounded using metrology for more general types of noise [Kubica and Demkowicz-Dobrzański, 2020; Zhou et al., 2020], based on the observation that a covariant code naturally induces a QEC protocol that estimates the Hamiltonian parameter and is therefore subject to metrological bounds. I will explain our work on this topic in this thesis as well.

## 1.4 Outline of the thesis

The main purpose of this thesis is to introduce novel QEC protocols that enhance quantum metrology and a general criterion that determines whether a quantum process follows the HL or the SQL. Moreover, by optimizing these QEC protocols using semidefinite programs, the ultimate estimation precision is attainable asymptotically. Reversely, the metrological bounds also restrict the QEC accuracy of covariant codes.

To proceed, I will present the preliminaries in Chapter 2 including the classical and quatum CR bound, QFI, other multi-parameter bounds and quantum channel estimation for quantum metrology and the QEC conditions and approximate QEC for quantum error correction.

In Chapter 3, I will discuss our works on error-corrected metrology for Hamiltonian estimation under Markovian noise, one of the most important sensing scenarios, where we focus on estimating the Hamiltonian parameter(s) in open quantum systems using fast and frequent ancilla-assisting quantum controls. We obtain a necessary and sufficient condition determining the attainability of the HL with respect to the probing time in both the single-parameter [Zhou et al., 2018] and the multi-parameter scenario [Górecki et al., 2020], where we use QEC protocols to prove the sufficiency part, and present numerical tools for the QEC optimization in both the HL [Zhou et al., 2018; Górecki et al., 2020] and SQL [Zhou and Jiang, 2020b] cases. We also seek the possibility of alleviating the ancilla-assisting

assumption in a special case where the Hamiltonian and noise commute [Layden et al., 2019].

In Chapter 4, the results in Chapter 3 are generalized to quantum channel estimation, where the channel is an arbitrary function of an unknown parameter [Zhou and Jiang, 2020a]. We consider the situation where we have a large number of channels and show that the asymptotic channel QFI is attainable using ancilla-assisting QEC protocols. Both the asymptotic QFI and the optimal QEC protocols are computable via semidefinite programs.

I will present our work on covariant QEC in Chapter 5, where we use quantum channel estimation theory to derive a powerful lower bound on the QEC accuracy of covariant error-correcting codes that is efficiently computable for general noise channels [Zhou and Jiang, 2020a].

Finally, I summarize and conclude the thesis in Chapter 6. I will discuss the open problems and future directions in the field of error-corrected quantum metrology.

# Chapter 2

# Preliminaries

## 2.1 Classical estimation theory

### 2.1.1 One-parameter estimation

In this thesis, we focus on local estimation of unknown parameters. In one-parameter estimation of an unknown parameter $\theta$, we consider the situation where $\theta$ is contained in a sufficiently small neighborhood $\Theta$ of its true value and the estimators of $\theta$ will only be optimized in the neighborhood of $\theta$ which may not be optimal globally. It is a reasonable assumption when the number of samples are sufficiently large, and we can use a negligible fraction of the samples to determine the small neighborhood in a pre-estimation step.

Given a family of probability distributions $p_\theta(x)$ of a random variable $X$, we consider estimation of $\theta$ by sampling from $X$. We will assume $X \in \Omega$ is discrete but the derivations below work for continuous $X$ as well. All theorems and proofs in Section 2.1 can be found in standard textbooks, e.g. [Kay, 1993; Lehmann and Casella, 2006; Kobayashi et al., 2011]. In order to infer the value of $\theta$ from a set of data $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\} \in \Omega^N$, we need an estimator $\hat{\theta}_N : \Omega^N \to \Theta$ for each $N$. The mean square error (MSE) is defined by

$$\Delta^2\hat{\theta}_N = \mathbb{E}[(\hat{\theta}_N - \theta)^2] = \sum_{\boldsymbol{x}\in\Omega} p_\theta(\boldsymbol{x})(\hat{\theta}_N(\boldsymbol{x}) - \theta)^2, \tag{2.1}$$

where $p_\theta(\boldsymbol{x}) = \prod_{i=1}^N p_\theta(x_i)$. Usually, the estimator is assumed to be unbiased, i.e.

$$\mathbb{E}[\hat{\theta}_N] = \sum_{\boldsymbol{x}\in\Omega^N} p_\theta(\boldsymbol{x})\hat{\theta}_N(\boldsymbol{x}) = \theta, \quad \forall\theta \in \Theta. \tag{2.2}$$

In local estimation, we could relax it to the local unbiasedness condition (at some $\theta \in \Theta$), i.e.

$$\mathbb{E}[\hat{\theta}_N] = \sum_{\boldsymbol{x}\in\Omega^N} p_\theta(\boldsymbol{x})\hat{\theta}_N(\boldsymbol{x}) = \theta, \tag{2.3}$$

$$\frac{\partial}{\partial\theta}\mathbb{E}[\hat{\theta}_N] = \sum_{\boldsymbol{x}\in\Omega^N} \frac{\partial p_\theta(\boldsymbol{x})}{\partial\theta}\hat{\theta}_N(\boldsymbol{x}) = 1, \tag{2.4}$$

which means that the estimation of $\theta$ is precise at its true value and also precise up to the first order approximation in the vicinity of it. Although in practice, biased estimators are sometimes more accurate than unbiased estimators, we ignore these scenarios in the setting of local estimation and $N \to \infty$ limit.

The Cramér–Rao (CR) bound is a key concept in classical estimation theory which provide a lower bound on the MSE for all unbiased estimators. It states that

**Theorem 2.1** (CR bound)**.** *For any locally unbiased estimator $\hat{\theta}_N$, the MSE is lower bounded by*

$$\Delta^2\hat{\theta}_N \geq \frac{1}{NF(p_\theta)}, \tag{2.5}$$

*where $F(p_\theta)$ is the (classical) Fisher information defined by*

$$F(p_\theta) = \sum_{x\in\Omega} \frac{1}{p_\theta(x)}\left(\frac{\partial p_\theta(x)}{\partial\theta}\right)^2 = \sum_{x\in\Omega} p_\theta(x)\left(\frac{\partial \log p_\theta(x)}{\partial\theta}\right)^2, \tag{2.6}$$

*where the summation over all terms such that $p_\theta(x) \neq 0$.*

*Proof.* Since $\sum_{\boldsymbol{x}\in\Omega^N} p_\theta(\boldsymbol{x}) = 1$ and its derivative is zero, according to the local unbiasedness condition,

$$\sum_{\boldsymbol{x}\in\Omega^N} \frac{\partial p_\theta(\boldsymbol{x})}{\partial\theta}(\hat{\theta}_N(\boldsymbol{x}) - \theta) = 1 \tag{2.7}$$

Using the Cauchy–Schwarz inequality, we have

$$\left(\sum_{\boldsymbol{x}\in\Omega^N}\frac{1}{p_\theta(\boldsymbol{x})}\left(\frac{\partial p_\theta(\boldsymbol{x})}{\partial\theta}\right)^2\right)\left(\sum_{\boldsymbol{x}\in\Omega^N}p_\theta(\boldsymbol{x})(\hat{\theta}_N(\boldsymbol{x})-\theta)^2\right)\geq 1. \tag{2.8}$$

Moreover,

$$\left(\sum_{\boldsymbol{x}\in\Omega^N}\frac{1}{p_\theta(\boldsymbol{x})}\left(\frac{\partial p_\theta(\boldsymbol{x})}{\partial\theta}\right)^2\right)=NF(p_\theta), \tag{2.9}$$

proving Eq. (2.5). □

Certainly, the CR-bound only holds under certain regularity conditions, e.g. $\partial\log p_\theta(x)/\partial\theta$ exists and the summation can be interchanged with the derivative. From the proof above, we see that the CR bound at $\theta_0$ is saturated if and only if

$$\frac{1}{p_\theta(\boldsymbol{x})}\frac{\partial p_\theta(\boldsymbol{x})}{\partial\theta}\bigg|_{\theta=\theta_0}=NF(p_{\theta_0})(\hat{\theta}_N(\boldsymbol{x})-\theta_0),\quad\forall\boldsymbol{x}\in\Omega^N, \tag{2.10}$$

and any estimator satisfying this condition, i.e. saturating the CR bound, is called efficient.

Since we consider the case where the probability distribution $p_\theta(x)$ is known, a reasonable procedure to find the value of $\theta$ is to use the maximum-likelihood estimator (MLE):

$$\hat{\theta}_N^{\mathrm{ML}}(\boldsymbol{x})=\arg\max_\theta p_\theta(\boldsymbol{x}). \tag{2.11}$$

Although this value is not always unique, when we know it exists and is unique, we call it the MLE. Now we show that the MLE is asymptotically ($N\to\infty$) unbiased and efficient. First, note that if the likelihood function $p_\theta(\boldsymbol{x})$ is differentiable, the MLE must satisfy

$$\ell(\boldsymbol{x};\hat{\theta}_N^{\mathrm{ML}}):=\frac{\partial\log p_\theta(\boldsymbol{x})}{\partial\theta}\bigg|_{\theta=\hat{\theta}_N^{\mathrm{ML}}}=0, \tag{2.12}$$

because the logarithmic function is monotonic. It in general has multiple solutions and we must select the one which yields the largest value of $p_\theta(\boldsymbol{x})$. Suppose the true value of $\theta$ is $\theta_0$,

$$\ell(\boldsymbol{x};\theta)=\ell(\boldsymbol{x};\theta_0)+\partial_\theta\ell(\boldsymbol{x};\theta_0)(\theta-\theta_0)+O((\theta-\theta_0)^2). \tag{2.13}$$

By setting the LHS to be zero, we find that

$$\hat{\theta}_N^{\mathrm{ML}} \approx \theta_0 + \frac{1}{\partial_\theta \ell(\boldsymbol{x};\theta_0)} \ell(\boldsymbol{x};\theta_0). \tag{2.14}$$

First, note that $\mathbb{E}[\ell(\boldsymbol{x};\theta)] = 0$, by the law of large numbers, the MLE is asymptotically unbiased (therefore also locally unbiased) because

$$\lim_{N\to\infty} \mathbb{E}[\hat{\theta}_N^{\mathrm{ML}}(\boldsymbol{x})] = \theta_0. \tag{2.15}$$

Second, note that $\mathbb{E}[-\partial_\theta \ell(\boldsymbol{x};\theta_0)] = NF(p_{\theta_0})$,

$$\mathbb{E}[(\hat{\theta}_N^{\mathrm{ML}} - \theta_0)^2] \approx \mathbb{E}\Big[\Big(\frac{1}{\partial_\theta \ell(\boldsymbol{x};\theta_0)} \ell(\boldsymbol{x};\theta_0)\Big)^2\Big] \approx \frac{1}{NF(p_{\theta_0})}. \tag{2.16}$$

Therefore, the CR bound is asymptotically saturated using the MLE estimator and represents the fundamental estimation limit in classical local estimation theory.

### 2.1.2 Multi-parameter estimation

The discussion above can be generalized to multi-parameter estimation. We are not going to provide detailed proofs, but only state the results. In multi-parameter estimation, we consider a family of probability distributions $p_{\boldsymbol{\theta}}(x)$ which is a function of $P$ unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)$. The MSE is not longer a real number. Instead, for locally unbiased estimators $\hat{\boldsymbol{\theta}}_N(\boldsymbol{x})$, we use the mean square matrix (or covariance matrix):

$$\mathbb{V}[\hat{\boldsymbol{\theta}}_N] = \mathbb{E}\left[(\hat{\boldsymbol{\theta}}_N(\boldsymbol{x}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_N(\boldsymbol{x}) - \boldsymbol{\theta})^T\right], \tag{2.17}$$

i.e. $\mathbb{V}[\hat{\boldsymbol{\theta}}_N(\boldsymbol{x})]_{ij} = \mathbb{E}\left[(\hat{\theta}_{N,i}(\boldsymbol{x}) - \theta_i)(\hat{\theta}_{N,j}(\boldsymbol{x}) - \theta_j)\right]$, to quantify the estimation error.

**Theorem 2.2** (Multi-parameter CR bound)**.** *For any locally unbiased estimator $\hat{\boldsymbol{\theta}}_N$, the MSE is lower bounded by*

$$\mathbb{V}[\hat{\boldsymbol{\theta}}_N] \geq \frac{F(p_{\boldsymbol{\theta}})^{-1}}{N}, \tag{2.18}$$

*where $F(p_{\boldsymbol{\theta}})$ is the (classical) Fisher information matrix defined by*

$$F(p_{\boldsymbol{\theta}})_{ij} = \sum_{x\in\Omega} \frac{1}{p_{\boldsymbol{\theta}}(x)} \frac{\partial p_{\boldsymbol{\theta}}(x)}{\partial \theta_i} \frac{\partial p_{\boldsymbol{\theta}}(x)}{\partial \theta_j} = \sum_{x\in\Omega} p_{\boldsymbol{\theta}}(x) \left(\frac{\partial \log p_{\boldsymbol{\theta}}(x)}{\partial \theta_i}\right) \left(\frac{\partial \log p_{\boldsymbol{\theta}}(x)}{\partial \theta_j}\right), \tag{2.19}$$

*where the summation over all terms such that $p_{\boldsymbol{\theta}}(x) \neq 0$. Here $A \geq B$ means $A - B$ is positive semidefinite and $^{-1}$ is matrix inverse.*

Moreover, the MLE for multiple parameters $\hat{\boldsymbol{\theta}}_N^{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}}(\boldsymbol{x})$ is also asymptotically unbiased and efficient in the $N \to \infty$ limit.

## 2.2 Quantum estimation theory

In quantum metrology, instead of considering a family of parametrized probability distributions, we consider a family of parametrized quantum state $\{\rho_\theta\}_{\theta\in\Theta} \subseteq \mathscr{S}(\mathcal{H})$ where $\rho_\theta$ are density operators satisfying $\rho_\theta \geq 0$, $\rho_\theta^\dagger = \rho_\theta$ and $\mathrm{Tr}(\rho_\theta) = 1$. We will use $\mathscr{S}(\mathcal{H})$ to denote the set of density operators on the Hilbert space $\mathcal{H}$. As before, we will first focus on one-parameter estimation and then discuss multi-parameter estimation.

### 2.2.1 Quantum Fisher information

Quantum measurements (or positive operator-valued measures) in a Hilbert space $\mathcal{H}$ are described a set of operators $\{M_x\}_{x\in\Omega} \subseteq \mathscr{L}(\mathcal{H})$ (where $\mathscr{L}(\mathcal{H})$ denotes bounded linear operators on $\mathcal{H}$), satisfying

$$\sum_{x\in\Omega} M_x = I, \quad M_x = M_x^\dagger \geq 0. \tag{2.20}$$

Note that we will assume $\Omega$ is discrete, but for a general $\Omega$ equipped with a $\sigma$-algebra, we could replace all summation with integration and consider measurements as a function from the $\sigma$-algebra to $\mathscr{L}(\mathcal{H})$ which satisfies the countable additivity condition, and the discussion still carries on.

Given a fixed quantum measurement, the probability of getting a measurement outcome $x$ is $p_\theta(x) = \mathrm{Tr}(\rho_\theta M_x)$. In order to find the optimal measurement which minimizes the MSE, we would like to maximize the classical Fisher information (FI) $F(p_\theta)$ over all measurements. Before we do so, we first introduce a useful Hermitian operator called the

symmetric logarithmic derivative (SLD) defined by

$$\frac{1}{2}\left(L_\theta \rho_\theta + \rho L_\theta\right) = \partial_\theta \rho_\theta \tag{2.21}$$

Though the solution may not be unique when $\rho_\theta$ is not full-rank, all solutions will lead to the same results and we do not distinguish among them. Note that $L_\theta$ classically reduces to $\ell(x;\theta)$ (Eq. (2.12)).

We now calculate the maximum classical Fisher information [Braunstein and Caves, 1994]. We note that

$$\begin{aligned}
&F(p_\theta(x)) \\
&= \sum_{x:\mathrm{Tr}(M_x\rho_\theta)\neq 0} \frac{\left(\mathrm{Tr}(M_x \partial_\theta \rho_\theta)\right)^2}{\mathrm{Tr}(M_x\rho_\theta)} = \sum_{x:\mathrm{Tr}(M_x\rho_\theta)\neq 0} \frac{\left(\mathrm{Re}[\mathrm{Tr}(M_x L_\theta \rho_\theta)]\right)^2}{\mathrm{Tr}(M_x\rho_\theta)} \\
&\leq \sum_{x:\mathrm{Tr}(M_x\rho_\theta)\neq 0} \frac{\left(|\mathrm{Tr}(M_x L_\theta \rho_\theta)|\right)^2}{\mathrm{Tr}(M_x\rho_\theta)} \leq \sum_{x:\mathrm{Tr}(M_x\rho_\theta)\neq 0} \mathrm{Tr}(M_x L_\theta \rho_\theta L_\theta) \leq \mathrm{Tr}(L_\theta^2 \rho_\theta),
\end{aligned} \tag{2.22}$$

where the first equality holds true when

$$\mathrm{Im}[\mathrm{Tr}(M_x L_\theta \rho_\theta)] = 0, \text{ for all } x, \tag{2.23}$$

the second equality holds true when

$$M_x^{1/2}\rho_\theta^{1/2} = \lambda_x M_x^{1/2} L_\theta \rho_\theta^{1/2}, \ \lambda_x \in \mathbb{C}, \text{ for all } x, \tag{2.24}$$

based on the use of the Cauchy–Schwarz inequality, and the third equality holds true when

$$\forall x \text{ s.t. } \mathrm{Tr}(M_x\rho_\theta) = 0, \quad \mathrm{Tr}(M_x L_\theta \rho_\theta L_\theta) = 0. \tag{2.25}$$

The conditions Eq. (2.23) and Eq. (2.24) are simplified to

$$M_x^{1/2}\rho_\theta^{1/2} = \lambda_x M_x^{1/2} L_\theta \rho_\theta^{1/2}, \ \lambda_x \in \mathbb{R}, \text{ for all } x. \tag{2.26}$$

Clearly, both conditions Eq. (2.26) and Eq. (2.25) are satisfied by choosing $M_x$ to be rank-

one projectors onto the eigenstates of $L_\theta$. Therefore, the RHS of Eq. (2.22) is attainable and we define it to be the quantum Fisher information (QFI) [Helstrom, 1976; Holevo, 1982],

$$F(\rho_\theta) = \max_{\{M_x\}} F(p_\theta(x)) = \mathrm{Tr}(\rho_\theta L_\theta^2). \tag{2.27}$$

$F(\rho_\theta)$ can be directly computed when $\rho_\theta = \sum_k \lambda_{\theta,k} |\psi_{\theta,k}\rangle \langle\psi_{\theta,k}|$ is diagonalized,

$$L_\theta = \sum_{\substack{j,k \\ \lambda_{\theta,j}+\lambda_{\theta,k}\neq 0}} \frac{2}{\lambda_{\theta,j}+\lambda_{\theta,k}} \langle\psi_{\theta,j}| \partial_\theta\rho_\theta |\psi_{\theta,k}\rangle |\psi_{\theta,j}\rangle \langle\psi_{\theta,k}|, \tag{2.28}$$

$$F(\rho_\theta) = \sum_{\substack{j,k \\ p_{\theta,j}+p_{\theta,k}\neq 0}} \frac{2}{\lambda_{\theta,j}+\lambda_{\theta,k}} |\langle\psi_{\theta,j}| \partial_\theta\rho_\theta |\psi_{\theta,k}\rangle|^2. \tag{2.29}$$

There are other approaches to define the QFI, e.g. using quantum fidelity [Hübner, 1992]

$$\frac{1}{4}F(\rho_\theta)d\omega^2 = d_B^2(\rho_\omega, \rho_{\omega+d\omega}) = 2 - 2f_B(\rho_\omega, \rho_{\omega+d\omega}), \tag{2.30}$$

where $d_B(\rho,\sigma) = \sqrt{2-2f_B(\rho,\sigma)}$ is the Bures distance and the fidelity $f_B(\rho,\sigma) = \mathrm{Tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})$. Though these two definitions will be different at singularity points if there exists $k$ such that $\lambda_{\theta,k} = 0$ and $\partial_\theta^2\lambda_{\theta,k} \neq 0$ [Safranek, 2017; Zhou and Jiang, 2019]. To avoid singularities, we assume in this thesis that for all $\lambda_{\theta,k} = 0$, $\partial_\theta^2\lambda_{\theta,k} = 0$.

QFI is an information measure quantifying the amount of information $\rho_\theta$ carries about $\theta$ with many nice properties:

1. **(Non-negativity).** $F(\rho_\theta) \geq 0$ and $F(\rho_\theta) = 0$ only when $\partial_\theta\rho_\theta = 0$.

2. **(Monotonicity).** $F(\mathcal{N}(\rho_\theta)) < F(\rho_\theta)$, where $\mathcal{N}$ is an arbitrary $\theta$-independent quantum channel (or CPTP map).

   This is obvious from the fact that the QFI is the optimized classical Fisher information.

3. **(Additivity).** $F(\rho_\theta \otimes \sigma_\theta) = F(\rho_\theta) + F(\sigma_\theta)$.

   This can be proven by noting that the SLD operator is local.

4. **(Convexity).** $F(p\rho_\theta + (1-p)\sigma_\theta) \leq pF(\rho_\theta) + (1-p)F(\sigma_\theta)$.

This can be proven by noting that $F(p\rho_\theta \oplus (1-p)\sigma_\theta) = F(p\rho_\theta) + F((1-p)\sigma_\theta)$ and that there is a channel mapping $p\rho_\theta \oplus (1-p)\sigma_\theta$ to $p\rho_\theta + (1-p)\sigma_\theta$.

There are also other types of QFI [Petz and Ghinea, 2010] which also satisfies the properties above, e.g. the RLD QFI [Yuen and Lax, 1973], but the SLD QFI is the minimal QFI among all definitions. We are going to focus only on the SLD QFI (Eq. (2.27)) in this thesis, and we will always be referring to the SLD QFI when we talk about the QFI.

### 2.2.2 Quantum Cramér–Rao bound

For multi-parameter estimation, the QFI matrix will no longer necessarily be equal to the classical FI matrix maximized over all measurements, because the optimal measurement with respect to different parameters may not be compatible with each other and but it is still possible to prove the SLD QFI matrix is an upper bound of the classical FI matrix using the same technique. We will have the following multi-parameter quantum Cramér–Rao bound based on the SLD QFI matrix [Helstrom, 1976; Holevo, 1982].

**Theorem 2.3** (Quantum CR bound)**.** *For any (locally) unbiased estimate,*

$$\mathbb{V}[\hat{\boldsymbol{\theta}}] \geq F(\rho_{\boldsymbol{\theta}})^{-1}, \tag{2.31}$$

*where $F(\rho_{\boldsymbol{\theta}})$ is the QFI matrix defined by*

$$F(\rho_{\boldsymbol{\theta}})_{ij} = \mathrm{Re}[\mathrm{Tr}\,(\rho_{\boldsymbol{\theta}} L_i L_j)], \tag{2.32}$$

*and $L_i$ is the SLD operator satisfying $\partial_i \rho_{\boldsymbol{\theta}} = \frac{1}{2}(L_i \rho_{\boldsymbol{\theta}} + \rho_{\boldsymbol{\theta}} L_i)$. For one-parameter estimation, the bound is saturable asymptotically.*

*Proof.* This proof can be found in e.g. [Petz and Ghinea, 2010]. For any quantum measurement $\{M_x\}$ and estimators $\hat{\boldsymbol{\theta}}$, according Theorem 2.2

$$\mathbb{V}[\hat{\boldsymbol{\theta}}] \geq F(p_{\boldsymbol{\theta}})^{-1}, \tag{2.33}$$

where

$$F(p_{\boldsymbol{\theta}})_{ij} = \sum_{x\in\Omega} \frac{1}{p_{\boldsymbol{\theta}}(x)} \frac{\partial p_{\boldsymbol{\theta}}(x)}{\partial \theta_i} \frac{\partial p_{\boldsymbol{\theta}}(x)}{\partial \theta_j} \tag{2.34}$$

and $p_{\boldsymbol{\theta}}(x) = \mathrm{Tr}(M_x \rho_{\boldsymbol{\theta}})$. Let $\boldsymbol{v} \in \mathbb{R}^P$ be an arbitrary vector,

$$\begin{aligned}
\boldsymbol{v}^T F(p_{\boldsymbol{\theta}}) \boldsymbol{v} &= \sum_{x:p_{\boldsymbol{\theta}}(x)\neq 0} \frac{1}{p_{\boldsymbol{\theta}}(x)} \frac{\partial p_{\boldsymbol{\theta}}(x)}{\partial \theta_i} \frac{\partial p_{\boldsymbol{\theta}}(x)}{\partial \theta_j} = \sum_{x:p_{\boldsymbol{\theta}}(x)\neq 0} \frac{1}{p_{\boldsymbol{\theta}}(x)} \left(\boldsymbol{v}\cdot \nabla p_{\boldsymbol{\theta}}(x)\right)^2 \\
&= \sum_{x:p_{\boldsymbol{\theta}}(x)\neq 0} \frac{\left(\sum_{i=1}^P v_i \,\mathrm{Re}[\mathrm{Tr}(M_x L_i \rho_{\boldsymbol{\theta}})]\right)^2}{p_{\boldsymbol{\theta}}(x)} \\
&\leq \sum_{x:p_{\boldsymbol{\theta}}(x)\neq 0} \mathrm{Tr}\left(M_x \left(\sum_{i=1}^P v_i L_i\right) \rho_{\boldsymbol{\theta}} \left(\sum_{j=1}^P v_j L_j\right)\right) \\
&= \sum_{i,j=1}^P v_i v_j \mathrm{Tr}(L_i \rho_{\boldsymbol{\theta}} L_j) = \sum_{i,j=1}^P v_i v_j \mathrm{Re}[\mathrm{Tr}(L_i \rho_{\boldsymbol{\theta}} L_j)] = \boldsymbol{v}^T F(\rho_{\boldsymbol{\theta}}) \boldsymbol{v},
\end{aligned} \tag{2.35}$$

where we use the Cauchy–Schwarz inequality and the fact that $\sum_{i,j=1}^P v_i v_j \mathrm{Tr}(L_i \rho_{\boldsymbol{\theta}} L_j)$ is real. For all $v \in \mathbb{R}^P$, $v^T F(p_{\boldsymbol{\theta}}) v \leq v^T F(\rho_{\boldsymbol{\theta}}) v$, then we must have $F(p_{\boldsymbol{\theta}}) \leq F(\rho_{\boldsymbol{\theta}})$ and $F(p_{\boldsymbol{\theta}})^{-1} \geq F(\rho_{\boldsymbol{\theta}})^{-1}$. Combing it with the classical CR bound gives the quantum CR bound. □

As in the classical case, we could consider the quantum CR bound of unbiased measurements and estimators on $N$ samples and get $\mathbb{V}[\hat{\boldsymbol{\theta}}_N] \geq F(\rho_{\boldsymbol{\theta}})^{-1}/N$, using the additivity of $F(\rho_{\boldsymbol{\theta}})$. Similar argument works for the bounds introduced later as well. Althouth the optimal measurement depends on the true parameter, one could use adaptive measurements to attain the lower bound [Barndorff-Nielsen and Gill, 2000; Gill and Massar, 2000; Hayashi, 2005].

### 2.2.3 Holevo bound and Matsumoto bound

The quantum CR bound based on the SLD QFI matrix is not saturable in general. We will provide two more informative bounds in this section: the Holevo bound [Holevo, 1982] and the Matsumoto bound [Matsumoto, 2002]. We already know from Section 2.2.1 that individual measurements on i.i.d. copies are sufficient to achieve the QFI for one-parameter estimation. The Holevo bound was proven to be asymptotically saturable using the experimentally more challenging collective measurements (measurements that might be correlated

| Bounds | Subject | Asymptotic Attainability |
|---|---|---|
| Classical CR bound | $p_{\boldsymbol{\theta}}(x)$ | Yes |
| (SLD) Quantum CR bound | $\rho_{\boldsymbol{\theta}}$ | One-parameter: Yes (individual measurements)<br>Not always in general |
| Holevo bound | $\rho_{\boldsymbol{\theta}}$ | Yes (collective measurements) |
| Matsumoto bound | $\|\psi_{\boldsymbol{\theta}}\rangle$ | Yes (individual measurements) |

Table 2.1: Summary of the MSE bounds under the local unbiasedness condition

among the i.i.d. copies of parameterized quantum states) [Hayashi and Matsumoto, 2008; Kahn and Guţă, 2009; Yamagata et al., 2013; Yang et al., 2019]. The Matsumoto bound is a reformulation of the Holevo bound in the case of pure states, but is asymptotically saturable using individual measurements. It will be useful in Section 3.4.2 where we discuss optimal QEC protocols for multi-parameter estimation.

In multi-parameter estimation, instead of having a mean square error $\Delta^2\hat{\theta}$, we have a mean square matrix $\mathbb{V}[\hat{\boldsymbol{\theta}}]$. In order to find the optimal measurement, we will need a real function as a figure of merit to minimize. Usually, one choose the weighted mean square error

$$\Delta_W^2\hat{\boldsymbol{\theta}} = \mathrm{Tr}(W\mathbb{V}[\hat{\boldsymbol{\theta}}]), \tag{2.36}$$

where $W$ is a real positive matrix that determines the weight we associate with each parameter. The larger the weight is on a parameter, the more we care about its accuracy. For simplcity, we will also call $\Delta_W^2\hat{\boldsymbol{\theta}}$ the MSE in the multi-parameter case.

**Theorem 2.4** (Holevo bound)**.** *For any (locally) unbiased estimate,*

$$\Delta_W^2\hat{\boldsymbol{\theta}} \geq \min_{\{X_i\}} \mathrm{Tr}(W\mathrm{Re}[V]) + \|\sqrt{W}\mathrm{Im}[V]\sqrt{W}\|_1, \tag{2.37}$$

*where* $V_{ij} = \mathrm{Tr}(X_iX_j\rho_{\boldsymbol{\theta}})$ *for Hermitian* $X_i \in \mathscr{L}(\mathcal{H})$*, satisfying* $\mathrm{Tr}(X_i\partial_j\rho_{\boldsymbol{\theta}}) = \delta_{ij}$*.* $\|\cdot\|_1$ *denotes the trace norm.*

*Proof.* The proof can be found in e.g. [Hayashi, 2005, Chapter 8]. For unbiased estimation,

let $X_i = \sum_{x\in\Omega}(\hat{\theta}_i(x) - \theta_i)M_x$. Then the local unbiasedness condition reads

$$\mathrm{Tr}(\rho_{\boldsymbol{\theta}} X_j) = 0, \quad \mathrm{Tr}(X_j \partial_i \rho_{\boldsymbol{\theta}}) = \delta_{ij}, \quad \forall i, j. \tag{2.38}$$

For an arbitrary complex vector $\boldsymbol{u} \in \mathbb{R}^P$, define function $g$: $\mathbb{R}^P \to \mathbb{C}$ and the operator $G$ by

$$g(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{P} v_i(\hat{\theta}_i - \theta_i), \quad G = \sum_{i=1}^{P} v_i X_i. \tag{2.39}$$

Clearly, $\sum_{x\in\Omega} g(\hat{\boldsymbol{\theta}}(x))M_x = G$ and

$$\sum_{x\in\Omega} (g(\hat{\boldsymbol{\theta}}(x)) - G)M_x(g(\hat{\boldsymbol{\theta}}(x)) - G)^\dagger \geq 0 \quad \Rightarrow \quad \sum_{x\in\Omega} \left|g(\hat{\boldsymbol{\theta}}(x))\right|^2 M_x \geq GG^\dagger. \tag{2.40}$$

Then

$$\sum_{x\in\Omega} \left|g(\hat{\boldsymbol{\theta}}(x))\right|^2 \mathrm{Tr}(\rho_{\boldsymbol{\theta}} M_x) \geq \mathrm{Tr}(\rho_{\boldsymbol{\theta}} GG^\dagger) \quad \Rightarrow \quad \boldsymbol{u}^\dagger \mathbb{V}[\hat{\boldsymbol{\theta}}]\boldsymbol{u} \geq \boldsymbol{u}^\dagger V \boldsymbol{u}. \tag{2.41}$$

Note that $\mathbb{V}[\hat{\boldsymbol{\theta}}]$ is a real symmetric matrix, then $\mathbb{V}[\hat{\boldsymbol{\theta}}] \geq V$ and $\mathbb{V}[\hat{\boldsymbol{\theta}}] \geq V^T$. This leads to

$$\sqrt{W}(\mathbb{V}[\hat{\boldsymbol{\theta}}] - \mathrm{Re}[V])\sqrt{W} \geq \pm i\sqrt{W}\mathrm{Im}[V]\sqrt{W}, \tag{2.42}$$

and for any $\boldsymbol{v} \in \mathbb{C}^P$,

$$\boldsymbol{v}^\dagger(\sqrt{W}(\mathbb{V}[\hat{\boldsymbol{\theta}}] - \mathrm{Re}[V])\sqrt{W})\boldsymbol{v} \geq |\boldsymbol{v}^\dagger\sqrt{W}\mathrm{Im}[V]\sqrt{W}\boldsymbol{v}| \tag{2.43}$$

Taking $\boldsymbol{v}$ to be eigenvectors of $\sqrt{W}\mathrm{Im}[V]\sqrt{W}$ and adding them up, we get

$$\mathrm{Tr}(\sqrt{W}(\mathbb{V}[\hat{\boldsymbol{\theta}}] - \mathrm{Re}[V])\sqrt{W}) \geq \|\sqrt{W}\mathrm{Im}[V]\sqrt{W}\|_1, \tag{2.44}$$

proving the bound. Note that we don't need the constraint $\mathrm{Tr}(\rho_{\boldsymbol{\theta}} X_i) = 0$ because $V$ is always minimized at the point where $\mathrm{Tr}(\rho_{\boldsymbol{\theta}} X_i) = 0$ for all $i$ and this constraint is naturally satisfied. □

Note that if we drop the second term on the RHS of Eq. (2.37), it gives the SLD quantum CR bound. Moreover, the SLD bound (Theorem 2.3) is equivalent to the Holevo bound if

and only if if the weak commutation condition $\mathrm{Im}[\mathrm{Tr}(\rho_{\boldsymbol{\theta}} L_i L_j)] = 0$ [Ragy et al., 2016] is satisfied, and is thus saturable using collective measurements. The Holevo bound is recently found to be computable via semidefinite programs (SDP) [Albarelli et al., 2019].

In the case of pure states $\rho_{\boldsymbol{\theta}} = |\psi_{\boldsymbol{\theta}}\rangle \langle\psi_{\boldsymbol{\theta}}|$, the Holevo bound is exactly equivalent to the Matsumoto bound [Matsumoto, 2002]:

**Theorem 2.5** (Matsumoto bound)**.** *For any (locally) unbiased estimate,*

$$\Delta^2_W \hat{\boldsymbol{\theta}} \geq \min_{\{|x_i\rangle\}} \mathrm{Tr}(WV), \quad \textit{where} \;\; V_{ij} = \langle x_i | x_j \rangle \,, \tag{2.45}$$

*where* $|x_i\rangle \in \mathrm{span}\{|\psi_{\boldsymbol{\theta}}\rangle \,, \partial_1 |\psi_{\boldsymbol{\theta}}\rangle \,, ..., \partial_P |\psi_{\boldsymbol{\theta}}\rangle\} \oplus \mathbb{C}^P$ *satisfying* $2\mathrm{Re}[\langle x_i | \partial_j | \psi_{\boldsymbol{\theta}}\rangle] = \delta_{ij}$, $\langle x_i | \psi_{\boldsymbol{\theta}}\rangle = 0$ *and* $\mathrm{Im}[V] = 0$. *Moreover, the bound is saturable asymptotically using individual measurements.*

*Proof.* The proof can be found in [Matsumoto, 2002; Górecki et al., 2020]. According to the Naimark's theorem [Holevo, 1982], any general measurement $\{M_\ell\}_{\ell\in\Omega}$ on $\mathcal{H}$ there exists a projective measurement $\{E_\ell\}_{\ell\in\Omega}$ on an extended space $\mathcal{H}'$ (where $\mathcal{H} \subseteq \mathcal{H}'$) satisfying $M_\ell = \Pi_{\mathcal{H}} E_\ell \Pi_{\mathcal{H}}$ (where $\Pi_{\mathcal{H}}$ is the projection onto $\mathcal{H}$), and

$$\sum_{\ell\in\Omega} E_\ell = \mathbb{1}, \quad E_\ell E_{\ell'} = \delta_{\ell\ell'} E_\ell. \tag{2.46}$$

(Note that here we use $\ell$ instead of $x$ to denote measurement outcomes to avoid confusion with $|x_i\rangle$.) We now define a set of vectors $|x_i\rangle \in \mathcal{H}'$:

$$|x_i\rangle = \sum_\ell (\hat{\theta}_i(\ell) - \theta_i) E_\ell \, |\psi_{\boldsymbol{\theta}}\rangle \,. \tag{2.47}$$

One may see that, using $E_\ell E_{\ell'} = \delta_{\ell\ell'} E_\ell$, inner products of vectors $|x_i\rangle$ yield the mean square matrix:

$$V_{ij} = \langle x_i | x_j \rangle = \sum_{\ell,\ell'} \langle \psi_{\boldsymbol{\theta}}| \, (\hat{\theta}_i(\ell) - \theta_i) E_\ell E_{\ell'} (\hat{\theta}_j(\ell') - \theta_j) \, |\psi_{\boldsymbol{\theta}}\rangle = \mathbb{V}[\hat{\boldsymbol{\theta}}]. \tag{2.48}$$

Now, instead of minimizing $\mathrm{Tr}(W\mathbb{V}[\hat{\boldsymbol{\theta}}])$ over the measurement $\{M_\ell\}$ on $\mathcal{H}$, we can per-

form the minimization directly over the vectors $|x_i\rangle \in \mathcal{H}'$, imposing the following constraints:

$$\langle x_i|\psi_{\boldsymbol{\theta}}\rangle = 0, \quad 2\mathrm{Re}[\langle x_i|\partial_j|\psi_{\boldsymbol{\theta}}\rangle] = \delta_{ij}, \quad \mathrm{Im}(\langle x_i|x_j\rangle) = 0. \tag{2.49}$$

The first two constraints are the local unbiasedness condition and the last is implied by Eq. (2.48). At this point one may wonder how big the space $\mathcal{H}'$ should be (as for a general measurement it might be arbitrary large). However, we can always map $\mathrm{span}\{|\psi_{\boldsymbol{\theta}}\rangle, \{\partial_i |\psi_{\boldsymbol{\theta}}\rangle, |x_i\rangle\}_{i=1}^P\} \subseteq \mathcal{H}'$ isometrically to a $(2P+1)$-dimensional space. Therefore when looking for the bound, under the constraint Eq. (2.49), it is enough to perform the minimization over $|x_i\rangle \in \mathrm{span}\{|\psi_{\boldsymbol{\theta}}\rangle, \partial_1 |\psi_{\boldsymbol{\theta}}\rangle, \cdots, \partial_P |\psi_{\boldsymbol{\theta}}\rangle\} \oplus \mathbb{C}^P$.

Finally, we show that indeed for any set of $|x_i\rangle$ satisfying Eq. (2.49) there exists a proper projective measurement on $\mathcal{H} \oplus \mathbb{C}^P$ and a locally unbiased estimator satisfying Eq. (2.47), and consequently there exists a measurement on $\mathcal{H}$ saturating the bound. To see this, notice that since $\forall i$, $\langle \psi_{\boldsymbol{\theta}}|x_i\rangle = 0$ and $\forall i, j$, $\langle x_i|x_j\rangle \in \mathbb{R}$, one may choose an orthonormal basis $\{|b_i\rangle\}$ of $\mathrm{span}\{|\psi_{\boldsymbol{\theta}}\rangle, |x_1\rangle, \ldots, |x_P\rangle\}$ satisfying: $\forall i$, $\langle \psi_{\boldsymbol{\theta}}|b_i\rangle \in \mathbb{R}\backslash\{0\}$ and $\forall i, j$, $\langle x_i|b_j\rangle \in \mathbb{R}$. Then one can define a projective measurement:

$$E_\ell = |b_\ell\rangle\langle b_\ell| \ \ (\ell = 1, \ldots, P+1), \quad E_0 = \mathbb{1}_{\dim(\mathcal{H}')} - \sum_{\ell=1}^{P+1} |b_\ell\rangle\langle b_\ell|, \tag{2.50}$$

(we use $\mathbb{1}_d$ to denote identity operator on a $d$-dimensional Hilbert space) with the corresponding estimator:

$$\hat{\theta}_i(\ell) = \frac{\langle b_\ell|x_i\rangle}{\langle b_\ell|\psi_{\boldsymbol{\theta}}\rangle} + \theta_i, \ \ell \geq 1, \quad \hat{\theta}_i(0) = 0, \tag{2.51}$$

which is locally unbiased and satisfies

$$|x_i\rangle = \sum_{\ell=1}^{P+1} (\hat{\theta}_i(\ell) - \theta_i) E_\ell |\psi_{\boldsymbol{\theta}}\rangle. \tag{2.52}$$

□

All the MSE lower bounds introduced in Section 2.1 and Section 2.2 are summarized in Table 2.1 in terms of its scope of application and its asymptotic attainability.

## 2.3 Quantum channel estimation

In Section 2.2, we introduced quantum estimation theory and presented several well-known Cramér–Rao type bounds that will be useful in the thesis. However, we focused on quantum state estimation where quantum measurements are optimized to minimize the estimation error. In this section, we will focus on optimization of the input state of one-parameter quantum channels such that the QFI of the output state is maximized. We will mostly follow the discussion in [Fujiwara and Imai, 2008].

### 2.3.1 Channel QFI

Before we discuss channel QFI, we will first prove a useful lemma which represents the QFI using the purification technique [Fujiwara and Imai, 2008, Theorem 1].

**Lemma 2.1** (Purification-based QFI)**.**

$$F(\rho_\theta) = 4 \min_{|\psi_\theta\rangle : \mathrm{Tr}_E(|\psi_\theta\rangle\langle\psi_\theta|)=\rho_\theta} \langle\dot{\psi}_\theta|\dot{\psi}_\theta\rangle\,, \tag{2.53}$$

*where* $|\psi_\theta\rangle \in \mathscr{S}(\mathcal{H}_S \otimes \mathcal{H}_E)$ *are purifications of* $\rho_\theta \in \mathscr{S}(\mathcal{H}_S)$*.* $\dot{\star}$ *denotes* $\frac{\partial \star}{\partial \theta}$*. The minimum is attained at* $|\dot{\psi}_\theta\rangle = \frac{1}{2}(L_\theta \otimes \mathbb{1})\,|\psi_\theta\rangle$*.*

*Proof.* Assume $\dim \mathcal{H}_S = \dim \mathcal{H}_E = d$. We will use the vectorization of matrices $|\star\rangle\!\rangle = \sum_{jk}(\star)_{jk}\,|j\rangle\,|k\rangle$, satisfying $A \otimes C|B\rangle\!\rangle = |ABC^T\rangle\!\rangle = (ABC^T) \otimes \mathbb{1}|\mathbb{1}\rangle\!\rangle = \mathbb{1} \otimes (CB^T A^T)|\mathbb{1}\rangle\!\rangle$, $\langle\!\langle A|B\rangle\!\rangle = \mathrm{Tr}(A^\dagger B)$, $\mathrm{Tr}_E(|A\rangle\!\rangle\langle\!\langle B|) = AB^\dagger$ for arbitrary $A$, $B$ and $C$. Then every purification of $\rho_\theta$ could be written as

$$|\psi_\theta\rangle = |R_\theta\rangle\!\rangle = |\rho_\theta^{1/2} U_\theta\rangle\!\rangle, \tag{2.54}$$

where $U_\theta$ is a unitary operator. When the rank of $\rho_\theta$ does not change in the neighborhood of $\theta$ (as we have assumed), there is an operator $J_\theta$ such that $\dot{\rho}_\theta^{1/2} = \frac{1}{2} J_\theta \rho_\theta^{1/2}$, and

$$\dot{R}_\theta = \frac{1}{2} J_\theta R_\theta + R_\theta U_\theta^\dagger \dot{U}_\theta \tag{2.55}$$

On the other hand, $J_\theta$ is a logarithmic derivative in the sense that

$$\dot{\rho}_\theta = \dot{\rho}_\theta^{1/2}\rho_\theta^{1/2} + \dot{\rho}_\theta^{1/2}\rho_\theta^{1/2} = \frac{1}{2}(J_\theta\rho_\theta + \rho_\theta J_\theta^\dagger). \tag{2.56}$$

Let $D_\theta = J_\theta - L_\theta$, we have

$$\langle\dot{\psi}_\theta|\dot{\psi}_\theta\rangle = \mathrm{Tr}(\dot{R}_\theta\dot{R}_\theta^\dagger) = \frac{1}{4}\mathrm{Tr}(\rho_\theta L_\theta^2) + \mathrm{Tr}\left(\left(\frac{1}{2}D_\theta R_\theta + R_\theta U_\theta^\dagger\dot{U}_\theta\right)\left(\frac{1}{2}D_\theta R_\theta + R_\theta U_\theta^\dagger\dot{U}_\theta\right)^\dagger\right), \tag{2.57}$$

where we use $D_\theta\rho_\theta + \rho_\theta D_\theta^\dagger = 0$ and $U_\theta^\dagger\dot{U}_\theta + \dot{U}_\theta^\dagger U_\theta = 0$. Then clearly, the RHS is larger than $\frac{1}{4}F(\rho_\theta)$ because the second term is non-negative. Moreover, the second term is zero if and only if

$$\frac{1}{2}D_\theta R_\theta + R_\theta U_\theta^\dagger\dot{U}_\theta = 0 \quad\Leftrightarrow\quad \frac{1}{2}L_\theta R_\theta = \dot{R}_\theta \quad\Leftrightarrow\quad |\dot{\psi}_\theta\rangle = \frac{1}{2}(L_\theta\otimes\mathbb{1})\,|\psi_\theta\rangle\,. \tag{2.58}$$

□

The purification-based definition of QFI is especially useful in quantum channel estimation. We consider a quantum channel $\mathcal{E}_\theta(\rho) = \sum_{i=1}^r K_i\rho K_i^\dagger$, where $r$ is the rank of the channel. The entanglement-assisted QFI of $\mathcal{E}_\theta$ (see Figure 2.1(a)) is defined by [Fujiwara and Imai, 2008; Kołodyński and Demkowicz-Dobrzański, 2013],

$$\mathfrak{F}_1(\mathcal{E}_\theta) := \max_{\rho\in\mathscr{S}(\mathcal{H}_S\otimes\mathcal{H}_A)} F((\mathcal{E}_\theta\otimes\mathbb{1})(\rho)). \tag{2.59}$$

Here we utilize the entanglement between the probe and an arbitrarily large ancillary system $\mathcal{H}_A$. We will usually omit the word "entanglement-assisted" in the definitions in this thesis for simplicity. When $\mathcal{E}_\theta$ satisfies certain regularity conditions, the channel QFI $\mathfrak{F}_1(\mathcal{E}_\theta)$ can be calculated through the following theorem [Fujiwara and Imai, 2008]:

**Theorem 2.6** (Channel QFI)**.**

$$\mathfrak{F}_1(\mathcal{E}_\theta) = 4 \min_{\substack{\{K_i'\}_{i=1}^{r'}: \\ \mathcal{E}_\theta(\rho)=\sum_i K_i'\rho K_i'^\dagger}} \left\| \mathbf{K}'^\dagger \mathbf{K}' \right\|, \tag{2.60}$$

*where the norm is minimized over all possible Kraus representation of* $\mathcal{E}_\theta$*.* $\|\cdot\|$ *denotes the operator norm and* $\mathbf{K}' = \begin{pmatrix} K_1' \\ K_2' \\ \vdots \\ K_{r'}' \end{pmatrix}$.

*Proof.* Because of the convexity of the QFI, we can assume the input state $|\psi\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A$ is pure without loss of generality.

$$(\mathcal{E}_\theta \otimes \mathbb{1})(|\psi\rangle\langle\psi|) = \sum_{i=1}^{r} (K_i \otimes \mathbb{1}) |\psi\rangle\langle\psi| (K_i{}^\dagger \otimes \mathbb{1}). \tag{2.61}$$

Its purifications in $\mathcal{H}_{SA} \otimes \mathcal{H}_E$ where $\dim \mathcal{H}_E = r'$ are

$$|\Psi\rangle = \sum_{i=1}^{r} (K_i \otimes \mathbb{1}) |\psi\rangle_{SA} \otimes U |i\rangle_E = \sum_{j=1}^{r'} \left( \sum_{i=1}^{r} u_{ji} K_i \otimes \mathbb{1} \right) |\psi\rangle_{SA} \otimes |j\rangle_E = \sum_{j=1}^{r'} \left( K_j' \otimes \mathbb{1} \right) |\psi\rangle_{SA} \otimes |j\rangle_E, \tag{2.62}$$

where $U$ is an arbitrary isometric operator satisfying $U^\dagger U = \mathbb{1}$ and $\langle i|U|j\rangle = u_{ij}$, and $\mathbf{K}' = u\mathbf{K}$ are the corresponding Kraus operators (we use $u$ as a shorthand for $u \otimes \mathbb{1}$ for simplicity).

According Lemma 2.1,

$$\begin{aligned} \mathfrak{F}_1((\mathcal{E}_\theta \otimes \mathbb{1})(|\psi\rangle\langle\psi|)) &= 4 \max_{|\psi\rangle} \min_{U: U^\dagger U = \mathbb{1}} \langle \dot{\Psi} | \dot{\Psi} \rangle \\ &= 4 \max_{|\psi\rangle} \min_{u: u^\dagger u = \mathbb{1}} \mathrm{Tr}\left( |\psi\rangle\langle\psi| (\mathbf{K}'^\dagger \mathbf{K}' \otimes \mathbb{1}) \right) \\ &= 4 \max_{\sigma \in \mathscr{S}(\mathcal{H}_S)} \min_{u: u^\dagger u = \mathbb{1}} \mathrm{Tr}(\sigma \mathbf{K}'^\dagger \mathbf{K}') \end{aligned} \tag{2.63}$$

$$= 4 \max_{\sigma \in \mathscr{S}(\mathcal{H}_S)} \min_{h \in \mathbb{H}_r} \mathrm{Tr}(\sigma(\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K}))$$

$$= 4 \min_{h \in \mathbb{H}_r} \max_{\sigma \in \mathscr{S}(\mathcal{H}_S)} \mathrm{Tr}(\sigma(\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})),$$

where in the fourth line we introduce a new operator $h = iu^\dagger \dot{u} \in \mathbb{H}_r$ (we use $\mathbb{H}_r$ to denote $r \times r$ Hermitian matrices) [Demkowicz-Dobrzański et al., 2012]. In the last step we interchange the maximization and the minimization using Sion's minimax theorem [Komiya, 1988; do Rosário Grossinho and Tersian, 2001]: for convex compact sets $\mathfrak{P} \subset \mathbb{R}^m$ and $\mathfrak{Q} \subset \mathbb{R}^n$ and $g : P \times Q \to \mathbb{R}$ such that $g(x, y)$ is a continuous convex (concave) function in $x$ ($y$) for every fixed $y$ ($x$), then

$$\max_{y \in \mathfrak{Q}} \min_{x \in \mathfrak{P}} g(x, y) = \min_{x \in \mathfrak{P}} \max_{y \in \mathfrak{Q}} g(x, y). \tag{2.64}$$

Here we could simply choose $x = h \in \mathbb{H}_r$ and $y = \mathrm{Tr}_A(|\psi\rangle \langle\psi|) \in \mathscr{S}(\mathcal{H}_S)$. Although $\mathbb{H}_r$ is not a compact set, we could always restrict it to a compact set as long as $\mathfrak{F}_1$ is finite, using the same argument in e.g. [Zhou and Jiang, 2020b, Appx. D]. □

It was later proven that the channel QFI $\mathfrak{F}_1(\mathcal{E}_\theta)$ is computable using SDP [Demkowicz-Dobrzański et al., 2012; Kołodyński and Demkowicz-Dobrzański, 2013] and that the optimal input states are also solvable using SDP [Zhou and Jiang, 2020a] (see Section 4.4.3). We will see the details in Chapter 4. In the following, we will always use the following formula to calculate the channel QFI

$$\mathfrak{F}_1(\mathcal{E}_\theta) = 4 \min_h \|\alpha\|, \tag{2.65}$$

where $h \in \mathbb{H}_r$ and $\alpha = (\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})$ is a Hermitian operator on $\mathcal{H}_S$.

### 2.3.2 HL vs. SQL

Consider $N$ identical copies of the quantum channel $\mathcal{E}_\theta$ (see Figure 2.1(b)), let

$$\mathfrak{F}_N(\mathcal{E}_\theta) := \mathfrak{F}_1(\mathcal{E}_\theta^{\otimes N}) = \max_\rho F((\mathcal{E}_\theta^{\otimes N} \otimes \mathbb{1})(\rho)). \tag{2.66}$$

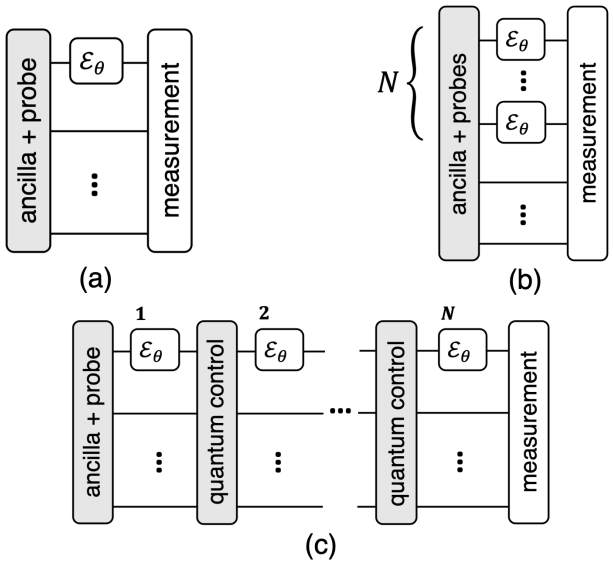


Figure 2.1: (a) The channel QFI $\mathfrak{F}_1(\mathcal{E}_\theta) = \max_\rho F((\mathcal{E}_\theta \otimes \mathbb{1})(\rho))$. The ancillary system is assumed to be at least as large as the probe system. (b) Parallel strategies. $\mathfrak{F}_N(\mathcal{E}_\theta) = \mathfrak{F}_1(\mathcal{E}_\theta^{\otimes N}) = \max_\rho F((\mathcal{E}_\theta^{\otimes N} \otimes \mathbb{1})(\rho))$ for $N$ identical copies of $\mathcal{E}_\theta$. (c) Sequential strategies. Let $F_N(\mathcal{E}_\theta, \mathfrak{S})$ be the QFI of the output state, given a sequential strategy $\mathfrak{S}$ which contains both an input state and quantum controls acting between $\mathcal{E}_\theta$. $\mathfrak{F}_N^{(\mathrm{seq})}(\mathcal{E}_\theta) = \max_{\mathfrak{S}} F_N(\mathcal{E}_\theta, \mathfrak{S})$ is the optimal QFI maximized over all sequential strategies. $\mathfrak{F}_N^{(\mathrm{seq})}(\mathcal{E}_\theta) \geq \mathfrak{F}_N(\mathcal{E}_\theta)$.

Clearly, $\mathfrak{F}_N(\mathcal{E}_\theta) \geq N\mathfrak{F}_1(\mathcal{E}_\theta)$ using the additivity of the QFI. The situation where $\mathfrak{F}_N = \Theta(N)$ is usually the standard quantum limit (SQL) in quantum metrology, which as achievable using product states. In contrast, the Heisenberg limit (HL) $\mathfrak{F}_N(\mathcal{E}_\theta) = \Theta(N^2)$ is only achievable using entangled states. For example, when $\mathcal{E}_\theta$ is unitary, $\mathfrak{F}_N(\mathcal{E}_\theta) = \Theta(N^2)$ is achievable using GHZ-type states [Giovannetti et al., 2006].

The key question in quantum channel estimation is to determine the scaling of the $N$-channel QFI for a given parametrized quantum channel. An upper bound on $\mathfrak{F}_N(\mathcal{E}_\theta)$ could be derived using Theorem 2.6 [Fujiwara and Imai, 2008; Demkowicz-Dobrzański et al., 2012]:

**Theorem 2.7** ($N$-Channel QFI upper bound)**.**

$$\mathfrak{F}_N(\mathcal{E}_\theta) \leq 4 \min_h \left( N \|\alpha\| + N(N-1) \|\beta\|^2 \right), \tag{2.67}$$

*where* $h \in \mathbb{H}_r$ *and* $\beta = i\mathbf{K}^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})$.

*Proof.* Let $K_i^{(1)} = K_i$ for $i \in [r]$, where $[r] = \{1, 2, \ldots, r\}$. Inductively, let

$$K_\iota^{(n+1)} = K_{\iota_1}^{(n)} \otimes K_{\iota_2}^{(1)}, \quad \forall \iota = (\iota_1, \iota_2) \in [r]^n \times [r]. \tag{2.68}$$

$\{K_\iota^{(n)}\}_{\iota \in [r]^n}$ is a Kraus representation of $\mathcal{E}_\theta^{\otimes n}$ for all $n$. Then let $\alpha_n = \sum_{\iota_1} \dot{K}_{\iota_1}^{(n)\dagger} \dot{K}_{\iota_1}^{(n)}, \beta_n = i \sum_{\iota_1} K_{\iota_1}^{(n)\dagger} \dot{K}_{\iota_1}^{(n)}$, we have

$$\alpha_{n+1} = \sum_{\iota_1, \iota_2} \left( \frac{\partial (K_{\iota_1}^{(n)} \otimes K_{\iota_2}^{(1)})}{\partial \theta} \right)^\dagger \left( \frac{\partial (K_{\iota_1}^{(n)} \otimes K_{\iota_2}^{(1)})}{\partial \theta} \right) = \alpha_n \otimes I + 2\beta_n \otimes \beta_1 + I \otimes \alpha_1, \tag{2.69}$$

$$\beta_{n+1} = i \sum_{\iota_1, \iota_2} \left( \frac{\partial (K_{\iota_1}^{(n)} \otimes K_{\iota_2}^{(1)})}{\partial \theta} \right)^\dagger (K_{\iota_1}^{(n)} \otimes K_{\iota_2}^{(1)}) = \beta_n \otimes I + I \otimes \beta_1. \tag{2.70}$$

The solution is $\beta_N = \sum_{k=0}^{N-1} I^{\otimes k} \otimes \beta_1 \otimes I^{\otimes N-1-k}$ and

$$\alpha_N = \sum_{k=0}^{N-1} I^{\otimes k} \otimes \alpha_1 \otimes I^{\otimes N-1-k} + 2 \sum_{k_1=0}^{N-2} \sum_{k_2=0}^{N-2-k_1} I^{\otimes k_1} \otimes \beta_1 \otimes I^{\otimes k_2} \otimes \beta_1 \otimes I^{\otimes N-2-k_1-k_2}. \tag{2.71}$$

Therefore, $\mathfrak{F}_N(\mathcal{E}_\theta) \leq 4\|\alpha_N\| \leq 4N\|\alpha_1\| + 4N(N-1)\|\beta_1\|^2$ and the inequality holds for any Kraus representation of $\mathcal{E}_\theta$. We can choose $\mathbf{K}' = u\mathbf{K}$, then

$$\mathfrak{F}_N(\mathcal{E}_\theta) \leq 4 \min_h \left( N\|\alpha\| + N(N-1)\|\beta\|^2 \right), \tag{2.72}$$

where $h = iu^\dagger \dot{u}$ is an arbitrary Hermitian matrix, $\alpha = \dot{\mathbf{K}}'^\dagger \dot{\mathbf{K}}' = (\dot{\mathbf{K}} - ih\mathbf{K})^\dagger (\dot{\mathbf{K}} - ih\mathbf{K})$ and $\beta = i\mathbf{K}'^\dagger \dot{\mathbf{K}}' = i\mathbf{K}^\dagger (\dot{\mathbf{K}} - ih\mathbf{K})$ . □

If there is an $h$ such that $\beta = 0$,

$$\mathfrak{F}_N(\mathcal{E}_\theta) \leq 4 \min_{h:\beta=0} N \|\alpha\|, \tag{2.73}$$

and $\mathfrak{F}_N(\mathcal{E}_\theta)$ follows the SQL asymptotically. Therefore, it is only possible to achieve the HL if $H \notin \mathcal{S}$ [Zhou and Jiang, 2020a], where

$$H = i\mathbf{K}^\dagger \dot{\mathbf{K}}, \quad \mathcal{S} = \mathrm{span}_{\mathbb{H}}\{K_i^\dagger K_j, \forall i, j\}. \tag{2.74}$$

Here $\text{span}_{\mathbb{H}}\{\cdot\}$ represents all Hermitian operators which are linear combinations of operators in $\{\cdot\}$. We call it the HNKS condition, an acronym for "Hamiltonian-not-in-Kraus-span". One can check that $H$ and $\beta$ are always Hermitian by taking the derivative of $\mathbf{K}^\dagger\mathbf{K} = I$. Note that different Kraus representations may lead to different $H$, but it does not affect the validity of $H \notin \mathcal{S}$. For a unitary channel $r = 1$ and $K_1 = U_\theta = e^{-iH\theta}$, $H = iU_\theta^\dagger \dot{U}_\theta$ is exactly the Hamiltonian for $\theta$, explaining its name. The HL is achievable for unitary channels because $\mathcal{S} = \text{span}_{\mathbb{H}}\{I\}$ and we always have $H \notin \mathcal{S}$ for nontrivial $H$.

The metrological protocols we considered in Figure 2.1(b) are usually called parallel strategies where $N$ identical quantum channels act in parallel on a quantum state [Demkowicz-Dobrzański and Maccone, 2014]. Researchers also consider sequential strategies where we allow quantum controls (arbitrary quantum operations) between each of the quantum channels (see Figure 2.1(c)). Sequential strategies are more powerful than parallel strategies because they can simulate parallel strategies using the same input states and swap operators as quantum controls. The QFI optimized over all sequential strategies, i.e. inputs and quantum controls, has the upper bound [Demkowicz-Dobrzański and Maccone, 2014; Sekatski et al., 2017],

$$\mathfrak{F}_N^{(\text{seq})}(\mathcal{E}_\theta) \leq 4 \min_h \left( N \|\alpha\| + N(N-1) \|\beta\| \left( \|\beta\| + 2\sqrt{\|\alpha\|} \right) \right). \tag{2.75}$$

Therefore, HNKS is also a necessary condition to achieve the HL for sequential strategies. When violated, there exists an $h$ such that $\beta = 0$ and $\mathfrak{F}_N^{(\text{seq})}(\mathcal{E}_\theta)$ has the same upper bound (Eq. (2.73)) as $\mathfrak{F}_N(\mathcal{E}_\theta)$.

To conclude, when $H \in \mathcal{S}$,

$$\mathfrak{F}_N(\mathcal{E}_\theta) \leq \mathfrak{F}_N^{(\text{seq})}(\mathcal{E}_\theta) \leq 4 \min_{h:\beta=0} N \|\alpha\| . \tag{2.76}$$

Therefore, the HL is not achievable for many typical noise channels including depolarization, dephasing, spontaneous emission and photon loss [Escher et al., 2011; Demkowicz-Dobrzański et al., 2012; Demkowicz-Dobrzański and Maccone, 2014], though it is still not known for which channels the HL is achievable (besides unitary channels) and whether these

bounds are achievable. We will discuss in Chapter 4 the asymptotic attainability of these bounds, in particular, we will show $H \notin \mathcal{S}$ is in fact a necessary and sufficient condition to achieve the HL.

## 2.4 Quantum error correction

Finally, we introduce the theory of quantum error correction (QEC) in this chapter. The basic idea of QEC is to encode a logical system into some subspace $\mathcal{H}_C$ of a large physical system $\mathcal{H}_S$ and correct noise using the redundancy in the encoding. Formally, let $\Pi$ be the projector onto the code subspace, $\mathcal{N}$ be the noise channel (CPTP map), then the code is error-correcting if $\mathcal{N}$ is invertible inside the code subspace, i.e. if there exists a recovery channel $\mathcal{R}$ such that

$$\mathcal{R} \circ \mathcal{N}(\rho) = \rho, \quad \forall \rho = \Pi \rho \Pi. \tag{2.77}$$

The study of QEC focuses on finding suitable encoding and recovery channels $(\Pi, \mathcal{R})$ with nice properties.

### 2.4.1 Knill–Laflamme condition

The Knill–Laflamme condition, proven independently by [Bennett et al., 1996] and [Knill and Laflamme, 1997] is a simple set of equations which can be checked to determine whether a quantum error-correcting code $\Pi$ corrects a particular type of noise $\mathcal{N}$.

**Theorem 2.8** (Knill–Laflamme condition)**.** *Let $\Pi$ be a projector onto a quantum code. Suppose $\mathcal{N}$ is a noise channel discribed by $\mathcal{N}(\cdot) = \sum_{i=1}^{r} E_i(\cdot)E_i^\dagger$. A necessary and sufficient condition for the existence of a recovery channel $\mathcal{R}$ satisfying Eq. (2.77) is that*

$$\Pi E_i^\dagger E_j \Pi = \kappa_{ij} \Pi, \tag{2.78}$$

*where $\kappa \in \mathbb{H}_r$.*

*Proof.* The proof can be found in e.g. [Nielsen and Chuang, 2010, Section 10.3].

**(Sufficiency).** We first prove sufficiency by showing that if Eq. (2.78) is satisfied, there exists a recovery channel $\mathcal{R}$ such that $\mathcal{R} \circ \mathcal{N}(\rho) = \rho$, $\forall \rho = \Pi \rho \Pi$. Suppose $\kappa$ is diagonalized

by $\kappa = u\lambda u^\dagger$ where $\lambda$ is a diagonal matrix in $\mathbb{C}^{r\times r}$ and $u$ is unitary. Moreover, $\mathrm{Tr}(\kappa) = \mathrm{Tr}(\lambda) = 1$. Then $\mathbf{F} = u^T\mathbf{E}$ is also a set of Kraus operators for $\mathcal{N}$.

$$\Pi F_k^\dagger F_\ell \Pi = \sum_{ij} u_{ki}^* u_{j\ell} \Pi E_i^\dagger E_j \Pi = (u^\dagger \kappa u)_{k\ell}\Pi = \lambda_k \Pi. \tag{2.79}$$

Using polar decomposition, $F_k\Pi = \sqrt{\lambda_k}U_k\Pi$ for some unitary $U_k$, which allows us to define orthogonal projectors $\Pi_k = U_k\Pi U_k^\dagger$ because when $k \neq \ell$, $\Pi_k\Pi_\ell \propto U_k\Pi F_k^\dagger F_\ell \Pi U_\ell^\dagger = 0$. In the case where $\sum_k \Pi_k \neq \mathbb{1}$, we automatically add $\Pi_{r+1} = \mathbb{1} - \sum_{k=1}^r \Pi_k$ into the set. We could take $U_{r+1} = \mathbb{1}$ and $\lambda_{r+1} = 0$. Then the recovery channel reads

$$\mathcal{R}(\sigma) = \sum_k U_k^\dagger \Pi_k \sigma \Pi_k U_k. \tag{2.80}$$

It satisfies for any $\rho = \Pi\rho\Pi$,

$$\mathcal{R}(\mathcal{N}(\rho)) = \sum_{k\ell} U_k^\dagger P_k F_\ell \rho F_\ell^\dagger P_k U_k = \sum_k U_k^\dagger P_k F_k \rho F_k^\dagger P_k U_k = \sum_k \lambda_k \rho = \rho. \tag{2.81}$$

**(Necessity).** To prove the necessity, suppose there exists $\mathcal{R}(\cdot) = \sum_j R_j(\cdot)R_j^\dagger$ such that

$$\mathcal{R}\circ\mathcal{N}(\Pi\rho\Pi) = \Pi\rho\Pi, \quad \forall\rho. \tag{2.82}$$

Then $\Pi$ and $\{R_jE_i\Pi\}$ are the Kraus operators for the same CP map. We must have

$$R_jE_i\Pi = c_{ji}\Pi \tag{2.83}$$

for some $c_{ij} \in \mathbb{C}$. Then

$$\Pi E_i^\dagger E_j \Pi = \Pi E_i^\dagger \sum_k R_k^\dagger R_k E_j \Pi = \sum_k c_{ki}^* c_{kj} \Pi = \kappa_{ij}\Pi. \tag{2.84}$$

Moreover, $\kappa = c^\dagger c$ is Hermitian. □

The Knill–Laflamme condition will be especially useful in error-corrected metrology when the HL is achievable. In that case, the Knill–Laflamme condition must be satisfied for

the QEC codes and the effective channel after QEC is unitary. Moreover, in Section 3.4.2, we formulate the Knill–Laflamme condition for ancilla-assisted error-correcting code using a semidefinite constraint which might be of independent interest.

### 2.4.2 Approximate QEC

In certain scenarios, it might be difficult to obtain perfect QEC codes satisfying the Knill–Laflamme condition and we might need to consider approximate error-correcting codes.

To characterize approximately error correcting codes, we first define the encoding channel $\mathcal{E}_{S\leftarrow L}(\cdot) = V(\cdot)V^\dagger$, where $V = \sum_{i=1}^{d_L} |\mathfrak{c}_i\rangle_S \langle i|_L$ is an isometry from the logical system $\mathcal{H}_L$ to the physical system $\mathcal{H}_S$ and $\{|\mathfrak{c}_i\rangle\}_{i=1}^{d_L}$ is the codewords. The encoding channel in general could be an arbitrary CPTP map (though CPTP maps are not always reversible), but we usually consider only isometric channels. Then a quantum code is error-correcting under a noise channel $\mathcal{N}_S$ if there exists a CPTP map $\mathcal{R}_{L\leftarrow S}$ such that $\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L} = \mathbb{1}_L$ and approximately error-correcting if $\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L}$ is close to $\mathbb{1}_L$ [Leung et al., 1997]. For example, one could use the (worst-case) entanglement fidelity $f_B(\Phi_1, \Phi_2)$ [Schumacher, 1996; Gilchrist et al., 2005] defined by

$$f_B(\Phi_1, \Phi_2) = \min_\rho f_B((\Phi_1 \otimes \mathbb{1})(\rho), (\Phi_2 \otimes \mathbb{1})(\rho)) \tag{2.85}$$

for two quantum channels $\Phi_1$ and $\Phi_2$ to characterize the infidelity of an approximate QEC code. After optimizing over recovery channels $\mathcal{R}_{L\leftarrow S}$, the infidelity of a code $\mathcal{E}_{S\leftarrow L}$ is defined by

$$\varepsilon(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L}) = 1 - \max_{\mathcal{R}_{L\leftarrow S}} f_B^2(\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L}, \mathbb{1}_L). \tag{2.86}$$

We say a code is $\varepsilon$-correctable if $\varepsilon \geq \varepsilon(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L})$. One could also consider the average infidelity $\bar{\varepsilon}(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L})$ by replacing $f_B(\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L}, \mathbb{1}_L)$ with the average entanglement fidelity $\bar{f}_B(\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L}, \mathbb{1}_L) = f_B(\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L}(|\Psi\rangle\langle\Psi|), |\Psi\rangle\langle\Psi|)$ where $|\Psi\rangle = \sum_{i=1}^{\dim \mathcal{H}_L} |i\rangle |i\rangle$ is the maximally entangled state. The (worst-case) infidelity is clearly larger than the average infidelity and is also upper bounded by it through $\varepsilon \leq \dim \mathcal{H}_L \cdot (\bar{\varepsilon})^{1/2}$ [Albert et al., 2018].

The central problem for approximate QEC is to find the optimal recovery channel and the code infidelity and there have been many research works on this topic (see e.g. [Barnum and Knill, 2002; Schumacher and Westmoreland, 2002; Audenaert and De Moor, 2002; Reimpell and Werner, 2005; Fletcher et al., 2007; Kosut and Lidar, 2009; Bény and Oreshkov, 2010; Ng and Mandayam, 2010; Tyson, 2010]). The average infidelity can be efficiently computed using convex optimization algorithms [Reimpell and Werner, 2005; Fletcher et al., 2007; Kosut and Lidar, 2009], but there are no known efficient algorithms to compute the worst-case infidelity. However, there are some constructions of analytical recovery channels that are near-optimal [Barnum and Knill, 2002; Ng and Mandayam, 2010; Tyson, 2010].

Here we simply state a lemma from [Bény and Oreshkov, 2010] which will be useful later to compute upper bounds on the code infidelity in Chapter 5:

**Lemma 2.2** ([Bény and Oreshkov, 2010])**.** *A code defined by its projector* $\Pi$ *is* $\varepsilon$*-correctable under a noise channel* $\mathcal{N}(\cdot) = \sum_{i=1}^{r} E_i(\cdot)E_i^\dagger$ *if and only if* $\Pi E_i^\dagger E_j \Pi = A_{ij}\Pi + \Pi \delta A_{ij} \Pi$ *for some* $A_{ij}$ *and* $\delta A_{ij}$ *where* $A_{ij}$ *are the components of a density operator, and* $1 - f_B^2(\mathcal{A} + \delta\mathcal{A}, \mathcal{A}) \leq \varepsilon$ *where* $\mathcal{A}(\rho) = \sum_{ij} A_{ij} \mathrm{Tr}(\rho) \left|i\right\rangle \left\langle j\right|$ *and* $(\mathcal{A}+\delta\mathcal{A})(\rho) = \mathcal{A}(\rho) + \sum_{ij} \mathrm{Tr}(\rho \delta A_{ij}) \left|i\right\rangle \left\langle j\right|$.

This lemma does not guarantee an efficient method to compute the code infidelity, since it does not specify how to find an optimal set of coefficients $(A, \delta A)$. Still, in some cases, a reasonable upper bound on the code infidelity can be found by choosing a good guess for $(A, \delta A)$.

# Chapter 3

# Hamiltonian Estimation under Markovian Noise

The HL is the ultimate parameter estimation limit allowed by quantum mechanics, characterized by a precision scaling $O(1/N^2)$ where $N$ is the number of probes, or $O(1/t^2)$ where $t$ is the total probing time, in contrast to the SQL $O(1/N)$ (or $O(1/t)$) governed by the central limit theorem. Here by "probes" we mean quantum systems which are subject to the parametrized quantum dynamics (see Figure 2.1(b) and Figure 2.1(c)). The quest for such a quadratic enhancement is of central importance in quantum metrology. However, we often need an idealized (highly coherent and controllable) quantum device to achieve it and it was not well understood how such a quantum enhancement was affected by quantum noise, except in some specific cases.

In this chapter, we first study the estimation of one Hamiltonian parameter under general Markovian noise [Ozeri, 2013; Dür et al., 2014; Arrad et al., 2014; Kessler et al., 2014; Sekatski et al., 2017] and found a necessary and sufficient condition [Zhou et al., 2018; Demkowicz-Dobrzański et al., 2017] such that when it is violated, the system is always subject to the SQL, even if arbitrary quantum controls are allowed; when it is satisfied, the HL is achievable using an explicit QEC protocol. It is a linear algebraic condition between the signal Hamiltonian and the Lindblad operators describing the noise, called the "Hamiltonian-not-in-Lindblad-span" (HNLS) condition. The HNLS condition rules out the possibility of achieving the HL in many typical quantum systems, e.g. lossy optical

interferometers; when satisfied, it also provides optimal error-correcting codes recovering the HL, e.g. detecting the Kerr effect in lossy bosonic systems. The optimal QEC code is solvable using SDP.

We then generalize the HNLS condition to multi-parameter cases Górecki et al. [2020], which cover more general metrological problems such as vector field sensing [Imai and Fujiwara, 2007], imaging [Tsang et al., 2016], multiple-arm interferometry [Humphreys et al., 2013; Gessner et al., 2018], etc. Multi-parameter estimation is notoriously difficult to deal with due to incompatibility of optimal measurements for different parameters, even in the noiseless case [Ragy et al., 2016; Yuan, 2016; Kura and Ueda, 2018]. Therefore, to find the optimal QEC protocol for multi-parameter estimation, we need to optimize not only the code, but also the input state and the final measurement. Surprisingly, when the multi-parameter HNLS is satisfied, we found a SDP solving the optimal QEC protocol which includes the QEC sensing conditions as a semidefinite constraint and the Matsumoto bound to optimize the final measurement. Moreover, while previous work dealt with specific types of noiseless dynamics, our algorithm works for arbitrary dynamics with or without noise.

We have established that when the HNLS condition is violated, e.g. local Hamiltonians under depolarizing noise, the HL is unachievable. Usually, when the noise is suppressed, the signal must also be suppressed as well, and it seems that QEC would be useless in this case. To address this issue, we study approximate QEC in quantum metrology and propose a new coding technique called the perturbation coding and prove the optimal estimation precision is always achievable using perturbation codes [Zhou and Jiang, 2020b]. Although the quantum enhancement here is only a constant factor instead of quadratic, the protocol works in a much wider range of scenarios.

Finally, we seek the possibility of alleviating the ancilla-assisting assumption for QEC [Layden et al., 2019]. For one-parameter estimation, we show that when the signal Hamiltonian commutes with the noise, ancilla-free codes are sufficient. We also propose new types of bosonic QEC codes called Chebyshev codes which are especially using in sensing under photon loss.

## 3.1 Sequential strategy

We assume that the probes used for parameter estimation are subject to noise described by a Markovian master equation [Gorini et al., 1976; Lindblad, 1976; Breuer et al., 2002]. In addition to the probe system, the experimentalist also has noiseless ancilla qubits at her disposal. She can apply fast, noiseless quantum gates which act jointly on the ancilla and probe; she can also perform perfect ancilla measurements, and reset the ancillas after measurement.

We endow the experimentalist with these powerful tools because we wish to address, as a matter of principle, how effectively QEC can overcome the deficiencies of the noisy probe system. Our scenario may be of practical interest as well, in hybrid quantum systems where ancillas are available which have a much longer coherence time than the probe. For example, sensing of a magnetic field with a probe electron spin can be enhanced by using a quantum code which takes advantage of the long coherence time of a nearby (ancilla) nuclear spin in diamond [Unden et al., 2016]. In cases where noise acting on the ancilla is weak but not completely negligible, we may be able to use QEC to enhance the coherence time of the ancilla, thus providing better justification for our idealized setting in which the ancilla is effectively noiseless. Our assumption that quantum processing is much faster than characteristic decoherence rates is necessary for QEC to succeed in quantum computing as well as in quantum metrology, and recent experimental progress indicates that this assumption is applicable in at least some realistic settings. For example, in superconducting devices QEC has reached the break-even point where the lifetime of an encoded qubit exceeds the natural lifetime of the constituents of the system [Ofek et al., 2016]; one- and two-qubit logical operations have also been demonstrated [Heeres et al., 2017; Rosenblum et al., 2018]. Moreover, if sensing could be performed using a probe encoded within a noiseless subspace or subsystem [Lidar et al., 1998], then active error correction would not be needed to protect the probe, making the QEC scheme more feasible using near-term technology.

In accord with our assumptions, we adopt the sequential strategy for quantum metrology [Demkowicz-Dobrzański and Maccone, 2014; Sekatski et al., 2017; Yuan, 2016] (see Figure 3.1(a)). In this strategy, a single noisy probe senses the unknown parameter for

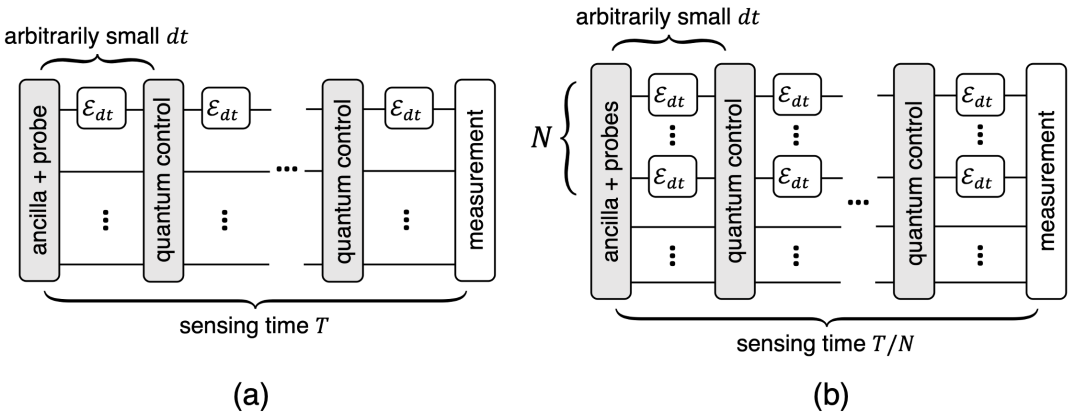


Figure 3.1: (a) Sequential strategies. One probe sequentially senses the parameter for time $T$, with arbitrary quantum controls applied every $dt$. (b) Parallel strategies. $N$ probes sense the parameter for time $T/N$ in parallel.

many rounds, where each round lasts for a short time interval $dt$, and the total number of rounds is $T/dt$, where $T$ is the total sensing time. In between rounds, an arbitrary (noiseless) quantum operation can be applied instantaneously, which acts jointly on the probe and the noiseless ancillas. The rapid operations between rounds empower us to perform QEC, suppressing the damaging effects of the noise on the probe. Note that sequential strategies for a sensing time $T$ can simulate parallel strategies (Figure 3.1(b)), in which $N$ probes simultaneously sense the parameter for time $T/N$ using swap gates between the probe and ancillas. Therefore, we will focus only on sequential strategies in this chapter.

## 3.2 "Hamiltonian-not-in-Lindblad-span" condition

In this section, we will present one of our key theorems—the necessary and sufficient condition to achieve the HL using sequential strategies in terms of one-parameter Hamiltonian estimation under Markovian noise.

We denote the $d$-dimensional Hilbert space of our probe by $\mathcal{H}_S$, and we assume a state $\rho \in \mathcal{H}_S$ of the probe evolves according to a time-homogeneous Lindblad master equation of the form (with $\hbar = 1$) [Gorini et al., 1976; Lindblad, 1976; Breuer et al., 2002],

$$\frac{d\rho}{dt} = -i[H(\theta), \rho] + \sum_{k=1}^{r} (L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}), \tag{3.1}$$

where $H(\theta) = \theta H$ is the probe's Hamiltonian, $\{L_k\}$ are the Lindblad operators, and $r$ is the "rank" of the noise channel acting on the probe (the smallest number of Lindblad operators needed to describe the channel). We assume the Hamiltonian $\theta H$ depends linearly on a parameter $\theta$, and our goal is to estimate $\theta$. The discussion on multi-parameter estimation will be delayed to Section 3.3. Note that our arguments may apply more generally when the Hamiltonian is not a linear function of $\theta$ by intepreting the operator $H$ as the derivative of the Hamiltonian $H(\theta)$ with respect to $\theta$ and by including in the protocol an inverse Hamiltonian evolution step $\exp\left(iH(\hat{\theta})dt\right)$ applied to the probe, where $\hat{\theta}$ is the currently estimated value of $\theta$. We can justify the linear approximation when $\hat{\theta}$ is sufficiently accurate, which works when we update $\hat{\theta}$ adaptively after each round of the estimation [Barndorff-Nielsen and Gill, 2000; Gill and Massar, 2000; Hayashi, 2005; Fujiwara, 2006].

We denote by $\mathcal{H}_A$ the $d$-dimensional Hilbert space of a noiseless ancilla system, whose evolution is determined solely by our fast and accurate quantum controls. Over the small time interval $dt$, during which no controls are applied, the ancilla evolves trivially, and the joint state $\rho \in \mathcal{H}_S \otimes \mathcal{H}_A$ of probe and ancilla evolves according to the quantum channel

$$\mathcal{E}_{dt}(\rho) = \rho - i\theta[H, \rho]dt + \sum_{k=1}^{r}(L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\})dt + O(dt^2), \tag{3.2}$$

where $H$, $L_k$ are shorthand for $H \otimes \mathbb{1}$, $L_k \otimes \mathbb{1}$ respectively. We assume that this time interval $dt$ is sufficiently small that corrections higher order in $dt$ can be neglected. In between rounds of sensing, each lasting for time $dt$, control operations acting on $\rho$ are applied instantaneously.

Our conclusions about the HL and the SQL of parameter estimation make use of an linear algebraic condition on the master equation which we will refer to often, and it will therefore be convenient to have a name for this condition. We will call it the HNLS condition, or simply HNLS, an acronym for "Hamiltonian not in Lindblad span." We denote by $\mathcal{S}$ the linear span of Hermitian matrices generated by the operators $\mathbb{1}, L_k, L_k^\dagger, L_k^\dagger L_j$ (for all $k$ and $j$ ranging from 1 to $r$), i.e.

$$\mathcal{S} = \mathrm{span}_{\mathbb{H}}\{\mathbb{1}, L_k, L_k^\dagger, L_k^\dagger L_j, \forall j, k\}, \tag{3.3}$$

Here $\text{span}_{\mathbb{H}}\{\cdot\}$ represents all Hermitian operators which are linear combinations of operators in $\{\cdot\}$. We say that the Hamiltonian $H$ obeys the HNLS condition, or simply HNLS, if $H$ is not contained in the Lindblad span $\mathcal{S}$. We will also use the HLS condition, or simply HLS, to indicate the situation where $H \in \mathcal{S}$. Now we can state our main conclusion about parameter estimation using fast and accurate quantum controls as Theorem 3.1 [Demkowicz-Dobrzański et al., 2017; Zhou et al., 2018].

**Theorem 3.1** (HNLS). *Consider a finite-dimensional probe with Hamiltonian $H(\theta) = \theta H$, subject to Markovian noise described by a Lindblad master equation with Lindblad operators $\{L_k\}$. Then $\theta$ can be estimated with the HL using sequential strategies if and only if $H$ and $\{L_k\}$ obey the HNLS (Hamiltonian-not-in-Lindblad-span) condition.*

Theorem 3.1 applies if the ancilla is noiseless, and also for an ancilla subject to Markovian noise obeying suitable conditions, as we discuss later in Section 3.2.3. We present an example of the QEC protocol in Section 3.2.1, prove the necessity of HNLS in Section 3.2.2, and the sufficiency of HNLS in Section 3.2.3 using an explicit QEC protocol.

### 3.2.1 Qubit probe

To illustrate how Theorem 3.1 works, let's look at the case where the probe is a qubit, which has been discussed in detail in [Sekatski et al., 2017]. Suppose one of the Lindblad operators is $L_1 \propto \mathbf{n} \cdot \boldsymbol{\sigma}$, where $\mathbf{n} = \mathbf{n_r} + i\mathbf{n_i}$ is a normalized complex 3-vector and $\mathbf{n_r}$, $\mathbf{n_i}$ are its real and imaginary parts, so that $L_1^\dagger L_1 \propto (\mathbf{n}^* \cdot \boldsymbol{\sigma})(\mathbf{n} \cdot \boldsymbol{\sigma}) = \mathbb{1} + 2(\mathbf{n_i} \times \mathbf{n_r}) \cdot \boldsymbol{\sigma}$. We use $^*$ to denote complex conjugate. If $\mathbf{n_r}$ and $\mathbf{n_i}$ are not parallel vectors, then $\mathbf{n_r}$, $\mathbf{n_i}$ and $\mathbf{n_i} \times \mathbf{n_r}$ are linearly independent, which means that $\mathbb{1}$, $L_1$, $L_1^\dagger$, and $L_1^\dagger L_1$ span the four-dimensional space of linear operators acting on the qubit. Hence HNLS cannot be satisfied by any qubit Hamiltonian, and therefore parameter estimation with the HL is not possible according to Theorem 3.1. We conclude that for the HL to be achievable, $\mathbf{n_r}$ and $\mathbf{n_i}$ must be parallel, which means that (after multiplying $L_1$ by a phase factor if necessary) we can choose $L_1$ to be Hermitian [Sekatski et al., 2017]. Moreover, if $L_1$ and $L_2$ are two linearly independent Hermitian traceless Lindblad operators, then $\{\mathbb{1}, L_1, L_2, L_1 L_2\}$ span the space of qubit linear operators and the HL cannot be achieved. In fact, for a qubit

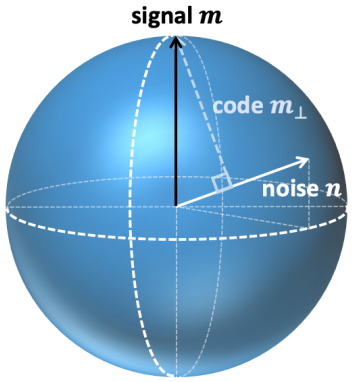


Figure 3.2: The relation between the Hamiltonian, the noise and the QEC code on the Bloch sphere for a qubit probe.

probe, HNLS can be satisfied only if there is a single Hermitian (not necessarily traceless) Lindblad operator $L$, and the Hamiltonian does not commute with $L$.

We will describe below how to achieve the HL for any master equation that satisfies HNLS, by constructing a two-dimensional QEC code which protects the probe from the Markovian noise. To see how the code works for a qubit probe, suppose $H = \frac{1}{2}\mathbf{m}\cdot\boldsymbol{\sigma}$ and $L \propto \mathbf{n}\cdot\boldsymbol{\sigma}$ where $\mathbf{m}$ and $\mathbf{n}$ are unit vectors in $\mathbb{R}^3$ (see Figure 3.2). Then the basis vectors for the QEC code may be chosen to be

$$|\mathfrak{c}_0\rangle = |\mathbf{m}_\perp, +\rangle_S \otimes |0\rangle_A\,, \quad |\mathfrak{c}_1\rangle = |\mathbf{m}_\perp, -\rangle_S \otimes |1\rangle_A\,; \tag{3.4}$$

here $|0\rangle_A$, $|1\rangle_A$ are basis states for the ancilla qubit, and $|\mathbf{m}_\perp, \pm\rangle_S$ are the eigenstates with eigenvalues $\pm 1$ of $\mathbf{m}_\perp\cdot\boldsymbol{\sigma}$ where $\mathbf{m}_\perp$ is the (normalized) component of $\mathbf{m}$ perpendicular to $\mathbf{n}$. In particular, if $\mathbf{m}\perp\mathbf{n}$ (perpendicular noise), then $|\mathfrak{c}_0\rangle = |\mathbf{m}, +\rangle_S \otimes |0\rangle_A$ and $|\mathfrak{c}_1\rangle = |\mathbf{m}, -\rangle_S \otimes |1\rangle_A$, the coding scheme was previously discussed in [Kessler et al., 2014; Arrad et al., 2014; Dür et al., 2014; Ozeri, 2013].

In the case of perpendicular noise, we estimate $\theta$ by tracking the evolution in the code space of a state initially prepared as (in a streamlined notation) $|\psi(0)\rangle = (|+,0\rangle + |-,1\rangle)/\sqrt{2}$; neglecting the noise, this state evolves in time $t$ to

$$|\psi(t)\rangle = \frac{1}{\sqrt{2}}\left(e^{-i\theta t/2}|+,0\rangle + e^{i\theta t/2}|-,1\rangle\right). \tag{3.5}$$

If a jump then occurs at time $t$, the state is transformed to

$$|\psi'(t)\rangle = \frac{1}{\sqrt{2}}\left(e^{-i\theta t/2}|-,0\rangle + e^{i\theta t/2}|+,1\rangle\right). \tag{3.6}$$

Jumps are detected by performing a two-outcome measurement which projects onto either the span of $\{|+,0\rangle, |-,1\rangle\}$ (the code space) or the span of $\{|-,0\rangle, |+,1\rangle\}$ (orthogonal to the code space), and when detected they are immediately corrected by flipping the probe. Because errors are immediately corrected, the error-corrected evolution matches perfectly the ideal evolution (without noise), for which the HL is possible.

When the noise is not perpendicular to the signal, then not just the jumps but also the Hamiltonian evolution can rotate the joint state of probe and ancilla away from the code space. However, after evolution for the short time interval $dt$ the overlap with the code space remains large, so that the projection onto the code space succeeds with probability $1 - O(dt^2)$. Neglecting $O(dt^2)$ corrections, then, the joint probe-ancilla state rotates noiselessly in the code space, at a rate determined by the component of the Hamiltonian evolution along the code space. As long as this component is nonzero, the HL can be achieved.

### 3.2.2 Proof of necessity

Here we show that the QFI of the final state $F(\rho(T))$ is at most asymptotically linear in $T$ when $H \in \mathcal{S}$, which means that the SQL cannot be surpassed in this case and proves the necessity part in Theorem 3.1 (see also [Demkowicz-Dobrzański et al., 2017]).

Suppose the quantum channel describing the joint evolution of probe and ancilla has a Kraus operator representation $\mathcal{E}_{dt}(\rho) = \sum_{k=1}^{r'} K_k \rho K_k^\dagger$, and in terms of these Kraus operators define

$$\alpha_{dt} = (\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K}), \quad \beta_{dt} = i(\dot{\mathbf{K}} - ih\mathbf{K})^\dagger\mathbf{K}, \tag{3.7}$$

where we express the Kraus operators in vector notation

$$\mathbf{K} := \begin{pmatrix} K_1 \\ K_2 \\ \vdots \\ K_{r'} \end{pmatrix} \tag{3.8}$$

and $h$ is the shorthand for $h \otimes \mathbb{1}$ where $h \in \mathbb{C}^{r' \times r'}$ is a Hermitian operator.

Let $\rho_{\text{in}}$ be the initial joint state of probe and ancilla at time $t = 0$, and $\rho(T)$ is the corresponding state at time $T$, then the upper bound on the QFI

$$F(\rho(T)) \leq 4\frac{T}{dt}\|\alpha_{dt}\| + 4\left(\frac{T}{dt}\right)^2 \|\beta_{dt}\| \left((\|\beta_{dt}\| + 2\sqrt{\|\alpha_{dt}\|}\right) \tag{3.9}$$

derived by the channel-extension method holds for any choice of $\rho_{\text{in}}$ even when fast and accurate quantum controls are applied during the evolution [Sekatski et al., 2017] (Theorem 2.7). The upper bound on the QFI provides a lower bound on the estimation precision according to the quantum Cramér–Rao bound (Theorem 2.3).

Here we prove that the QFI scales linearly with the evolution time $T$ in the case where the HNLS condition is violated. We follow the proof in [Sekatski et al., 2017], which applies when the probe is a qubit, and generalize their proof to the case where the probe is $d$-dimensional.

First, we approximate the quantum channel

$$\mathcal{E}_{dt}(\rho) = \rho - i\theta[H, \rho]dt + \sum_{k=1}^{r}(L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\})dt + O(dt^2) \tag{3.10}$$

by the following one:

$$\tilde{\mathcal{E}}_{dt}(\rho) = \sum_{k=0}^{r} K_k \rho K_k^\dagger, \tag{3.11}$$

where $K_0 = I + (-i\theta H - \frac{1}{2}\sum_{k=1}^{r} L_k^\dagger L_k)dt$ and $K_k = L_k\sqrt{dt}$ for $k \geq 1$. The approximation is valid because the distance between $\mathcal{E}_{dt}$ and $\tilde{\mathcal{E}}_{dt}$ is $O(dt^2)$ and the sensing time is divided into $\frac{T}{dt}$ segments, meaning the error $O(\frac{T}{dt} \cdot dt^2) = O(Tdt)$ introduced by this approximation

in calculating the QFI vanishes as $dt \to 0$. Next we calculate the operators $\alpha_{dt} = (\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})$ and $\beta_{dt} = i(\dot{\mathbf{K}} - ih\mathbf{K})^\dagger\mathbf{K}$ for the channel $\tilde{\mathcal{E}}_{dt}(\rho)$, and expand these operators as a power series in $\sqrt{dt}$:

$$\alpha_{dt} = \alpha^{(0)} + \alpha^{(1)}\sqrt{dt} + \alpha^{(2)}dt + O(dt^{3/2}), \tag{3.12}$$

$$\beta_{dt} = \beta^{(0)} + \beta^{(1)}\sqrt{dt} + \beta^{(2)}dt + \beta^{(3)}dt^{3/2} + O(dt^2). \tag{3.13}$$

We will now search for a Hermitian matrix $h$ that sets low-order terms in each power series to zero.

Expanding $h$ as $h = h^{(0)} + h^{(1)}\sqrt{dt} + h^{(2)}dt + h^{(3)}dt^{3/2} + O(dt^2)$ in $\sqrt{dt}$, and using the notation $\mathbf{K} = \mathbf{K}^{(0)} + \mathbf{K}^{(1)}dt^{1/2} + \mathbf{K}^{(2)}dt$, we find

$$\alpha^{(0)} = \mathbf{K}^{(0)\dagger}h^{(0)}h^{(0)}\mathbf{K}^{(0)} = \sum_{k=0}^{r}\left|h_{0k}^{(0)}\right|^2 \mathbb{1} = 0 \Longrightarrow h_{0k}^{(0)} = 0,\ 0 \le k \le r. \tag{3.14}$$

Therefore $h^{(0)}\mathbf{K}^{(0)} = \mathbf{0}$ and $\alpha^{(1)} = \beta^{(0)} = 0$ are automatically satisfied. Then,

$$\beta^{(1)} = -\mathbf{K}^{(0)\dagger}h^{(1)}\mathbf{K}^{(0)} = -h_{00}^{(1)}\mathbb{1} = 0 \Longrightarrow h_{00}^{(1)} = 0. \tag{3.15}$$

and

$$\begin{aligned}\beta^{(2)} &= i\dot{\mathbf{K}}^{(2)\dagger}\mathbf{K}^{(0)} - \mathbf{K}^{(1)\dagger}h^{(0)}\mathbf{K}^{(1)} \\ &\quad - \mathbf{K}^{(0)\dagger}h^{(1)}\mathbf{K}^{(1)} - \mathbf{K}^{(1)\dagger}h^{(1)}\mathbf{K}^{(0)} - \mathbf{K}^{(0)\dagger}h^{(2)}\mathbf{K}^{(0)} \\ &= H - \sum_{k,j=1}^{r} h_{jk}^{(0)}L_k^\dagger L_j - \sum_{k=1}^{r}(h_{0k}^{(1)}L_k + h_{k0}^{(1)}L_k^\dagger) - h_{00}^{(2)}\mathbb{1},\end{aligned} \tag{3.16}$$

which can be set to zero if and only if $H$ is a linear combination of $\mathbb{1}, L_k, L_k^\dagger$ and $L_k^\dagger L_j$ $(0 \le k, j \le r)$.

In addition,

$$\begin{aligned}\beta^{(3)} &= -\mathbf{K}^{(1)\dagger}h^{(1)}\mathbf{K}^{(1)} - \mathbf{K}^{(0)\dagger}h^{(2)}\mathbf{K}^{(1)} \\ &\quad - \mathbf{K}^{(1)\dagger}h^{(2)}\mathbf{K}^{(0)} - \mathbf{K}^{(0)\dagger}h^{(3)}\mathbf{K}^{(0)} \\ &= -\sum_{k,j=1}^{r} h^{(1)}_{jk}L_k^\dagger L_j - \sum_{k=1}^{r}(h^{(2)}_{0k}L_k + h^{(2)}_{k0}L_k^\dagger) - h^{(3)}_{00}\mathbb{1} = 0\end{aligned} \tag{3.17}$$

can be satisfied by setting the above parameters (which do not appear in the expressions for $\alpha^{(0,1)}$ and $\beta^{(0,1,2)}$) all to zero (other terms in $\beta^{(3)}$ are zero because of the constraints on $h^{(0)}$ and $h^{(1)}$ in Eq. (3.14) and Eq. (3.15)). Therefore, when $H$ is a linear combination of $\mathbb{1}, L_k, L_k^\dagger$ and $L_k^\dagger L_j$, there exists an $h$ such that $\alpha_{dt} = O(dt)$ and $\beta_{dt} = O(dt^2)$ for the quantum channel $\tilde{\mathcal{E}}_{dt}$; therefore the QFI obeys

$$\begin{aligned}F(\rho(T)) &\leq 4\frac{T}{dt}\|\alpha_{dt}\| + 4\left(\frac{T}{dt}\right)^2\|\beta_{dt}\|\left(\|\beta_{dt}\| + 2\sqrt{\|\alpha_{dt}\|}\right) \\ &= 4\|\alpha^{(2)}\|T + O(\sqrt{dt}),\end{aligned} \tag{3.18}$$

in which $\alpha^{(2)} = (h^{(1)}\mathbf{K}^{(0)} + h^{(0)}\mathbf{K}^{(1)})^\dagger(h^{(1)}\mathbf{K}^{(0)} + h^{(0)}\mathbf{K}^{(1)})$ under the constraint $\beta^{(2)} = 0$.

We therefore have $\alpha_{dt} = O(dt)$ and $\beta_{dt} = O(dt^2)$, so that the second term in the RHL of Eq. (3.9) vanishes as $dt \to 0$:

$$F(\rho(T)) \leq 4\|\alpha^{(2)}\|T, \tag{3.19}$$

proving that the SQL cannot be surpassed when HNLS is violated (the necessary condition in Theorem 3.1). We require the probe to be finite dimensional in the statement of Theorem 3.1 because otherwise the norm of $\alpha_{dt}$ or $\beta_{dt}$ could be infinite. The theorem can be applied to the case of a probe with an infinite-dimensional Hilbert space if the state of the probe is confined to a finite-dimensional subspace even for asymptotically large $T$.

### 3.2.3 Proof of sufficiency: Code construction

To prove that HNLS is a sufficient condition for achieving the HL, we show that a QEC code achieving the HL can be explicitly constructed if $H \notin \mathcal{S}$ [Zhou et al., 2018].

### The QEC sensing conditions

We first derive three QEC sensing conditions that our QEC code needs to satisfy in order to achieve the HL. Our discussion of the qubit probe indicates how a QEC code can be used to achieve the HL for estimating the parameter $\theta$. The code allows us to correct quantum jumps whenever they occur, and in addition the noiseless error-corrected evolution in the code space depends nontrivially on $\theta$. Similar considerations apply to higher-dimensional probes. Let $\Pi_{\mathfrak{c}}$ denote the projection onto the code space. Jumps are correctable if the code satisfies the error correction conditions [Knill and Laflamme, 1997; Gottesman, 2009; Nielsen and Chuang, 2010], namely:

$$[[1]] \quad \Pi_{\mathfrak{c}} L_k \Pi_{\mathfrak{c}} = \lambda_k \Pi_{\mathfrak{c}}, \ \forall k, \tag{3.20}$$

$$[[2]] \quad \Pi_{\mathfrak{c}} L_k^\dagger L_j \Pi_{\mathfrak{c}} = \mu_{kj} \Pi_{\mathfrak{c}}, \ \forall k, j, \tag{3.21}$$

for some complex numbers $\lambda_k$ and $\mu_{kj}$. The error-corrected joint state of probe and ancilla evolves according to the unitary channel (asymptotically)

$$\frac{d\rho}{dt} = -i\theta[H_{\text{eff}}, \rho] \tag{3.22}$$

where $H_{\text{eff}} = \Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}}$. There is a code state for which the evolution depends nontrivially on $\theta$ provided that

$$[[3]] \quad \Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}} \neq \text{constant} \times \Pi_{\mathfrak{c}}. \tag{3.23}$$

For this noiseless evolution with effective Hamiltonian $\theta H_{\text{eff}}$, the QFI of the encoded state at time $t$ is

$$F(\rho(T)) = 4t^2 \Big[ \text{Tr}(\rho_{\text{in}} H_{\text{eff}}^2) - \big(\text{Tr}(\rho_{\text{in}} H_{\text{eff}})\big)^2 \Big], \tag{3.24}$$

where $\rho_{\text{in}}$ is the initial state at time $t = 0$. The QFI is maximized by choosing the initial pure state

$$|\psi_{\text{in}}\rangle = \frac{1}{\sqrt{2}}(|\lambda_{\min}\rangle + |\lambda_{\max}\rangle), \tag{3.25}$$

where $|\lambda_{\min}\rangle$, $|\lambda_{\max}\rangle$ are the eigenstates of $H_{\text{eff}}$ with the minimum and maximal eigenvalues;

with this choice the QFI is

$$F(\rho(T)) = T^2(\lambda_{\max} - \lambda_{\min})^2. \tag{3.26}$$

By measuring in the appropriate basis at time $T$, we can estimate $\theta$ with a precision that saturates the Cramér–Rao bound in the asymptotic limit of a large number of measurements, hence realizing the HL.

Now we justify the above three conditions. Suppose that a QEC code obeys the conditions [[1]] and [[2]] in Eq. (3.20) and Eq. (3.21), where $\Pi_{\mathfrak{c}}$ is the orthogonal projector onto the code space. We will construct a recovery operator such that the error-corrected time evolution is unitary to linear order in $dt$, governed by the effective Hamiltonian $\theta H_{\text{eff}} = \theta \Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}}$.

For a density operator $\rho = \Pi_{\mathfrak{c}} \rho \Pi_{\mathfrak{c}}$ in the code space, conditions [[1]] and [[2]] imply

$$\Pi_{\mathfrak{c}} \mathcal{E}_{dt}(\rho) \Pi_{\mathfrak{c}} = \rho - i\theta[\Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}}, \rho]dt + \sum_{k=1}^{r} (|\lambda_k|^2 - \mu_{kk})\rho dt + O(dt^2), \tag{3.27}$$

$$\Pi_{\mathfrak{c}}^{\perp} \mathcal{E}_{dt}(\rho) \Pi_{\mathfrak{c}}^{\perp} = \sum_{k=1}^{r} (L_k - \lambda_k)\rho(L_k^{\dagger} - \lambda_k^*)dt + O(dt^2), \tag{3.28}$$

where $\Pi_{\mathfrak{c}}^{\perp} = I - \Pi_{\mathfrak{c}}$. When acting on a state in the code space, $\Pi_{\mathfrak{c}}^{\perp} \mathcal{E}_{dt}(\cdot) \Pi_{\mathfrak{c}}^{\perp}$ is an operation with Kraus operators $K_k = (I - \Pi_{\mathfrak{c}})\, L_k \Pi_{\mathfrak{c}} \sqrt{dt}$, which obey the normalization condition

$$\sum_{k=1}^{r} K_k^{\dagger} K_k = \sum_{k=1}^{r} \Pi_{\mathfrak{c}} L_k^{\dagger} (\mathbb{1} - \Pi_{\mathfrak{c}})\, L_k \Pi_{\mathfrak{c}} dt = \sum_{k=1}^{r} \left(\mu_{kk} - |\lambda_k|^2\right) dt, \tag{3.29}$$

where we have used conditions [[1]] and [[2]]. Therefore, if $\rho$ is in the code space, then a recovery channel $\mathcal{R}_E(\cdot)$ such that

$$\mathcal{R}_E(\Pi_{\mathfrak{c}}^{\perp} \mathcal{E}_{dt}(\rho) \Pi_{\mathfrak{c}}^{\perp}) = -\sum_{k=1}^{r} (|\lambda_k|^2 - \mu_{kk})\rho dt + O(dt^2) \tag{3.30}$$

can be constructed, provided that the operators $\{L_k - \lambda_k\}_{k=1}^{r}$ satisfy the Knill–Laflamme conditions (Theorem 2.8). Indeed these conditions are satisfied because $\Pi_{\mathfrak{c}}(L_k^{\dagger} - \lambda_k^*)(L_j -$

$\lambda_j)\Pi_{\mathfrak{c}} = (\mu_{kj} - \lambda_k^* \lambda_j)\Pi_{\mathfrak{c}}$, for all $k, j$. Therefore, the quantum channel

$$\mathcal{R}(\sigma) = \Pi_{\mathfrak{c}} \sigma \Pi_{\mathfrak{c}} + \mathcal{R}_E(\Pi_{\mathfrak{c}}^{\perp} \sigma \Pi_{\mathfrak{c}}^{\perp}) \tag{3.31}$$

completely reverses the effects of the noise. The channel describing time evolution for time $dt$ followed by an instantaneous recovery step is

$$\mathcal{R}(\mathcal{E}_{dt}(\rho)) = \rho - i\theta[\Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}}, \rho]dt + O(dt^2), \tag{3.32}$$

a noiseless unitary channel with effective Hamiltonian $\theta \Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}}$ if $O(dt^2)$ corrections are neglected.

The dependence of the Hamiltonian on $\theta$ can be detected, for a suitable initial code state $\rho_{\text{in}}$, if and only if $\Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}}$ has at least two distinct eigenvalues. Thus for nontrivial error-corrected sensing we require condition [[3]]: $\Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}} \neq \text{constant} \times \Pi_{\mathfrak{c}}$.

**Error-correctable noisy ancillas**

Above, we assume that a noiseless ancilla system is available for the purpose of constructing the QEC code. Here we deviate a bit from the proof of sufficiency to relax the noiseless ancilla assumption for experimental purpose. We suppose instead that the ancilla is subject to Markovian noise, which is uncorrelated with noise acting on the probe. Hence the joint evolution of probe and ancilla during the infinitesimal time interval $dt$ is described by the quantum channel

$$\begin{aligned}\mathcal{E}_{dt}(\rho) = \rho - i\theta[H \otimes \mathbb{1}, \rho]dt + \sum_{k=1}^{r} \left((L_k \otimes \mathbb{1})\rho(L_k^\dagger \otimes \mathbb{1}) - \frac{1}{2}\{L_k^\dagger L_k \otimes \mathbb{1}, \rho\}\right)dt \\ + \sum_{k'=1}^{r'} \left((\mathbb{1} \otimes L'_k)\rho(\mathbb{1} \otimes L'^\dagger_k) - \frac{1}{2}\{\mathbb{1} \otimes L'^\dagger_k L'_k, \rho\}\right)dt + O(dt^2),\end{aligned} \tag{3.33}$$

where $\{L_k\}$ are Lindblad operators acting on the probe, and $\{L'_k\}$ are Lindblad operators acting on the ancilla.

In this case, we may be able to protect the probe using a code $\bar{C}$ scheme with two layers — an "inner code" $C'$ and an "outer code" $C$. Assuming as before that arbitrarily fast and

accurate quantum processing can be performed, and that the Markovian noise acting on the ancilla obeys a suitable condition, an effectively noiseless encoded ancilla can be constructed using the inner code. Then the QEC scheme that achieves the HL can be constructed using the same method as in the main text, but with the encoded ancilla now playing the role of the noiseless ancilla used in our previous construction.

Errors on the ancilla can be corrected if the projector $\Pi_{\mathfrak{c}'}$ onto the inner code $C'$ satisfies the conditions.

$$[[1']] \quad \Pi_{\mathfrak{c}'} L'_k \Pi_{\mathfrak{c}'} = \lambda'_k \Pi_{\mathfrak{c}'}, \ \forall k, \tag{3.34}$$

$$[[2']] \quad \Pi_{\mathfrak{c}'} L'^{\dagger}_j L'_k \Pi_{\mathfrak{c}'} = \mu'_{jk} \Pi_{\mathfrak{c}'}, \ \forall k, j. \tag{3.35}$$

Eq. (3.34) and Eq. (3.35) resemble Eq. (3.20) and Eq. (3.21), except that the inner code $C'$ is supported only on the ancilla system $\mathcal{H}_A$, while the code $C$ in Eq. (3.20) and Eq. (3.21) is supported on the joint system $\mathcal{H}_S \otimes \mathcal{H}_A$ of probe and ancilla. To search for a suitable inner code $C'$ we may use standard QEC methods; namely we seek an encoding of the logical ancilla with sufficient redundancy for Eq. (3.34) and Eq. (3.35) to be satisfied.

Given a code $C$ that satisfies Eq. (3.20), Eq. (3.21) and Eq. (3.23) for the case of a noiseless ancilla, and a code $C'$ supported on a noisy ancilla that satisfies Eq. (3.34) and Eq. (3.35), we construct the code $\bar{C}$ which achieves the HL for a noisy ancilla system by "concatenating" the inner code $C'$ and the outer code $C$. That is, if the basis states for the code $C$ are $\{|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle\}$, where

$$|\mathfrak{c}_i\rangle = \sum_{j,k=1}^{d} A_{i,jk} \, |j\rangle_S \otimes |k\rangle_A \, , \tag{3.36}$$

then the corresponding basis states for the code $\bar{C}$ are $|\bar{\mathfrak{c}}_0\rangle, |\bar{\mathfrak{c}}_1\rangle$, where

$$|\bar{\mathfrak{c}}_i\rangle = \sum_{j,k=1}^{d} A_{i,jk} \, |j\rangle_S \otimes |\mathfrak{c}'_k\rangle_A \, , \tag{3.37}$$

and $|\mathfrak{c}'_k\rangle$ denotes the basis state of $C'$ which encodes $|k\rangle$. Using our fast quantum controls, the code $C'$ protects the ancilla against the Markovian noise, and the code $\bar{C}$ then protects

the probe, so that the HL is achievable.

In fact the code that achieves the HL need not have this concatenated structure; any code that corrects both the noise acting on the probe and the noise acting on the ancilla will do. For Markovian noise acting independently on probe and ancilla as in Eq. (3.33), the conditions Eq. (3.20) and Eq. (3.21) on the QEC code should be generalized to

$$\Pi_{\bar{\mathfrak{c}}}(O \otimes O')\Pi_{\bar{\mathfrak{c}}} \propto \Pi_{\bar{\mathfrak{c}}}, \quad \forall O \in \mathcal{S} \text{ and } O' \in \mathcal{S}'; \tag{3.38}$$

here $\mathcal{S} = \text{span}\{\mathbb{1}, L_k, L_k^\dagger, L_j^\dagger L_k, \forall k, j\}$, $\mathcal{S}' = \text{span}\{\mathbb{1}, L'_k, L'^\dagger_k, L'^\dagger_j L'_k, \forall k, j\}$ and $\Pi_{\bar{\mathfrak{c}}}$ is the projector onto the code $\bar{C}$ supported on $\mathcal{H}_S \otimes \mathcal{H}_A$. The condition Eq. (3.23) remains the same as before, but now applied to the code $\bar{C}$: $\Pi_{\bar{\mathfrak{c}}}(H \otimes \mathbb{1})\Pi_{\bar{\mathfrak{c}}} \neq \text{constant } \Pi_{\bar{\mathfrak{c}}}$. When these conditions are satisfied, the noise acting on probe and ancilla is correctable; rapidly applying QEC makes the evolution of the probe effectively unitary (and nontrivial), to linear order in $dt$.

### A code satisfying the three QEC sensing conditions

To prove the sufficient condition in Theorem 3.1, we will now show that a code with conditions [[1]]–[[3]] can be constructed whenever HNLS is satisfied. In this code construction we make use of a noiseless ancilla system, but as we discuss in the previous part, the construction can be extended to the case where the ancilla system is subject to Markovian noise obeying suitable conditions.

To see how the code is constructed, note that the $d$-dimensional Hermitian matrices form a real Hilbert space where the inner product of two matrices $A$ and $B$ is defined to be $\text{Tr}(AB)$ (the Hilbert–Schmidt norm) and $H$ has a unique decomposition into $H = H_\parallel + H_\perp$, where $H_\parallel \in \mathcal{S}$ and $H_\perp \perp \mathcal{S}$.

If HNLS holds, then $H_\perp$ is nonzero. It must also be traceless, in order to be orthogonal to $I$, which is contained in $\mathcal{S}$. Therefore, using the spectral decomposition, we can write $H_\perp = \frac{1}{2}\|H_\perp\|_1 (\rho_0 - \rho_1)$, where $\rho_0$ and $\rho_1$ are trace-one positive matrices with orthogonal support.Our QEC code is chosen to be the two-dimensional subspace of $\mathcal{H}_S \otimes \mathcal{H}_A$ spanned by $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$, which are normalized purifications of $\rho_0$ and $\rho_1$ respectively, with orthogonal

support in $\mathcal{H}_A$. (If the probe is $d$-dimensional, a $d$-dimensional ancilla can purify its state.) Because the code basis states have orthogonal support on $\mathcal{H}_A$, it follows that, for any $O$ acting on $\mathcal{H}_S$,

$$\langle \mathfrak{c}_0 | O \otimes \mathbb{1} | \mathfrak{c}_1 \rangle = 0 = \langle \mathfrak{c}_1 | O \otimes \mathbb{1} | \mathfrak{c}_0 \rangle, \tag{3.39}$$

and furthermore

$$\mathrm{Tr}((|\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| - |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|)(O \otimes \mathbb{1})) = \mathrm{Tr}((\rho_0 - \rho_1)O) = \frac{2\ \mathrm{Tr}(H_\perp O)}{\|H_\perp\|_1}. \tag{3.40}$$

In particular, for any $O$ in the span $\mathcal{S}$ we have $\mathrm{Tr}(H_\perp O) = 0$, and therefore

$$\langle \mathfrak{c}_0 | \, (O \otimes \mathbb{1}) \, | \mathfrak{c}_0 \rangle = \langle \mathfrak{c}_1 | \, (O \otimes \mathbb{1}) \, | \mathfrak{c}_1 \rangle. \tag{3.41}$$

Code conditions [[1]]–[[3]] now follow from Eq. (3.39) and Eq. (3.41). For this two-dimensional code, the projector onto the code space is $\Pi_\mathfrak{c} = |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| + |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|$, and therefore

$$\Pi_\mathfrak{c} \, (O \otimes \mathbb{1}) \, \Pi_\mathfrak{c} = \langle \mathfrak{c}_0 | \, (O \otimes \mathbb{1}) \, | \mathfrak{c}_0 \rangle \Pi_\mathfrak{c} \tag{3.42}$$

for $O \in \mathcal{S}$, which implies conditions [[1]] and [[2]] because $L_k$ and $L_k^\dagger L_j$ are in $\mathcal{S}$. Condition [[3]] is also satisfied by the code, because $\langle \mathfrak{c}_0 | H | \mathfrak{c}_0 \rangle - \langle \mathfrak{c}_1 | H | \mathfrak{c}_1 \rangle = 2\ \mathrm{Tr}(H_\perp^2) / \|H_\perp\|_1 > 0$, which means that the diagonal elements of $\Pi_\mathfrak{c} H \Pi_\mathfrak{c}$ are not equal when projected onto the code space. Thus we have demonstrated the existence of a code with conditions [[1]]–[[3]].

## 3.3 Multi-parameter HNLS

We proceed in this section to generalize the HNLS condition to multi-parameter estimation. Again, we assume the dynamics of a $d$-dimensional probe system $\mathcal{H}_S$ is given by:

$$\frac{d\rho}{dt} = -i[H(\boldsymbol{\theta}), \rho] + \sum_{k=1}^{r} (L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}), \tag{3.43}$$

where the parameters to be estimated $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)$ enter linearly into the Hamiltonian of the evolution via Hermitian generators $\boldsymbol{H} = (H_1, \ldots, H_P)$ so that

$$H(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \boldsymbol{H} = \sum_{k=1}^{P} \theta_k H_k. \tag{3.44}$$

We say that the HL in a multi-parameter estimation problem is achieved when there exists an adaptive protocol such that for every $W > 0$, the weighted MSE (Eq. (2.36)) $\Delta_W^2 \hat{\boldsymbol{\theta}} \propto 1/T^2$ in the limit $T \to \infty$. This is equivalent to a requirement that all parameters (and any combination of parameters) are estimated with precision that scales like the HL. The following theorem generalize the HNLS condition proven in Section 3.2 to the multi-parameter case.

**Theorem 3.2** (Multi-parameter HNLS)**.** *The HL can be achieved in a multi-parameter estimation problem using sequential strategies if and only if* $\{(H_i)_\perp, i = 1, \ldots, P\}$ *are linearly independent operators. Here* $(H_i)_\perp$ *are orthogonal projections of* $H_i$ *onto space* $\mathcal{S}^\perp$ *which is the orthogonal complement of the Lindblad span* $\mathcal{S}$*.*

*Proof.* Recall that the necessary and sufficient condition to achieve the HL for one-parameter estimation is $H \notin \mathcal{S}$, or in other words that $H_\perp \neq 0$. In particular, following Section 3.2.3, an explicit construction of the optimal QEC code was provided, where the code space $\mathcal{H}_C \subseteq \mathcal{H}_S \otimes \mathcal{H}_A$ is defined on the Hilbert space of the probe system $\mathcal{H}_S$ extended by an ancillary space $\mathcal{H}_A \cong \mathcal{H}_S$. The code space satisfies the QEC sensing conditions [[1]] and [[2]], i.e.

$$\Pi_{\mathfrak{c}}(S \otimes \mathbb{1})\Pi_{\mathfrak{c}} \propto \Pi_{\mathfrak{c}}, \quad \forall S \in \mathcal{S}, \tag{3.45}$$

where the operator $S$ acting on $\mathcal{H}_S$ was tensored with identity on $\mathcal{H}_A$ and now we use $\Pi_{\mathfrak{c}}$ to denote the projection onto $\mathcal{H}_C$. Metrological sensitivity is guaranteed by the QEC sensing condition condition [[3]]:

$$H_{\text{eff}} = \Pi_{\mathfrak{c}}(H \otimes \mathbb{1})\Pi_{\mathfrak{c}} \not\propto \Pi_{\mathfrak{c}}, \tag{3.46}$$

where we obtain a noiseless unitary evolution generated by $H_{\text{eff}}$ leading to the HL in the estimation precision of $\theta$.

**(Necessity).** Suppose $(H_i)_\perp$'s are linearly dependent. Then there exists a linear (invert-

ible) transformation on the parameter space $A \in \mathbb{R}^{P\times P}$: $\boldsymbol{\theta'} = \boldsymbol{\theta} A^{-1}$, (where we also modify accordingly the generators $\boldsymbol{H'} = A\boldsymbol{H}$ and the cost matrix $W' = AWA^T$, so that $H$ and $\Delta^2_W \hat{\boldsymbol{\theta}}$ remain unchanged), such that $(H'_i)_\perp = 0$ for some $i$. Then, from the one-parameter result, $\theta'_i$ cannot be estimated with precision better than $\Delta^2\hat{\theta}'_i \sim 1/T$ which contradicts the HL requirements.

**(Sufficiency).** Suppose $(H_i)_\perp$'s are linearly independent. We assume the ancillary space to be a direct sum of $P$ subspaces $\mathcal{H}_{A_i}$ so that the whole Hilbert space is $\mathcal{H}_S \otimes (\mathcal{H}_{A_1} \oplus \cdots \oplus \mathcal{H}_{A_P})$. We may construct separate code spaces for each parameter using orthogonal ancillary subspace $\mathcal{H}_{C_i} \subseteq \mathcal{H}_S \otimes \mathcal{H}_{A_i}$ so that the QEC conditions Eq. (3.45) are satisfied within each code space $\mathcal{H}_{C_i}$ separately (see Figure 3.4(a)). While constructing the code space for the $i$-th parameter, we include all the remaining generators $H_j$ $(j \neq i)$ in the Lindblad span, so effectively treating them as noise, i.e. $\mathcal{S}_i = \mathrm{span}_{\mathbb{H}}\{\mathbb{1}, L_k, L_k^\dagger, L_k^\dagger L_{k'}, \forall k, k', H_j, \forall j \neq i\}$. As a result thanks to the QEC condition it follows that $\forall_{i\neq j} \Pi_{\mathfrak{c}_i} (H_j \otimes \mathbb{1}) \Pi_{\mathfrak{c}_i} \propto \Pi_{\mathfrak{c}_i}$ and hence within a given subspace only one parameter is being sensed via the effective generator $H_{\mathrm{eff},i} = \Pi_{\mathfrak{c}_i}(H_i \otimes \mathbb{1})\Pi_{\mathfrak{c}_i}$, while all other generators act trivially. If $|\psi_i\rangle \in \mathcal{H}_{C_i}$ is the optimal state for measuring $\theta_i$, the state to be used in order to obtain HL for all parameters which is not affected by noise reads $\rho_{\mathrm{in}} = \frac{1}{P}\sum_{i=1}^P |\psi_i\rangle\langle\psi_i| \in \mathcal{H}_S \otimes \left(\bigoplus_{i=1}^P \mathcal{H}_{A_i}\right)$—then the optimal measurements with respect to different parameters are compatible (commute with each other) because different parameters are encoded on orthogonal subspaces. □

## 3.4 Optimal QEC protocol: HNLS

In this section, we consider optimization of the QEC protocols when the (multi-parameter) HNLS condition is satisfied. As we will see below, in the one-parameter case, the code optimization is straightforward: we only need to optimize the two-dimensional code $\mathcal{H}_C = \mathrm{span}\{|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle\}$ such that the Hamiltonian has the maximum gap between the minimum and the maximum eigenvalues. In the multi-parameter case, the situation is much more complicated as we will need to optimize not only the code itself but also the input state and the final measurement, which even in noiseless systems does not have an efficient solution before our work.

### 3.4.1 One-parameter: Code optimization

When HNLS is satisfied, we can use our QEC code, along with fast and accurate quantum control, to achieve noiseless evolution of the error-corrected probe, governed by the effective Hamiltonian $\theta\Pi_{\mathfrak{c}}H\Pi_{\mathfrak{c}} = \theta H_{\text{eff}}$ where $\Pi_{\mathfrak{c}}$ is the orthogonal projection onto the code space $\mathcal{H}_C$. Because the optimal initial state Eq. (3.25) is a superposition of just two eigenstates of $H_{\text{eff}}$, a two-dimensional QEC code suffices for achieving the best possible precision. For a code with basis states $\{|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle\}$,

$$H_{\text{eff}} = |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0|H_\perp|\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| + |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|H_\perp|\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|. \tag{3.47}$$

Here we have ignored the contribution due to $H_\parallel$, which is an irrelevant additive constant if the code satisfies condition [[1]] and [[2]]; moreover, we can ignore the off-diagonal terms because we can always choose the basis freely such that $H_{\text{eff}}$ is diagonal. We have seen how to construct a code for which

$$\lambda_{\max} - \lambda_{\min} = 2\,\frac{\text{Tr}(H_\perp^2)}{\|H_\perp\|_1}. \tag{3.48}$$

It is possible, though, that a larger value of this difference of eigenvalues could be achieved using a different code, improving the precision by a constant factor (independent of the time $T$).

To search for a better code, with basis states $\{|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle\}$, define

$$\rho_0 = \text{Tr}_A(|\mathfrak{c}_0\rangle\langle\mathfrak{c}_0|), \quad \rho_1 = \text{Tr}_A(|\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|), \tag{3.49}$$

where we use $\text{Tr}_A$ to represent partial trace where the state in $\mathcal{H}_A$ is traced out. Consider $\tilde{C} = \rho_0 - \rho_1$. Conditions [[1]]–[[2]] on the code imply

$$\text{Tr}(\tilde{C}O) = 0,\ \forall O \in \mathcal{S}, \tag{3.50}$$

and we want to maximize

$$\lambda_{\max} - \lambda_{\min} = \mathrm{Tr}(H_{\mathrm{eff}}\tilde{C}) = \mathrm{Tr}(H_{\perp}\tilde{C}), \tag{3.51}$$

over matrices $\tilde{C}$ of the form $\tilde{C} = \rho_0 - \rho_1$ subject to Eq. (3.50). Note that $\tilde{C}$ is the difference of two normalized density operators, and therefore satisfies $\|\tilde{C}\|_1 \leq 2$. In fact, though, if $\tilde{C}$ obeys the constraint Eq. (3.50), then the constraint is still satisfied if we rescale $\tilde{C}$ by a real constant greater than one, which increases $\mathrm{Tr}(H_{\perp}\tilde{C})$; hence the maximum of $\mathrm{Tr}(H_{\perp}\tilde{C})$ is achieved for $\|\tilde{C}\|_1 = 2$, which means that $\rho_0$ and $\rho_1$ have orthogonal support.

There is a description of the code optimization, with a pleasing geometrical interpretation. As we will see below, the optimization can be formulated as a SDP and the optimal QFI is given by [Zhou et al., 2018]

$$F(\rho(T)) = 4T^2 \min_{S\in\mathcal{S}} \|H_{\perp} - S\|^2. \tag{3.52}$$

In this sense, the QFI is determined by the minimum distance between $H_{\perp}$ and $\mathcal{S}$ in terms of spectral norm (see Figure 3.3(b)).

In the noiseless case ($\mathcal{S} = \mathrm{span}\{\mathbb{1}\}$), the minimum in Eq. (3.52) occurs when the maximum and minimum eigenvalues $H_{\perp} - S$ have the same absolute value, and then the operator norm is half the difference of the maximum and minimum eigenvalues of $H_{\perp}$. Hence we recover the result Eq. (3.26). When noise is introduced, $\mathcal{S}$ swells and the minimum distance shrinks, lowering the QFI and reducing the precision of parameter estimation. If HNLS fails, then the minimum distance is zero, and no QEC code can achieve the HL, in accord with Theorem 3.1.

To obtain Eq. (3.52), we first note that the code optimization can be formulated as the following optimization problem:

$$\begin{aligned} &\text{maximize } \mathrm{Tr}(\tilde{C}H_{\perp}) \\ &\text{subject to } \|\tilde{C}\|_1 \leq 2 \text{ and } \mathrm{Tr}(\tilde{C}S) = 0,\ \forall S \in \mathcal{S}. \end{aligned} \tag{3.53}$$

This optimization problem is convex (because $\|\cdot\|_1$ is convex) and satisfies the Slater's

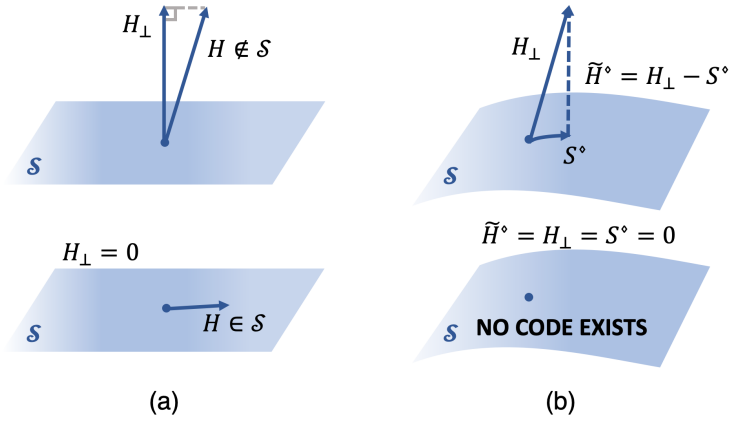


Figure 3.3: Geometric illustration of HNLS and code optimization. (1) $H_\perp$ is the projection of $H$ onto $\mathcal{S}$ in the Hilbert space of Hermitian matrices equipped with the Hilbert–Schmidt norm. $H_\perp \neq 0$ if and only if $H \notin \mathcal{S}$, which is the HNLS condition. (2) $\tilde{H}^\diamond$ is the projection of $H$ onto $\mathcal{S}$ in the linear space of Hermitian matrices equipped with the operator norm. In general, the optimal QEC code can be constructed from $\tilde{H}^\diamond$ and $\tilde{H}^\diamond$ is not necessarily equal to $H_\perp$.

condition, so it can be solved by solving its Lagrange dual problem [Boyd and Vandenberghe, 2004]. The Lagrangian $L(\tilde{C}, \lambda, \nu)$ is defined for $\lambda \geq 0$ and $\nu_k \in \mathbb{R}$:

$$L(\tilde{C}, \lambda, \nu) = \mathrm{Tr}(\tilde{C} H_\perp) - \lambda(\|\tilde{C}\|_1 - 2) + \sum_k \nu_k \mathrm{Tr}(E_k \tilde{C}), \tag{3.54}$$

where $\{E_k\}$ is any basis of $\mathcal{S}$. The optimal value is obtained by taking the minimum of the dual

$$\begin{aligned} g(\lambda, \nu) &= \max_{\tilde{C}} L(\tilde{C}, \lambda, \nu) = \max_{\tilde{C}} \mathrm{Tr}\big((H_\perp + \sum_k \nu_k E_k)\tilde{C} - \lambda|\tilde{C}|\big) + 2\lambda \\ &= \begin{cases} 2\lambda & \lambda \geq \|H_\perp + \sum_k \nu_k E_k\| \\ \infty & \lambda \leq \|H_\perp + \sum_k \nu_k E_k\| \end{cases} \end{aligned} \tag{3.55}$$

over $\lambda$ and $\{\nu_k\}$. Hence the optimal value of the primal problem is

$$\min_{\lambda,\nu} g(\lambda, \nu) = 2 \min_{\nu_k} \|H_\perp + \sum_k \nu_k E_k\| = 2 \min_{S \in \mathcal{S}} \|H_\perp - S\|. \tag{3.56}$$

The optimization problem Eq. (3.56) is equivalent to the following SDP [Boyd and

Vandenberghe, 2004]:

$$\begin{aligned} &\text{minimize } s \\ &\text{subject to } \begin{pmatrix} s\mathbb{1} & H_\perp + \sum_k \nu_k E_k \\ H_\perp + \sum_k \nu_k E_k & s\mathbb{1} \end{pmatrix} \geq 0 \end{aligned} \tag{3.57}$$

for variables $\nu_k \in \mathbb{R}$ and $s \geq 0$. SDPs can be solved using the Matlab-based package CVX [Grant and Boyd].

Once we have the solution to the dual problem we can use it to find the solution to the primal problem. We denote by $\lambda^\diamond$ and $\nu^\diamond$ the values of $\lambda$ and $\nu$ where $g(\lambda, \nu)$ attains its minimum, and define

$$\tilde{H}^\diamond = H_\perp + \sum_k \nu_k^\diamond E_k. \tag{3.58}$$

The minimum $g(\lambda^\diamond, \nu^\diamond)$ matches the value of the Lagrangian $L(\tilde{C}, \lambda^\diamond, \nu^\diamond)$ when $\tilde{C} = \tilde{C}^\Delta$ is the value of $\tilde{H}$ which maximizes $\mathrm{Tr}\left(\tilde{H} H_\perp\right)$ subject to the constraints. This means that

$$\mathrm{Tr}\left(\tilde{C}^\Delta \tilde{H}^\diamond\right) = 2\|\tilde{H}^\diamond\|. \tag{3.59}$$

Since we require $\mathrm{Tr}(\tilde{C}^\Delta) = 0$ and $\|\tilde{C}^\Delta\|_1 = 2$, and because minimizing $g(\lambda, \nu)$ enforces that the maximum and minimum eigenvalues of $\tilde{H}^\diamond$ have the same absolute value and opposite sign, we conclude that

$$\tilde{C}^\Delta = \rho_0^\diamond - \rho_1^\diamond, \tag{3.60}$$

where $\rho_0^\diamond$ is a density operator supported on the eigenspace of $\tilde{H}^\diamond$ with the maximal eigenvalue, and $\rho_1^\diamond$ is a density operator supported on the eigenspace of $\tilde{H}^\diamond$ with the minimum eigenvalue. An $\tilde{C}^\Delta$ of this form which satisfies the constraints of the primal problem is guaranteed to exist and provides the optimal QEC codes.

### 3.4.2 Multi-parameter: Input, encoding and measurement optimization

In Section 3.3, we provided a QEC code where each parameter is sensed separately in different error-corrected subspaces (see Figure 3.4(a)). Such protocols will be referred as

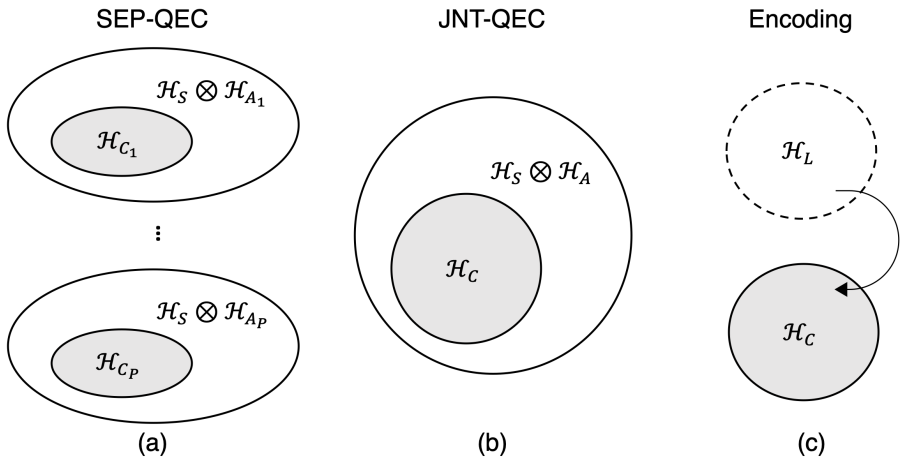


Figure 3.4: Schematic diagrams of relations between Hilbert spaces $\mathcal{H}_S$, $\mathcal{H}_A$, $\mathcal{H}_C$, $\mathcal{H}_{A_i}$, $\mathcal{H}_{C_i}$, $\mathcal{H}_{\text{eff}}$. (a) In SEP-QEC, we use $P$ mutually orthogonal ancillary subspaces $\mathcal{H}_{A_i}$ to sense each parameter $\theta_i$. $\mathcal{H}_A = \bigoplus_{i=1}^{P} \mathcal{H}_{A_i}$ and $\mathcal{H}_C = \bigoplus_{i=1}^{P} \mathcal{H}_{C_i}$. $\dim(\mathcal{H}_{A_i}) = \dim(\mathcal{H}_S) = d$ and $\dim(\mathcal{H}_{C_i}) = 2$. (b) In JNT-QEC, we use a single code space $\mathcal{H}_C \subseteq \mathcal{H}_S \otimes \mathcal{H}_A$ to estimate all parameters jointly. $\dim(\mathcal{H}_A) = (P+1)d$ and $\dim(\mathcal{H}_C) = P+1$. (c) We use $\mathcal{H}_L$ to represent the logical space $\text{span}\{|0\rangle, |1\rangle, \ldots, |P\rangle\}$ which is encoded into the physical space $\mathcal{H}_C = \text{span}\{|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle, \ldots, |\mathfrak{c}_P\rangle\}$.

separate-QEC protocols (SEP-QEC). In contrast to this construction, we will now consider QEC strategies which allow simultaneous estimation of all the parameters in a single coherent protocol by utilizing states within a single protected code space, which we will call the joint-parameter QEC protocol (JNT-QEC). In this section we provide a general method to identify the optimal JNT-QEC, while its potential advantages over the optimal SEP-QEC will be discussed at the end of this subsection.

From now on, we assume the multi-parameter HNLS condition is satisfied. Without loss of generality, in this section, we also assume the generators $\{H_i\}_{i=1}^{P} \subseteq \mathcal{S}^{\perp}$ are orthonormal, since the components in $\mathcal{S}$ do not contribute and there is always a linear transformation $A$ on parameters leading to orthonormality. The following theorem provides a recipe to find the optimal JNT-QEC [Górecki et al., 2020].

**Theorem 3.3** (Optimal JNT-QEC)**.** *Fix a cost matrix $W$. If the multi-parameter HNLS condition is satisfied with generators $\{H_i\}_{i=1}^{P} \subseteq S^{\perp}$, the minimum MSE $\Delta^2_W\hat{\boldsymbol{\theta}}$ that can be achieved in a JNT-QEC (asymptotically) reads*

$$\Delta^2_W\hat{\boldsymbol{\theta}} = \frac{P}{4T^2} \min_{H_{L,i}, B_i, \nu_i, K, w} w, \tag{3.61}$$

*subject to*

$$\mathbb{1}_{P+1} \otimes \frac{\mathbb{1}_d}{d} + \sum_{i=1}^{P} (H_{L,i})^T \otimes H_i + \sum_{i=P+1}^{P'} \nu_i \mathbb{1}_{P+1} \otimes S_i + \sum_{i=P'+1}^{d^2-1} B_i \otimes R_i \geq 0, \tag{3.62}$$

$$\Gamma_{ij} = \mathrm{Im}[H_{L,j}]_{i0}, \quad \begin{pmatrix} w\mathbb{1}_P & K \\ K & \mathbb{1}_P \end{pmatrix} \geq 0, \quad \begin{pmatrix} K & \mathbb{1}_P \\ \mathbb{1}_P & \Gamma\sqrt{W^{-1}} \end{pmatrix} \geq 0, \tag{3.63}$$

*where* $\mathbb{1}_d/\sqrt{d}$, $\{H_i\}_{i=1}^{P}$, $\{S_i\}_{i=P+1}^{P'}$, $\{R_i\}_{i=P'+1}^{d^2-1}$ *form an orthonormal basis of Hermitian operators acting on* $\mathcal{H}_S$ *such that* $\mathcal{S} = \mathrm{span}_{\mathbb{H}}\{\mathbb{1}_d, (S_i)_{i=P+1}^{P'}\}$. *Moreover,* $H_{L,i}$, $B_i$ *are Hermitian* $(P+1) \times (P+1)$ *matrices (with matrix indices taking values from* 0 *to* $P$*), and* $\nu_i \in \mathbb{R}$. $\Gamma$ *and* $K$ *are real* $P \times P$ *matrices (with matrix indices from* 1 *to* $P$*).*

The solution yields an explicit form of the optimal input state, QEC codes and measurements. No collective measurements are required on the output states. Our protocol goes beyond the typically used QFI-based formalism and overcomes all the challenges related with the multi-parameter aspect of the problem, including measurement incompatibility, input state optimization and formulating the QEC conditions. Our work reveals the advantage of QEC protocols in multi-parameter estimation and we expect that the SDP formulation of our problem will also be an inspiration for other research areas in quantum error correction and quantum metrology. The proof is divided into three parts. First, we tailor the Matsumoto bound (Theorem 2.5) in our QEC setting using the advantage of noiseless ancillas. Second, we formulate the optimization of ancilla-assisted code as a semidefinite constraint. Finally, we incorporate the first two steps and add one more semidefinite constraint using the symmetry in the objective function to obtain the SDP in Eq. (3.61).

### Reformulation of the Matsumoto bound

Recall from Theorem 2.5 that for unbiased estimates on pure states $|\psi_{\boldsymbol{\theta}}\rangle\langle\psi_{\boldsymbol{\theta}}| \in \mathcal{H}$,

$$\Delta_W^2 \hat{\boldsymbol{\theta}} \geq \min_{\{|x_i\rangle\}} \mathrm{Tr}(WV), \quad \text{where} \;\; V_{ij} = \langle x_i | x_j \rangle\,, \tag{3.64}$$

where $|x_i\rangle \in \mathcal{H}' = \text{span}\{|\psi_{\boldsymbol{\theta}}\rangle, \partial_1 |\psi_{\boldsymbol{\theta}}\rangle, ..., \partial_P |\psi_{\boldsymbol{\theta}}\rangle\} \oplus \mathbb{C}^P$ satisfying

$$\langle x_i|\psi_{\boldsymbol{\theta}}\rangle = 0, \quad 2\text{Re}[\langle x_i|\partial_j|\psi_{\boldsymbol{\theta}}\rangle] = \delta_{ij}, \quad \text{Im}(\langle x_i|x_j\rangle) = 0. \tag{3.65}$$

The bound is saturable asymptotically using independent measurements. Specifically, if $\dim(\mathcal{H}) \geq 2P + 1$ we may simply choose $\mathcal{H}' \subseteq \mathcal{H}$ and $\text{span}\{|\psi_{\boldsymbol{\theta}}\rangle, \partial_1 |\psi_{\boldsymbol{\theta}}\rangle, ..., \partial_P |\psi_{\boldsymbol{\theta}}\rangle\} \oplus \mathbb{C}^P$ as a subspace of $\mathcal{H}$ and optimize over $|x_i\rangle \in \mathcal{H}$.

In this case we may reformulate the Matsumoto bound in a slightly different form. First, note that any vectors $\{|x_i\rangle\}$ satisfying Eq. (3.65) need to be linearly independent. Let $\{|\mathfrak{c}_i\rangle\}_{i=1}^P$ be an orthonormal basis of $\text{span}\{|x_1\rangle, \ldots, |x_P\rangle\}$, satisfying $\forall_{i,j} \text{Im}\langle x_i|\mathfrak{c}_j\rangle = 0$. Such a set may be generated using the Gram-Schmidt orthonormalization procedure. (Note that we haven't related $\{|\mathfrak{c}_i\rangle\}_{i=1}^P$ to the codewords yet.) The local unbiasedness conditions may now be rewritten as:

$$2\text{Re}[\langle x_i|\partial_j|\psi_{\boldsymbol{\theta}}\rangle] = \sum_{k=1}^{P} 2\text{Re}[\langle x_i|\mathfrak{c}_k\rangle \langle \mathfrak{c}_k|\partial_j|\psi_{\boldsymbol{\theta}}\rangle] = \sum_{k=1}^{P} \langle x_i|\mathfrak{c}_k\rangle \, 2\text{Re}[\langle \mathfrak{c}_k|\partial_j|\psi_{\boldsymbol{\theta}}\rangle] = \delta_{ij}, \tag{3.66}$$

which (after introducing matrices $\mathcal{X}_{ki} = \langle \mathfrak{c}_k|x_i\rangle$, $\mathcal{Y}_{kj} = 2\text{Re}[\langle \mathfrak{c}_k|\partial_j|\psi_{\boldsymbol{\theta}}\rangle]$ is equivalent to the matrix equality $\mathcal{X}^T\mathcal{Y} = \mathbb{1}_P$. From $\mathcal{X}^T\mathcal{Y} = \mathbb{1}_P$ we have $\mathcal{X}^T = \mathcal{Y}^{-1} \Rightarrow \text{Tr}(W \cdot V) = \text{Tr}(W \cdot \mathcal{X}^T\mathcal{X}) = \text{Tr}(W \cdot (\mathcal{Y}^T\mathcal{Y})^{-1})$, which gives

$$\begin{aligned} &\min_{|\mathfrak{c}_1\rangle,\ldots,|\mathfrak{c}_P\rangle \in \mathcal{H}} \text{Tr}(W \cdot (\mathcal{Y}^T\mathcal{Y})^{-1})), \\ &\text{where } \mathcal{Y}_{ij} = 2\text{Re}[\langle \mathfrak{c}_i|\partial_j|\psi_{\boldsymbol{\theta}}\rangle], \quad \text{subject to } \langle \mathfrak{c}_i|\mathfrak{c}_j\rangle = \delta_{ij}. \end{aligned} \tag{3.67}$$

This formulation will be more convenient to use when we will formulate the QEC protocol optimization problem as a SDP.

### Error correction as a semidefinite constraint

Now we apply the reformulated Matsumoto bound to our task of identification of the optimal JNT-QEC. Consider a given input state $|\psi_{\text{in}}\rangle$. Let $\mathcal{H}_C$ be any code subspace of $\mathcal{H}_S \otimes \mathcal{H}_A$ containing $|\psi_{\text{in}}\rangle$ and satisfying the QEC conditions Eq. (3.45)—in order to be in accordance with the reformulated Matsumoto bound, this space may be required to be at least $2P +$

1 dimensional, but as we show in the following it will always be possible to reduce its dimensionality to $P+1$ effectively. Using QEC, our goal is to preserve an effective unitary evolution in the encoded space and coherently acquire the sensing signal. Therefore, we are effectively dealing with pure state $|\psi_{\boldsymbol{\theta}}\rangle$, which allows us to utilize the reformulate Matsumoto bound Eq. (3.67). The effective evolution after implementing QEC is given by

$$|\psi_{\boldsymbol{\theta}}\rangle = \exp\left(-iT\sum_{j=1}^{P}\theta_j \Pi_{\mathfrak{c}}\left(H_j \otimes \mathbb{1}_{\dim(\mathcal{H}_A)}\right)\Pi_{\mathfrak{c}}\right)|\psi_{\text{in}}\rangle\,. \tag{3.68}$$

We focus on the estimation around point $\boldsymbol{\theta} = [0,\ldots,0]$ (which can always be achieved by applying inverse Hamiltonian dynamics [Yuan, 2016]) and denote $|\mathfrak{c}_0\rangle = |\psi_{\boldsymbol{\theta}=0}\rangle$ for notational simplicity. Then for any $|\mathfrak{c}_i\rangle \in \mathcal{H}_C$ we have

$$2\text{Re}[\langle \mathfrak{c}_i|\partial_j|\psi_{\boldsymbol{\theta}=0}\rangle] = 2T\text{Im}[\langle \mathfrak{c}_i|(H_j \otimes \mathbb{1}_{\dim(\mathcal{H}_A)})|\mathfrak{c}_0\rangle], \tag{3.69}$$

and according to Eq. (3.67) the minimum achievable MSE for a fixed code space $\mathcal{H}_C$ is given by:

$$\begin{aligned} &\min_{|\mathfrak{c}_1\rangle,\ldots,|\mathfrak{c}_P\rangle\in\mathcal{H}_C} \text{Tr}(W\cdot(\mathcal{Y}^T\mathcal{Y})^{-1})), \\ &\text{where } \mathcal{Y}_{ij} = 2T\text{Im}[\langle \mathfrak{c}_i|(H_j\otimes\mathbb{1}_{\dim(\mathcal{H}_A)})|\psi_{\boldsymbol{\theta}}\rangle], \quad \text{subject to } \langle \mathfrak{c}_i|\mathfrak{c}_j\rangle = \delta_{ij}. \end{aligned} \tag{3.70}$$

From the above formulation it is clear that we may always reduce the code space $\mathcal{H}_C$ to $\text{span}\{|\mathfrak{c}_k\rangle\}_{k=0}^P$ without increasing the MSE. Hence, the problem of optimization over both probes and error-correction protocols is now equivalent to identification of the set $\{|\mathfrak{c}_k\rangle\}_{k=0}^P$ that minimizes the MSE with the constraint that $\mathcal{H}_C = \text{span}\{|\mathfrak{c}_k\rangle\}_{k=0}^P$ satisfies the QEC conditions.

To solve this problem, it will be convenient to formally extend the Hilbert space $\mathcal{H}_S\otimes\mathcal{H}_A$ by tensoring it with a $(P+1)$-dimensional reference space $\mathcal{H}_L = \text{span}\{|0\rangle_L,\ldots,|P\rangle_L\}$ (see Figure 3.4(c)). This reference space will be representing the effective evolution of the probe state that happens within the code space and it will allow us to encode QEC conditions in a compact and numerically friendly way.

First, we introduce a matrix $Q$ as a Hermitian operator in $\mathscr{L}(\mathcal{H}_L\otimes\mathcal{H}_S)$ that represents

a code

$$Q = \mathrm{Tr}_A \left( \begin{pmatrix} |\mathfrak{c}_0\rangle \\ \vdots \\ |\mathfrak{c}_P\rangle \end{pmatrix} \begin{pmatrix} \langle \mathfrak{c}_0| & \cdots & \langle \mathfrak{c}_P| \end{pmatrix} \right). \tag{3.71}$$

This matrix is proportional to the reduced density matrix of the maximum entangled state between $\mathcal{H}_L$ and $\mathcal{H}_C$. By its construction $Q \geq 0$ and contains all relevant information on the code states in $\mathcal{H}_C$.

Next, we introduce effective generators $H_{L,i}$ acting on $\mathcal{H}_L$ so that they represent properly the action of the physical generators on the code space $(H_{L,i})_{kl} = \langle \mathfrak{c}_k | H_i \otimes \mathbb{1}_{\dim(\mathcal{H}_A)} | \mathfrak{c}_l \rangle$. The effective evolution generators are related with the $Q$ matrix via:

$$(H_{L,i})^T = \mathrm{Tr}_S \left( Q(\mathbb{1}_{P+1} \otimes H_i) \right) \quad i = 1, \ldots, P. \tag{3.72}$$

Note that the identity operator here acts on the reference space $\mathcal{H}_L$, and *not* on the ancillary space $\mathcal{H}_A$. Taking into account the orthonormality of $|\mathfrak{c}_k\rangle$ and the QEC condition Eq. (3.45), we obtain the following constraints on $Q$

$$\mathrm{Tr}_S(Q) = \mathbb{1}_{P+1}, \ \forall_{S_i \in \mathcal{S}} \ \mathrm{Tr}_S \left( Q(\mathbb{1}_{P+1} \otimes S_i) \right) \propto \mathbb{1}_{P+1}. \tag{3.73}$$

Let $\mathbb{1}_d/\sqrt{d}$, $\{H_i\}_{i=1}^{P}$, $\{S_i\}_{i=P+1}^{P'}$, $\{R_i\}_{i=P'+1}^{d^2-1}$ form an orthonormal basis of Hermitian operators in $\mathscr{L}(\mathcal{H}_S)$ such that $\mathcal{S} = \mathrm{span}_{\mathbb{H}}\{\mathbb{1}_d, (S_i)_{i=P+1}^{P'}\}$. Any non-negative $Q$ satisfying Eqs. (3.72)-(3.73) has the following form:

$$Q = \mathbb{1}_{P+1} \otimes \frac{\mathbb{1}_d}{d} + \sum_{i=1}^{P} (H_{L,i})^T \otimes H_i + \sum_{i=P+1}^{P'} \nu_i \mathbb{1}_{P+1} \otimes S_i + \sum_{i=P'+1}^{d^2-1} B_i \otimes R_i \geq 0, \tag{3.74}$$

where $\nu_i \in \mathbb{R}$ and $B_i$ are Hermitian. Conversely, for any nonnegative $Q \geq 0$, we can consider its purification $|Q\rangle \in \mathcal{H}_L \otimes \mathcal{H}_S \otimes \mathcal{H}_A$, which when written as $|Q\rangle = \sum_{k=0}^{P} |k\rangle_L \otimes |\mathfrak{c}_k\rangle_{SA}$ yields the code states $|\mathfrak{c}_k\rangle$. Note that it implies that the rank of $Q$ corresponding to the dimension of the ancillary space. It is always sufficient to assume the dimension of the ancillary space to be $\dim \mathcal{H}_A = (P+1)d$. Therefore $\{H_{L,i}\}$ is an achievable set of effective generators

(satisfying the QEC condition Eq. (3.45)) if and only if there exist such $\nu_i \in \mathbb{R}$ and $B_i$, for which $Q \geq 0$.

Finally, in order to have an explicit dependence of the MSE on the total time parameter $T$, we introduce a matrix $\Gamma = \frac{1}{2T}\mathcal{Y}$, i.e. $\Gamma_{ij} = \mathrm{Im}[\langle \mathfrak{c}_i | H_j \otimes \mathbb{1}_{\dim(\mathcal{H}_A)} | \mathfrak{c}_0 \rangle] = \mathrm{Im}[(H_{L,j})_{i0}]$, and we end up with:

$$\begin{aligned} &\frac{1}{4T^2} \min_{H_{L,i}, B_i, \nu_i} \mathrm{Tr}\left(W(\Gamma^T\Gamma)^{-1}\right), \quad \text{where } \Gamma_{ij} = \mathrm{Im}[H_{L,j}]_{i0}, \\ &\text{subject to } \mathbb{1}_{P+1} \otimes \frac{\mathbb{1}_d}{d} + \sum_{i=1}^{P} (H_{L,i})^T \otimes H_i + \sum_{i=P+1}^{P'} \nu_i \mathbb{1}_{P+1} \otimes S_i + \sum_{i=P'+1}^{d^2-1} B_i \otimes R_i \geq 0. \end{aligned} \tag{3.75}$$

We also remark that our way of formulating the Knill–Laflamme conditions as a positive semidefinite constraint is novel and may have applications beyond error-corrected quantum metrology.

### Reduction to a SDP

In order to reformulate Eq. (3.75) as a SDP, we first show that we may assume without loss of generality that $\Gamma\sqrt{W^{-1}} \geq 0$. Note that for any full rank matrix $\Gamma$, the polar decomposition theorem implies that there always exists an orthonormal matrix $O$ such that $O\Gamma\sqrt{W^{-1}} \geq 0$. Next, as $\Gamma_{ij} = \mathrm{Im}[\langle i | H_{L,j} | 0 \rangle]$, multiplication $\Gamma$ by $O$ is equivalent to rotating the base in the reference space $\mathcal{H}_L$. Since, according to Eq. (3.71) such a rotation cannot change the non-negativity of $Q$ and at the same time it does not affect the figure of merit $\mathrm{Tr}\left(W(\Gamma^T\Gamma)^{-1}\right)$, the statement is proven. To put Eq. (3.75) in the form of a SDP, we introduce a positive matrix $K \in \mathbb{R}^{P\times P}$ and a positive real number $w$. Now, using the following two relations,

$$\begin{pmatrix} K & \mathbb{1}_P \\ \mathbb{1}_P & \Gamma\sqrt{W^{-1}} \end{pmatrix} \geq 0 \quad \Leftrightarrow \quad K \geq (\Gamma\sqrt{W^{-1}})^{-1}, \tag{3.76}$$

$$\begin{pmatrix} w\mathbb{1}_P & K \\ K & \mathbb{1}_P \end{pmatrix} \geq 0 \quad \Leftrightarrow \quad w\mathbb{1}_P \geq K^2, \tag{3.77}$$

we see that $P \min w = \min \mathrm{Tr}(K^2) = \min \mathrm{Tr}\left(W(\Gamma^T\Gamma)^{-1}\right)$ in Eq. (3.61), making it equivalent to Eq. (3.75). Hence the problem takes the form of a SDP.

### JNT-QEC vs. SEP-QEC

It should be remarked that JNT-QEC do not contain SEP-QEC as a subclass. In SEP-QEC, unlike in JNT-QEC, the noises are not fully corrected in the entire space, and the decoherence is only avoided by choosing a properly mixed state input. However, in the noiseless cases, JNT-QEC contains SEP-QEC and is indeed always optimal. In general, one could combine both these approaches in a unified framework by dividing the set of all parameters into smaller subsets and then applying JNT-QEC for each of these subset separately—in this approach SEP-QEC case would correspond to the situation where JNT-QEC optimization is applied to single parameter subsets. Such an optimization is in principle doable, but will involve much large numerical effort and it is not clear that it will lead to better protocols.

The advantage of SEP-QEC over JNT-QEC can be revealed through examples [Górecki et al., 2020]. In SEP-QEC, the input probe state $\rho_{\mathrm{in}} = \frac{1}{P}\sum_{i=1}^{P} |\psi_i\rangle\langle\psi_i|$ and each parameter is estimated separately. As a consequence, we effectively measure each parameter only once in every $P$ repetitions of an experiment (corresponding to the $1/P$ factor in the $\rho_{\mathrm{in}}$). Therefore for a fixed total number of measurements, the uncertainty of estimating a given parameter will grow proportionally to $P$. Intuitively, $\Delta^2_W\hat{\boldsymbol{\theta}}_{\mathrm{SEP}}$ will scale as $\Theta(P^2)$ in normal circumstances. On the other hand, in JNT-QEC, it is possible to estimate all parameters jointly. In fact, the largest possible advantage offered by JNT-QEC is $\Theta(P)$, achievable in a noiseless system example where $\Delta^2_W\hat{\boldsymbol{\theta}}_{\mathrm{SEP}} = \Theta(P^2)$ and $\Delta^2_W\hat{\boldsymbol{\theta}}_{\mathrm{JNT}} = \Theta(P)$. In noisy case, there is another example involving $SU(d)$ estimation where $\Delta^2_W\hat{\boldsymbol{\theta}}_{\mathrm{SEP}} = \Theta(P^2)$ and $\Delta^2_W\hat{\boldsymbol{\theta}}_{\mathrm{JNT}} = \Theta(P^{3/2})$.

It is also worth noting that, apart from the improved metrological performance provided by QEC protocols when dealing with noisy systems, the above algorithm is also applicable in the noiseless scenario when $\mathcal{S} = \mathrm{span}_{\mathbb{H}}\{\mathbb{1}_d\}$. In such a situation *no QEC is required* (for simplicity we may still use $\mathcal{H}_C$ for $\mathrm{span}\{|\mathfrak{c}_k\rangle\}_{k=0}^{P}$, but no recovery operation or projection $\Pi_{\mathfrak{c}}$ is needed during evolution), but the condition $\boldsymbol{\theta} = [0,\ldots,0]$ (which is achievable by

applying inverse Hamiltonian dynamics [Yuan, 2016]) is still required, as otherwise the derivatives of the state may not scale linearly with $T$. In such situations, the solution of JNT-QEC yields an optimally ancilla-assisted sensing protcol under arbitrary system dynamics (Hamiltonitans) that resolves the potential incompatibility issues between sensing of different parameters. It should be stressed that our approach is universal and unlike existing approaches [Imai and Fujiwara, 2007; Yuan, 2016; Kura and Ueda, 2018] does not assume any specific structures of the Hamiltonians.

## 3.5 Optimal QEC protocol: HLS

In this section, we present the optimal QEC protocol when the HNLS condition is violated, i.e. $H \in \mathcal{S}$. Instead of achieving the HL, here our goal is to find the optimal QEC protocol such that the QFI is maximized asymptotically. We prove that the optimal SQL coefficient using sequential strategies is achievable using QEC by showing that the upper bound obtained in Section 3.2.2 is attainable.

Recall from Section 3.2.2 that, when $H \in \mathcal{S}$, there exists Hermitian matrices $h^{(0)}$, $h^{(1)}$ and $h^{(2)} \in \mathbb{C}^{(r+1)\times(r+1)}$ satisfying $h^{(0)}_{0k} = 0$, $0 \leq k \leq r$ and $h^{(1)}_{00} = 0$ such that

$$\beta^{(2)} = H - \sum_{k,j=1}^{r} h^{(0)}_{jk} L^\dagger_k L_j - \sum_{k=1}^{r} (h^{(1)}_{0k} L_k + h^{(1)}_{k0} L^\dagger_k) - h^{(2)}_{00} \mathbb{1} = 0. \tag{3.78}$$

Then the QFI has at most a linear scaling with respect to $T$: $F(\rho(T)) \leq 4T\|\alpha^{(2)}\|$, where

$$\begin{aligned} \alpha^{(2)} &= (h^{(1)}\mathbf{K}^{(0)} + h^{(0)}\mathbf{K}^{(1)})^\dagger (h^{(1)}\mathbf{K}^{(0)} + h^{(0)}\mathbf{K}^{(1)}) \\ &= \sum_{j=1}^{r} \left| h^{(1)}_{0j} \right|^2 + \sum_{k,k'=1}^{r} h^{(0)*}_{jk} h^{(0)}_{jk'} L^\dagger_k L_{k'}. \end{aligned} \tag{3.79}$$

To simplify the notations, let $\alpha = \alpha^{(2)}$, $\beta = \beta^{(2)}$, $g = h_{00}^{(2)} \in \mathbb{R}$,

$$\mathbf{g} = \begin{pmatrix} h_{11}^{(1)} \\ \vdots \\ h_{1r}^{(1)} \end{pmatrix} \in \mathbb{C}^r, \quad \text{and } \mathfrak{g} = \begin{pmatrix} h_{11}^{(1)} & h_{12}^{(1)} & \cdots & h_{1r}^{(1)} \\ h_{21}^{(1)} & h_{22}^{(1)} & \cdots & h_{2r}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{r1}^{(1)} & \cdots & \cdots & h_{rr}^{(1)} \end{pmatrix} \in \mathbb{C}^{r\times r}. \tag{3.80}$$

Then we have

$$F(\rho(T)) \leq 4T \min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\|, \tag{3.81}$$

where $\|\cdot\|$ is the operator norm of a matrix, $g \in \mathbb{R}$, $\mathbf{g} \in \mathbb{C}^r$, $\mathfrak{g} \in \mathbb{C}^{r\times r}$ is hermitian,

$$\alpha = (\mathbf{g}\mathbb{1} + \mathfrak{g}\mathbf{L})^\dagger(\mathbf{g}\mathbb{1} + \mathfrak{g}\mathbf{L}), \tag{3.82}$$

$$\beta = H + g\mathbb{1} + \mathbf{g}^\dagger\mathbf{L} + \mathbf{L}^\dagger\mathbf{g} + \mathbf{L}^\dagger\mathfrak{g}\mathbf{L}, \tag{3.83}$$

where

$$\mathbf{L} := \begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_r \end{pmatrix}. \tag{3.84}$$

Again, $\mathbf{g}, \mathfrak{g}$ are shorthand for $\mathbf{g} \otimes \mathbb{1}$, $\mathfrak{g} \otimes \mathbb{1}$ and we omit the "$\otimes\mathbb{1}$" for simplicity.

We will introduce an approximate QEC strategy which (asymptotically) saturates the QFI upper bound up to an arbitrarily small error under arbitrary Markovian noise, that is [Zhou and Jiang, 2020b],

**Theorem 3.4** (Optimal SQL)**.** *Consider a finite-dimensional probe with Hamiltonian $H(\theta) = \theta H$, subject to Markovian noise described by a Lindblad master equation with Lindblad operators $\{L_k\}$. Then*

$$\sup_{\substack{\text{sequential} \\ \text{strategies}}} \lim_{T\to\infty} \frac{F(\rho(T))}{T} = 4 \min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\|. \tag{3.85}$$

*In particular, for any small $\eta > 0$, there exists an approximate QEC strategy such that*

$$\lim_{T\to 0} \frac{F(\rho(T))}{T} > 4 \min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\| - \eta. \tag{3.86}$$

The SQL coefficient

$$F_{\text{SQL}} := \lim_{T\to 0} \frac{F(\rho(T))}{T} \tag{3.87}$$

will be the objective function we maximize. Note that

$$\lim_{T\to 0} \frac{F(\rho(T))}{T} = \sup_{T>0} \frac{F(\rho(T))}{T} \tag{3.88}$$

because for any $T_0$ such that $F(T_0)/T_0 \approx \sup_{T>0} F(\rho(T))/T$, we can always find a sensing strategy such that $T = kT_0$ and $F(kT_0)/(kT_0) = F(T_0)/T_0$ for all integers $k$ by measuring and renewing the probing state every constant time $T_0$. Then $\lim_{T\to\infty} F(\rho(T))/T \geq \lim_{k\to\infty} F(kT_0)/(kT_0) \approx \sup_{T>0} F(\rho(T))/T$.

Intuitively, the QEC protocol when the HNLS is violated must be fundamentally different from the ones in the previous sections (Section 3.2.3 and Section 3.4), because if errors in the Lindblad span are fully corrected as before, then the signal will be fully corrected as well. Therefore, we must use approximate QEC strategies in which case the noise is only partially corrected while the signal is also partially preserved and the optimization is not only over the encoding but also the recovery as well. Normally, only suboptimal recovery channels are available in approximate QEC [Barnum and Knill, 2002; Fletcher et al., 2007; Bény and Oreshkov, 2010; Ng and Mandayam, 2010; Tyson, 2010; Albert et al., 2018]. In our setting, however, taking the advantage of noiseless ancillas, the exact solution exist. Below, we will first review the QFI upper bound in the SQL case, as derived in Section 3.2.2 using the channel-extension method, propose an approximate QEC protocol which achieves this bound asymptotically and then provide an efficient algorithm to solve it.

### 3.5.1 Attaining the upper bound

#### Approximate QEC

Let $\Pi_{\mathfrak{c}} = |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| + |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|$ is the projection on to the code space $\mathcal{H}_C$, where $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$ are the logical zero and one states. Applying the approximate QEC quantum operation $\mathcal{P}+\mathcal{R}\circ\mathcal{P}_\perp$ infinitely fast, the effective evolution would be (up to the first order of $dt$ [Zhou et al., 2018; Layden et al., 2019])

$$\frac{d\rho}{dt} = -i[\theta\mathcal{P}(H),\rho] + \sum_{i=1}^{r}\Big(\mathcal{P}(L_i\rho L_i^\dagger) + \mathcal{R}(\mathcal{P}_\perp(L_i\rho L_i^\dagger)) - \frac{1}{2}\{\mathcal{P}(L_i^\dagger L_i),\rho\}\Big), \tag{3.89}$$

where $\Pi_{\mathfrak{c}}^\perp = 1-\Pi_{\mathfrak{c}}$, $\mathcal{P}(\cdot) = \Pi_{\mathfrak{c}}(\cdot)\Pi_{\mathfrak{c}}$, $\mathcal{P}_\perp(\cdot) = \Pi_{\mathfrak{c}}^\perp(\cdot)\Pi_{\mathfrak{c}}^\perp$ and $\mathcal{R}$ is a CPTP map describing the approximate QEC recovery channel. We define the following class of approximate QEC codes

$$|\mathfrak{c}_0/\mathfrak{c}_1\rangle = \sum_{ij} A_{0/1,ij}\,|i\rangle_S\,|j,0/1\rangle_A\,, \tag{3.90}$$

where $A_0, A_1 \in \mathbb{C}^{d\times d}$, $A_{0,ij} = \sqrt{1-\varepsilon^2}C_{ij} + \varepsilon D_{ij}$ and $A_{1,ij} = \sqrt{1-\varepsilon^2}C_{ij} - \varepsilon D_{ij}$ satisfy $\mathrm{Tr}(A_0A_0^\dagger) = \mathrm{Tr}(A_1A_1^\dagger) = 1$ and $\mathrm{Tr}(C^\dagger D) = 0$. Here $C$ describes the part of the code which $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$ have in common and $D$ describes the part distinguishing $|\mathfrak{c}_0\rangle$ from $|\mathfrak{c}_1\rangle$ which generates non-zero signal and noise. In the special case where $\varepsilon = 0$, the effective signal and noise are zero. Let $\mathcal{H}_A = \mathcal{H}_{A'}\otimes\mathcal{H}_2$ where $\dim\mathcal{H}_{A'} = d$ and $\dim\mathcal{H}_2 = 2$, the last ancillary qubit in $\mathcal{H}_2$ makes the signal and noises both diagonal in the code space, i.e. $\langle\mathfrak{c}_0|H|\mathfrak{c}_1\rangle = \langle\mathfrak{c}_0|S|\mathfrak{c}_1\rangle = 0$ for all $S\in\mathcal{S}$. Later on, we will assume $\varepsilon$ is a small parameter and consider the perturbation expansion of the effective dynamics around $\varepsilon = 0$. We consider the recovery channel restricted to the structure (we will show that this type of recovery channels is sufficient for our purpose)

$$\mathcal{R}(\cdot) = \sum_m \left(|\mathfrak{c}_0\rangle\langle R_m,0| + |\mathfrak{c}_1\rangle\langle Q_m,1|\right)(\cdot)\left(|R_m,0\rangle\langle\mathfrak{c}_0| + |Q_m,1\rangle\langle\mathfrak{c}_1|\right), \tag{3.91}$$

where $\{|R_m\rangle\},\{|Q_m\rangle\} \subseteq \mathcal{H}_S\otimes\mathcal{H}_{A'}$ are two sets of orthonormal basis and $\mathcal{R}$ is CPTP. A few lines of calculation shows the effective channel (Eq. (3.89)) under the approximate QEC

code (Eq. (3.90)) and the recovery channel (Eq. (3.91)) is

$$\frac{d\rho}{dt} = -i\left[\frac{\theta \mathrm{Tr}(H\sigma_{z,\mathfrak{c}})}{2}\sigma_{z,\mathfrak{c}} + H_{\mathrm{shift}}, \rho\right] + \frac{\gamma(\mathcal{R})}{2}\left(\sigma_{z,\mathfrak{c}}\rho\sigma_{z,\mathfrak{c}} - \rho\right), \tag{3.92}$$

where $\sigma_{z,\mathfrak{c}} = |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| - |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|$, $H_{\mathrm{shift}}$ is independent of $\theta$, and

$$\gamma(\mathcal{R}) = -\mathrm{Re}\Big[\sum_{i=1}^{r}\langle\mathfrak{c}_0|\Big(\mathcal{R}(\mathcal{P}_\perp(L_i|\mathfrak{c}_0\rangle\langle\mathfrak{c}_1|L_i^\dagger)) + \mathcal{P}(L_i|\mathfrak{c}_0\rangle\langle\mathfrak{c}_1|L_i^\dagger) - \frac{1}{2}\{\mathcal{P}(L_i^\dagger L_i), |\mathfrak{c}_0\rangle\langle\mathfrak{c}_1|\}\Big)|\mathfrak{c}_1\rangle\Big]. \tag{3.93}$$

We can remove the term $H_{\mathrm{shift}}$ in Eq. (3.92) by applying a reverse Hamiltonian constantly [Sekatski et al., 2017]. For dephasing channels, the optimal $F_{\mathrm{SQL}}$ is reached using a special type of spin-squeezed state as the input [Kitagawa and Ueda, 1993; Huelga et al., 1997; Ulam-Orgikh and Kitagawa, 2001; Escher et al., 2011; Demkowicz-Dobrzański and Maccone, 2014], where we have

$$F_{\mathrm{SQL}} = \frac{\mathrm{Tr}(H\sigma_{z,\mathfrak{c}})^2}{2\gamma(\mathcal{R})}. \tag{3.94}$$

To simulate the evolution of multipartite spin-squeezed states using the sequential strategy where we have only a single probe, one could first prepare the desired spin-squeezed state in $\bigotimes_{i=1}^{N}\mathcal{H}_i$ by entangling the logical qubit in the effective dephasing channel ($\mathcal{H}_1 = \mathcal{H}_S \otimes \mathcal{H}_A$) with a large number of ancillas ($\bigotimes_{i=2}^{N}\mathcal{H}_i$) where $\dim\mathcal{H}_i = \dim\mathcal{H}_1$ for $2 \le i \le N$, and then perform swap operations between $\mathcal{H}_1$ and $\mathcal{H}_i$ for $i = 2, \ldots, N$ successively every time $T/N$. The optimal $F_{\mathrm{SQL}}$ in Eq. (3.94) is asymptotically attainable at $N \to \infty$ [Ulam-Orgikh and Kitagawa, 2001]. On the other hand, if we used a single logical qubit state $|\mathfrak{c}_+\rangle = \frac{|\mathfrak{c}_0\rangle+|\mathfrak{c}_1\rangle}{\sqrt{2}}$ as the input, the SQL coefficient will be reduced by a factor of $e$, in which case one can still achieve $F_{\mathrm{SQL}} \approx \frac{4}{e}\min_{g,\mathbf{g},\mathfrak{g}|\beta=0}\|\alpha\|$ for arbitrary Markovian noise.

For simplicity in furture calculations, we perform a two-step gauge transformation on the Lindblad operators $\{L_i\}_{i=1}^r$ to simplify the dynamics: (1) Let $L_i \leftarrow L_i - \mathrm{Tr}(C^\dagger L_i C)\cdot\mathbb{1}$, such that $L_i$ satisfies $\mathrm{Tr}(C^\dagger L_i C) = 0$ for all $L_i$. (2) Perform a unitary transformation $\mathbf{L} \leftarrow u\mathbf{L}$ ($u \in \mathbb{C}^{r\times r}$) such that $\mathrm{Tr}(C^\dagger L_i^\dagger L_j C)$ is a diagonal matrix. The transformations

above only induce another parameter-independent shift $H_{\mathrm{s}}$ in the Hamiltonian which could be eliminated by applying a reverse Hamiltonian. Now we have a new set of Lindblad operators $\{J_i\}_{i=1}^r$, satisfying

$$\mathrm{Tr}(C^\dagger J_i C) = 0, \quad \mathrm{Tr}(C^\dagger J_i^\dagger J_j C) = \lambda_i \delta_{ij}, \tag{3.95}$$

and we replace $\{L_i\}_{i=1}^r$ with $\{J_i\}_{i=1}^r$ in Eq. (3.93).

**Recovery optimization**

First, we maximize $F_{\mathrm{SQL}}$ over the recovery $\mathcal{R}$, which is equivalent to minimizing $\gamma(\mathcal{R})$ over $\mathcal{R}$. According to Eq. (3.93),

$$\begin{aligned}\gamma(\mathcal{R}) = -\mathrm{Re}\Big[\sum_{i=1}^r \langle \mathfrak{c}_0| \Big(\mathcal{R}(\mathcal{P}_\perp(J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger)) + \mathcal{P}(J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger) \\ - \frac{1}{2}\{\mathcal{P}(J_i^\dagger J_i), |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1|\}\Big) |\mathfrak{c}_1\rangle \Big].\end{aligned} \tag{3.96}$$

In order to calculate $\gamma = \min_{\mathcal{R}} \gamma(\mathcal{R})$, we only need to calculate the first term minimized over $\mathcal{R}$:

$$\begin{aligned}&- \max_{\mathcal{R}} \mathrm{Re}\Big[\sum_i \langle \mathfrak{c}_0| \mathcal{R}(\mathcal{P}_\perp(J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger)) |\mathfrak{c}_1\rangle\Big] \\ =& - \max_{|R_m\rangle, |Q_m\rangle} \mathrm{Re}\Big[\sum_{i,m} \langle R_m, 0| \Pi_{\mathfrak{c}\perp} J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger \Pi_{\mathfrak{c}\perp} |Q_m, 1\rangle\Big] \\ =& -\frac{1}{2} \max_{|R_m\rangle, |Q_m\rangle} \mathrm{Tr}\Big(\sum_m |R_m\rangle \langle Q_m| \cdot \sum_i \langle 0| \Pi_{\mathfrak{c}\perp} J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger \Pi_{\mathfrak{c}\perp} |1\rangle + h.c.\Big) \\ =& -\Big\| \sum_i \langle 0| \Pi_{\mathfrak{c}\perp} J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger \Pi_{\mathfrak{c}\perp} |1\rangle \Big\|_1 = -\Big\| \sum_i \Pi_{\mathfrak{c}\perp} J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger \Pi_{\mathfrak{c}\perp} \Big\|_1,\end{aligned} \tag{3.97}$$

where $h.c.$ denotes Hermitian conjugate and we have used $\max_{U:U^\dagger U=\mathbb{1}} \mathrm{Tr}(MU + M^\dagger U^\dagger) = 2\|M\|_1$ for arbitrary square matrices $M$ and $U$, which could be proven easily using the singular value decomposition of $M$.

Then the minimum noise rate $\gamma = \min_{\mathcal{R}} \gamma(\mathcal{R})$ is

$$\gamma = -\Big\| \sum_{i=1}^{r} \mathcal{P}_{\perp}(J_i \left|\mathfrak{c}_0\right\rangle \left\langle \mathfrak{c}_1\right| J_i^{\dagger}) \Big\|_1 - \\ \mathrm{Re}\Big[ \sum_{i=1}^{r} \left\langle \mathfrak{c}_0\right| \Big( \mathcal{P}(J_i \left|\mathfrak{c}_0\right\rangle \left\langle \mathfrak{c}_1\right| J_i^{\dagger}) - \frac{1}{2}\{\mathcal{P}(J_i^{\dagger} J_i), \left|\mathfrak{c}_0\right\rangle \left\langle \mathfrak{c}_1\right|\} \Big) \left|\mathfrak{c}_1\right\rangle \Big], \quad (3.98)$$

**Perturbative expansion**

Next, we would like to maximize $F_{\mathrm{SQL}}$ (Eq. (3.94)) over all possible approximate QEC codes (Eq. (3.90)), which is mathematcially difficult because of the trace norm in the denominator. To eliminate the trace norm, we further sacrifice the generality of our approximate QEC code and assume $\varepsilon \ll 1$. We call it the "perturbation" code in the sense that the signal and the noise are both infinitesimally small when $\varepsilon \to 0$. Under the limit $\varepsilon \to 0$, we have $\mathrm{Tr}(H\sigma_{z,\mathfrak{c}}) = 2\varepsilon \mathrm{Tr}(H\tilde{C}) + O(\varepsilon^2)$, where

$$\tilde{C} = CD^{\dagger} + DC^{\dagger}, \quad (3.99)$$

and ignoring all $o(\varepsilon^2)$ terms (where $f(\varepsilon) = o(\varepsilon^2)$ means $\lim_{\varepsilon \to 0} f(\varepsilon)/\varepsilon^2 = 0$), the noise rate is

$$\gamma = \varepsilon^2 \Big( \sum_i 2 \big|\mathrm{Tr}(J_i \tilde{C})\big|^2 + \sum_{ij:\lambda_i + \lambda_j \neq 0} \frac{|\mathrm{Tr}(J_i^{\dagger} J_j \tilde{C})|^2}{(\lambda_i + \lambda_j)} \Big). \quad (3.100)$$

The detailed derivations are contained in Appendix A.

Then we have the following expression of the SQL coefficient (up to the lowest order of $\varepsilon$)

$$F_{\mathrm{SQL}}(C, \tilde{C}) \approx \frac{\mathrm{Tr}(H\tilde{C})^2}{\sum_i \big|\mathrm{Tr}(J_i \tilde{C})\big|^2 + \sum_{ij:\lambda_i + \lambda_j \neq 0} \frac{|\mathrm{Tr}(J_i^{\dagger} J_j \tilde{C})|^2}{2(\lambda_i + \lambda_j)}}, \quad (3.101)$$

as a function of $\tilde{C}$ and $C$ (implicitly through the choice of $\{J_i\}_{i=1}^r$). The effective dynamics of the perturbation code has the feature that both the signal and the noises are equally weak and only the ratio between them matters. Therefore the exact value of $\varepsilon$ will not influence the SQL coefficient $F_{\mathrm{SQL}}$ as long as it is sufficiently small. On the other hand, it does influence how fast $F(\rho(T))/T$ reaches its optimum $F_{\mathrm{SQL}}$, characterized by a coherence time $O(1/\varepsilon^2)$.

### Code optimization

Now we maximize the SQL coefficient (up to the lowest order of $\varepsilon$) over $C$ and $\tilde{C}$ and show that the optimal $F_{\text{SQL}}$ is exactly equal to its upper bound in Eq. (3.81). The domain of $C$ is all complex matrices satisfying $\text{Tr}(C^\dagger C) = 1$. We assume the domain of $\tilde{C}$ is all traceless Hermitian matrices satisfying $\text{Tr}(J_i^\dagger J_j \tilde{C}) = 0$ for all $i, j \in \mathfrak{n} := \{i | \lambda_i = 0\}$. When $C$ is full-rank, $\mathfrak{n}$ is empty and for arbitrary traceless $\tilde{C}$, we could always take $D^\dagger = \frac{1}{2} C^{-1} \tilde{C}$ such that Eq. (3.99) is satisfied. When $C$ is singular, we could replace it with an approximate full-rank version (e.g. $C \leftarrow C + \delta \mathbb{1}$). In this case, $F_{\text{SQL}}$ will only be decreased by an infinitesimal small amount when $\varepsilon = o(\delta^2)$ because the numerator in Eq. (3.101) is only slightly perturbed after the replacement.

Consider the following optimization problem over $g, \mathbf{g}, \mathfrak{g}$ and $C$,

$$\begin{aligned} &\max_{C} \min_{g,\mathbf{g},\mathfrak{g}} \quad 4\text{Tr}(C^\dagger \alpha C), \\ &\text{subject to} \quad \beta = 0, \quad \text{Tr}(C^\dagger C) = 1, \end{aligned} \tag{3.102}$$

Fixing $C$, we introduce a Hermitian matrix $\tilde{C}$ as the Lagrange multiplier associated with the constraint $\beta = 0$ [Boyd and Vandenberghe, 2004]. Strong duality implies Eq. (3.102) has the same solution as its dual program, which we claim is

$$\begin{aligned} \max_{C,\tilde{C}} \; F_{\text{SQL}}(C, \tilde{C}), \;\; &\text{subject to } \text{Tr}(C^\dagger C) = 1, \; \text{Tr}(\tilde{C}) = 0, \\ &\text{and } \text{Tr}(J_i^\dagger J_j \tilde{C}) = 0, \forall i, j \in \mathfrak{n}. \end{aligned} \tag{3.103}$$

whose optimal value could be achieved using the perturbation code up to an infinitesimal small error according to the discussion above.

Now we show the Lagrange dual program of Eq. (3.102) is indeed Eq. (3.103). From the definition of $\alpha$ (Eq. (3.82)) and $\beta$ (Eq. (3.83)), we see that the upper bound in Eq. (3.81) is invariant under the transformation $\mathbf{L} \to \mathbf{J}$, that is, after the transformation $\mathbf{L} \to \mathbf{J}$ there is always another set of $(g, \mathbf{g}, \mathfrak{g})$ such that $\beta = 0$ and $\alpha$ is the same. Therefore we let

$$\alpha = (\mathbf{g}\mathbb{1} + \mathfrak{g}\mathbf{J})^\dagger (\mathbf{g}\mathbb{1} + \mathfrak{g}\mathbf{J}), \tag{3.104}$$

$$\beta = H + g\mathbb{1} + \mathbf{g}^\dagger \mathbf{J} + \mathbf{J}^\dagger \mathbf{g} + \mathbf{J}^\dagger \mathfrak{g} \mathbf{J}, \tag{3.105}$$

where $\mathbf{J} = (J_1, J_2, \ldots, J_r)^T$. To proceed, we simplify the notations by letting

$$\boldsymbol{\jmath}_i = \frac{\mathrm{Tr}(J_i \tilde{C})}{\mathrm{Tr}(H\tilde{C})}, \quad \mathfrak{j}_{ij} = \frac{\mathrm{Tr}(J_i^\dagger J_j \tilde{C})}{\mathrm{Tr}(H\tilde{C})}. \tag{3.106}$$

Note that the $r$-dimensional vector $\boldsymbol{\jmath}$ is to be distinguished from the index $j$, then we have

$$F_{\mathrm{SQL}}(C, \tilde{C}) = \left( \boldsymbol{\jmath}^\dagger \boldsymbol{\jmath} + \sum_{ij:\lambda_i+\lambda_j\neq 0} \frac{|\mathfrak{j}_{ij}|^2}{2(\lambda_i+\lambda_j)} \right)^{-1}, \tag{3.107}$$

and $4\mathrm{Tr}(C^\dagger \alpha C) = 4(\mathbf{g}^\dagger \mathbf{g} + \mathrm{Tr}(\Lambda \mathfrak{g}^2))$.

Fixing $C$, we introduce a Hermitian matrix $\tilde{C}$ as a Lagrange multiplier of $\beta = 0$ [Boyd and Vandenberghe, 2004], the Lagrange function is

$$L(\tilde{C}, g, \mathbf{g}, \mathfrak{g}) = 4(\mathbf{g}^\dagger \mathbf{g} + \mathrm{Tr}(\Lambda \mathfrak{g}^2)) + \mathrm{Tr}(\tilde{C}(H + g\mathbb{1} + \mathbf{J}^\dagger \mathbf{g} + \mathbf{g}^\dagger \mathbf{J} + \mathbf{J}^\dagger \mathfrak{g} \mathbf{J})). \tag{3.108}$$

Then the dual program of Eq. (3.102) is

$$\begin{aligned}
&\max_{\tilde{C}} \min_{g,\mathbf{g},\mathfrak{g}} L(\tilde{C}, g, \mathbf{g}, \mathfrak{g}) \\
=& \max_{\tilde{C}} \min_{g,\mathbf{g},\mathfrak{g}} 4(\mathbf{g}^\dagger \mathbf{g} + \mathrm{Tr}(\Lambda \mathfrak{g}^2)) + \mathrm{Tr}(\tilde{C}(H + g\mathbb{1} + \mathbf{J}^\dagger \mathbf{g} + \mathbf{g}^\dagger \mathbf{J} + \mathbf{J}^\dagger \mathfrak{g} \mathbf{J})) \\
=& \max_{\substack{\tilde{C}:\mathrm{Tr}(\tilde{C})=0, \\ \mathrm{Tr}(\tilde{C}H)\neq 0}} \min_{\mathbf{g},\mathfrak{g}} 4(\mathbf{g}^\dagger \mathbf{g} + \mathrm{Tr}(\Lambda \mathfrak{g}^2)) + \mathrm{Tr}(\tilde{C}H)(1 + \mathbf{g}^\dagger \boldsymbol{\jmath} + \boldsymbol{\jmath}^\dagger \mathbf{g} + \mathrm{Tr}(\mathfrak{g}^T \mathfrak{j})) \\
=& \max_{\substack{\tilde{C}:\mathrm{Tr}(\tilde{C})=0, \\ \forall_{i,j\in\mathfrak{n}} \mathrm{Tr}(\tilde{C} J_i^\dagger J_j)=0, \\ \mathrm{Tr}(\tilde{C}H)\neq 0}} -\frac{1}{4}\mathrm{Tr}(\tilde{C}H)^2 \boldsymbol{\jmath}^\dagger \boldsymbol{\jmath} - \frac{1}{8}\mathrm{Tr}(\tilde{C}H)^2 \sum_{ij:\lambda_i+\lambda_j\neq 0} \frac{|\mathfrak{j}_{ij}|^2}{\lambda_i+\lambda_j} + \mathrm{Tr}(\tilde{C}H) \\
=& \max_{\substack{\tilde{C}:\mathrm{Tr}(\tilde{C})=0, \\ \forall_{i,j\in\mathfrak{n}} \mathrm{Tr}(\tilde{C} J_i^\dagger J_j)=0,}} \left( \boldsymbol{\jmath}^\dagger \boldsymbol{\jmath} + \sum_{ij:\lambda_i+\lambda_j\neq 0} \frac{|\mathfrak{j}_{ij}|^2}{2(\lambda_i+\lambda_j)} \right)^{-1} = \max_{\substack{\tilde{C}:\mathrm{Tr}(\tilde{C})=0, \\ \forall_{i,j\in\mathfrak{n}} \mathrm{Tr}(\tilde{C} J_i^\dagger J_j)=0}} F_{\mathrm{SQL}}(C, \tilde{C}),
\end{aligned} \tag{3.109}$$

as in Eq. (3.103).

On the other hand, thanks to Sion's minimax theorem [Komiya, 1988; do Rosário Grossinho and Tersian, 2001], we can exchange the order of the maximization and minimization

in Eq. (3.102). The minimax theorem [do Rosário Grossinho and Tersian, 2001] states that for convex compact sets $P \subseteq \mathbb{R}^m$ and $Q \subseteq \mathbb{R}^n$ and $f : P \times Q \to \mathbb{R}$ such that $f(x, y)$ is a continuous convex (concave) function in $x$ ($y$) for every fixed $y$ ($x$), then

$$\max_{y\in Q} \min_{x\in P} f(x, y) = \min_{x\in P} \max_{y\in Q} f(x, y). \tag{3.110}$$

In Eq. (3.102), the objective function $4\mathrm{Tr}(C^\dagger \alpha C)$ is concave (linear) with respect to $CC^\dagger$ and convex (quadratic) with respect to $(\mathbf{g}, \mathfrak{g})$. The operator $CC^\dagger$ satisfying $\mathrm{Tr}(CC^\dagger) = 1$ is contained in a convex compact set, but the domain of $(g, \mathbf{g}, \mathfrak{g})$ is not compact. In fact, we could always confine $(\mathbf{g}, \mathfrak{g})$ in a convex and compact set such that the solution of Eq. (3.102) is not altered [Zhou and Jiang, 2020b, Appx. D]. As a result, the minimax theorem is applicable and we can exchange the order of the maximization and minimization in Eq. (3.102).

### 3.5.2 Efficient numerical algorithm

It is known that the upper bound in Eq. (3.81) could be calculated via a SDP [Demkowicz-Dobrzański et al., 2017; Czajkowski et al., 2019]. Based on that, now we provide an efficient numerical algorithm obtaining an optimal $(C^\diamond, \tilde{C}^\diamond)$ in three steps. The algorithm runs as follows:

(1) Solving $\min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\|$ using the SDP gives us an optimal $\alpha^\diamond$ (and corresponding $g^\diamond, \mathbf{g}^\diamond, \mathfrak{g}^\diamond$) satisfying $\|\alpha^\diamond\| = \min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\|$.

(2) Suppose $\Pi^\diamond$ is the projection onto the subspace spanned by all eigenstates corresponding to the largest eigenvalue of $\alpha^\diamond$, we find an optimal $C^\diamond C^{\diamond\dagger}$ satisfying $\Pi^\diamond C^\diamond C^{\diamond\dagger} \Pi^\diamond = C^\diamond C^{\diamond\dagger}$ and

$$\mathrm{Re}[\mathrm{Tr}(C^\diamond C^{\diamond\dagger}(\Delta\mathbf{g}\mathbb{1} + \Delta\mathfrak{g}\mathbf{L})^\dagger(\mathbf{g}^\diamond\mathbb{1} + \mathfrak{g}^\diamond\mathbf{L}))] = 0, \tag{3.111}$$

for all $(\Delta\mathbf{g}, \Delta\mathfrak{g})$ such that $\Delta g\mathbb{1} + \Delta\mathbf{g}^\dagger\mathbf{L} + \mathbf{L}^\dagger\Delta\mathbf{g} + \mathbf{L}^\dagger\Delta\mathfrak{g}\mathbf{L} = 0$ for some $\Delta g$. Note that this step is simply solving a system of linear equations.

(3) Find $\{J_i\}_{i=1}^r$ via the gauge transformation. Let $\mathcal{S}_0 = \mathrm{span}\{\mathbb{1}, J_i^\dagger J_j, \forall i,j \in \mathfrak{n}\}$. Decompose $M = J_i$ or $J_{ij}$ $(:= J_i^\dagger J_j)$ into $M = M^{\mathbb{H}} + iM^{\mathbb{AH}} + M_0^{\mathbb{H}} + iM_0^{\mathbb{AH}}$ where $M^{\mathbb{H},\mathbb{AH}}$, $M_0^{\mathbb{H},\mathbb{AH}}$ are Hermitian, $M_0^{\mathbb{H},\mathbb{AH}} \in \mathcal{S}_0$ and $M^{\mathbb{H},\mathbb{AH}} \perp \mathcal{S}_0$ (in terms of the Hilbert-Schmidt norm). Using the vectorization of matrices $|\cdot\rangle\!\rangle = \sum_{jk} \langle j| (\cdot) |k\rangle |j\rangle |k\rangle$, let

$$B = \sum_i |J_i^{\mathbb{H}}\rangle\!\rangle\langle\!\langle J_i^{\mathbb{H}}| + |J_i^{\mathbb{AH}}\rangle\!\rangle\langle\!\langle J_i^{\mathbb{AH}}| + \sum_{ij:\lambda_i+\lambda_j\neq 0} \frac{|J_{ij}^{\mathbb{H}}\rangle\!\rangle\langle\!\langle J_{ij}^{\mathbb{H}}| + |J_{ij}^{\mathbb{AH}}\rangle\!\rangle\langle\!\langle J_{ij}^{\mathbb{AH}}|}{2(\lambda_i+\lambda_j)}. \tag{3.112}$$

According to the Cauchy-Schwarz inequality,

$$\max_{\tilde{C}} F_{\mathrm{SQL}}(C^\diamond, \tilde{C}) = \max_{\tilde{C}} \frac{|\langle\!\langle H|\tilde{C}\rangle\!\rangle|^2}{\langle\!\langle \tilde{C}|B|\tilde{C}\rangle\!\rangle} = \langle\!\langle H^{\mathbb{H}}|B^{-1}|H^{\mathbb{H}}\rangle\!\rangle, \tag{3.113}$$

and the optimal $|\tilde{C}^\diamond\rangle\!\rangle = B^{-1}|H^{\mathbb{H}}\rangle\!\rangle$. Here $(\cdot)^{-1}$ denotes the Moore-Penrose pseudoinverse.

The validity of the algorithm is shown in Appendix B.

## 3.6 Ancilla-free QEC protocol

In this section, we consider the possibility of removing the noiseless ancilla assumption from the QEC protocols. In particular, we focus on one-parameter estimation when HNLS is satisfied and found a sufficient condition where ancilla-free codes exist, achieving the HL and even the same optimal HL coefficient when optimized over ancilla-assisted codes.

### 3.6.1 Commuting noise

Recall from Section 3.2.3 and Section 3.4.1, one can use noiseless ancillas to construct a QEC code, described by the projector $\Pi_{\mathfrak{c}} = |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| + |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|$ onto the code space, which asymptotically restores the unitary dynamics with non-vanishing signal

$$\frac{d\rho}{dt} = -i[\theta H_{\mathrm{eff}}, \rho], \tag{3.114}$$

where $H_{\text{eff}} = \Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}} \not\propto \Pi_{\mathfrak{c}}$, if and only if HNLS is satisfied. The optimal QFI optimized over all QEC protocols is

$$F_{\text{opt}}(T) := 4T^2 \min_{S \in \mathcal{S}} \|H - S\|^2 := 4T^2 \|H - \mathcal{S}\|^2 . \tag{3.115}$$

We address here the following open questions: (1) Under what conditions the noiseless sensing dynamics in Eq. (3.114) can be achieved with an ancilla-free QEC code. (2) Whether such code can achieve the same optimal QFI in Eq. (3.115). We give a partial answer to these questions in terms of a sufficient condition on the Hamiltonian and the Lindblad operators [Layden et al., 2019].

**Theorem 3.5** (Commuting noise)**.** *Suppose $H \notin \mathcal{S}$ and $[H, L_i] = [L_i, L_j] = [L_i^\dagger, L_j] = 0$, $\forall i, j$, i.e. every Hermitian operator in $\mathcal{S}$ commutes with each other. Then there exists a QEC code without noiseless ancilla that achieves the HL. Moreover, it achieves the same optimal asymptotic QFI [Eq. (3.115)] offered by noiseless ancilla.*

*Proof.* Recall that a QEC sensing code recovering Eq. (3.114) should satisfy the following three QEC sensing conditions:

$$\Pi_{\mathfrak{c}} H \Pi_{\mathfrak{c}} \not\propto \Pi_{\mathfrak{c}}, \tag{3.116}$$

$$\Pi_{\mathfrak{c}} L_i \Pi_{\mathfrak{c}} \propto \Pi_{\mathfrak{c}}, \quad \Pi_{\mathfrak{c}} L_i^\dagger L_j \Pi_{\mathfrak{c}} \propto \Pi_{\mathfrak{c}}, \tag{3.117}$$

We will say the code corrects the Lindblad span $\mathcal{S}$ if Eq. (3.117) satisfied. Without loss of generality, we consider only a two-dimensional code $|\mathfrak{c}_{0(1)}\rangle = \sum_{k=1}^{d} \sqrt{c_k^{0(1)}} |k\rangle$, where $\{|k\rangle\}_{k=1}^d$ is an orthonormal basis under which $H$ and $L_i$'s are diagonal. Define $d$-dimensional vectors $\mathbf{1}, \boldsymbol{h}, \boldsymbol{\ell}_i$, and $\boldsymbol{\ell}_{ij}$ such that $(\mathbf{1})_k = 1$, $(\boldsymbol{h})_k = \langle k|H|k\rangle$, $(\boldsymbol{\ell_i})_k = \langle k|L_i|k\rangle$ and $(\boldsymbol{\ell_{ij}})_k = \langle k|L_i^\dagger L_j|k\rangle$. Define the real subspace $\mathcal{S}_{\text{diag}} = \text{span}\{\mathbf{1}, \text{Re}[\boldsymbol{\ell_i}], \text{Im}[\boldsymbol{\ell_i}], \text{Re}[\boldsymbol{\ell_{ij}}], \text{Im}[\boldsymbol{\ell_{ij}}], \forall i, j\} \subseteq \mathbb{R}^d$. The optimal code can be identified from the optimal solution $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{c}}^{\mathbf{0}} - \tilde{\boldsymbol{c}}^{\mathbf{1}}$ of the following semidefinite program (SDP) [Boyd and Vandenberghe, 2004],

$$\text{maximize} \quad \langle \boldsymbol{c}, \boldsymbol{h} \rangle \tag{3.118}$$

$$\text{subject to} \quad \|\boldsymbol{c}\|_1 \leq 2, \text{ and } \langle \boldsymbol{c}, \boldsymbol{\ell} \rangle = 0, \ \forall \boldsymbol{\ell} \in \mathcal{S}_{\text{diag}}. \tag{3.119}$$

Here $\|\boldsymbol{x}\|_1 = \sum_{i=1}^d |x_i|$ denotes the one-norm in $\mathbb{R}^d$ and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^d x_i y_i$ the inner product. Choosing the optimal input quantum state $|\psi_0\rangle = \frac{1}{\sqrt{2}}(|\mathfrak{c}_0\rangle + |\mathfrak{c}_1\rangle)$, the QFI is $F(\rho(T)) = T^2 \left|\langle \boldsymbol{c^0} - \boldsymbol{c^1}, \boldsymbol{h} \rangle\right|^2$. Moreover, the optimal value of Eq. (3.118) is $2 \min_{\boldsymbol{\ell} \in \mathcal{S}_{\text{diag}}} \|\boldsymbol{h} + \boldsymbol{\ell}\|_\infty$ with the argument of the minimum denoted by $\boldsymbol{\ell}^\diamond$. Here $\|\cdot\|_\infty$ denotes the infinity/max norm, defined as the largest absolute value of elements in a vector. The optimal solution $\tilde{\boldsymbol{c}}^{\boldsymbol{0(1)}}$ can be obtained from the constraint that it is in the span of vectors $\boldsymbol{v}$ such that $\langle \boldsymbol{v}, \boldsymbol{h} + \boldsymbol{\ell}^\diamond \rangle$ is the largest (smallest) [Boyd and Vandenberghe, 2004]. In this case, $F(\rho(T)) = 4T^2 \|\boldsymbol{h} - \mathcal{S}_{\text{diag}}\|_\infty^2$ is the same as $F_{\text{opt}}$ in Eq. (3.115) for noiseless ancilla. Therefore, we conclude that $\tilde{\boldsymbol{c}}^{\boldsymbol{0(1)}}$ gives the optimal code. □

Theorem 3.5 reveals that the need for noiseless ancilla arises from the non-commuting nature of the Hamiltonian and Lindblad operators. Indeed, we can find a non-trivial example with $[H, L_i] \neq 0$ (see Appendix C) for which there exist no ancilla-free QEC codes—even when we can extend $d$ to arbitrarily large. It is known that when the system dimension ($\dim \mathcal{H}_S = d$) is sufficiently large compared to the dimension of the noise space ($\dim \mathcal{S}$), a QEC code satisfying the Knill–Laflamme condition always exists [Knill et al., 2000, Theorem 4]. Therefore, the role of noiseless ancilla in quantum sensing could not be replaced by a simple extension of the system dimension, as in traditional QEC.

### 3.6.2 Chebyshev code for photon loss

We now consider an explicit example of quantum sensors dominated by commuting noise—Hamiltonian estimation under photon loss, where the signal Hamiltonians are diagonal in Fock basis. Although in principle, photon annihilation operator does not commute with the Hamiltonian, they does when we restrict the code in a subspace whose photon numbers of Fock basis are separated from each other with a distance at least 3 and therefore the commuting noise result applies.

#### Kerr effect with photon loss

Before we delve into the details to general diagonal Hamiltonian estimation, we consider an example, which is the most relevant in practice—Kerr Hamiltonian [Walls and Milburn,

2007] estimation under photon loss [Chuang et al., 1997]:

$$\frac{d\rho}{dt} = -i\Big[\theta(a^\dagger a)^2, \rho\Big] + \kappa\Big(a\rho a^\dagger - \frac{1}{2}\{a^\dagger a, \rho\}\Big), \tag{3.120}$$

where $a$ is the photon annihilation operator. In this case the probe is infinite dimensional, but suppose we assume that the occupation number $n = a^\dagger a$ is bounded: $n \leq M$, where $M$ is even. The noise source is photon loss, with Lindblad operator $L \propto a$. Can we find a QEC code that protects the probe against loss and achieves the HL for estimation of $\theta$?

To solve the dual program, we find real parameters $\chi_0$, $\chi_1$, $\chi_+$, $\chi_-$ which minimize the operator norm of

$$\tilde{n}^2 := n^2 + \chi_1 n + \chi_- a + \chi_+ a^\dagger + \chi_0, \tag{3.121}$$

where $n \leq M$. Since $a$ and $a^\dagger$ are off-diagonal in the occupation number basis, we should set $\chi_\pm$ to zero for the purpose of minimizing the difference between the largest and smallest eigenvalue of $\tilde{n}^2$. After choosing $\chi_1$ such that $\tilde{n}^2$ is minimized at $n = M/2$, and choosing $\chi_0$ so that the maximum and minimum eigenvalues of $\tilde{n}^2$ are equal in absolute value and opposite in sign, we have the optimal

$$\left(\tilde{n}^2\right)^\diamond = \left(n - \frac{1}{2}M\right)^2 - \frac{1}{8}M^2, \tag{3.122}$$

which has operator norm $\left\|(\tilde{n}^2)^\diamond\right\| = M^2/8$; hence the optimal QFI after evolution time $t$ is $F_{\text{opt}}(T) = T^2M^4/16$, according to Eq. (3.115). For comparison, the minimum operator norm is $M^2/2$ for a noiseless bosonic mode with $n \leq M$. We see that loss reduces the precision of our estimate of $\theta$, but only by a factor of 4 if we use the optimal QEC code. the HL can still be maintained. The scaling $\delta\hat{\theta} \sim 1/M^2$ of the optimal precision arises from the nonlinear photon-photon interactions in the Hamiltonian Eq. (3.120) [Boixo et al., 2007].

To find the code states, we note that the eigenstate of $(\tilde{n}^2)^\diamond$ with the lowest eigenvalue $-M^2/8$ is $|n = M/2\rangle$, while the largest eigenvalue $+M^2/8$ has the two degenerate eigenstates $|n = 0\rangle$ and $|n = M\rangle$. The code condition [[2]] requires that both code vectors have

the same expectation value of $L^\dagger L \propto n$, and we therefore may choose

$$|\mathfrak{c}_0\rangle = |M/2\rangle_S \otimes |0\rangle_A,\ |\mathfrak{c}_1\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_S + |M\rangle_S\right) \otimes |1\rangle_A \tag{3.123}$$

as the code achieving optimal precision. For $M \geq 4$, the ancilla may be discarded, and we can use the simpler code

$$|\mathfrak{c}_0\rangle = |M/2\rangle_S,\ |\mathfrak{c}_1\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_S + |M\rangle_S\right), \tag{3.124}$$

which is easier to realize experimentally. Eq. (3.20) and Eq. (3.21) are still satisfied without the ancilla, because the states $\{|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle, a|\mathfrak{c}_0\rangle, a|\mathfrak{c}_1\rangle\}$ are all mutually orthogonal. This encoding Eq. (3.124) belongs to the family of "binomial quantum codes" which, as discussed in [Michael et al., 2016], can protect against loss of bosonic excitations.

An experimental realization of this coding scheme can be achieved using tools from circuit quantum electrodynamics, by coupling a single transmon qubit to two microwave waveguide resonators. For example, when $M$ is a multiple of 4, $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$ both have even photon parity while $a|\mathfrak{c}_0\rangle$ and $a|\mathfrak{c}_1\rangle$ both have odd parity. Then QEC can be carried out by the following procedure [Sun et al., 2014; Ofek et al., 2016; Heeres et al., 2017; Hu et al., 2019]: (1) A quantum non-demolition parity measurement is performed to check whether photon loss has occurred. (2) If photon loss is detected, the initial logical encoding is restored using optimal control pulses. (3) If there is no photon loss, the quantum state is projected onto the code space span$\{|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle\}$ [Shen et al., 2017]. The probability of an uncorrectable logical error becomes arbitrarily small if the QEC procedure is sufficiently fast compared to the photon loss rate. Meanwhile, the Kerr signal accumulates coherently in the relative phase of $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$, so that the HL can be attained for arbitrarily fast quantum control. For integer values of $M$ which are not a multiple of 4, coding schemes can still be constructed which protect against photon loss, as described in [Michael et al., 2016].

### Chebyshev code

Consider the following general Hamiltonian estimation under photon loss:

$$\frac{d\rho}{dt} = -i\Big[\sum_{i=1}^{s} \zeta_i (a^\dagger a)^i, \rho\Big] + \kappa\Big(a\rho a^\dagger - \frac{1}{2}\{a^\dagger a, \rho\}\Big). \tag{3.125}$$

We only consider Hamiltonians that are a function of the photon number $a^\dagger a$, applying a cutoff at the $s$-th power, where $s > 1$ is a positive integer. According to the HNLS condition, while $\zeta_1$ cannot be sensed at the Heisenberg limit, $\theta = \zeta_s$ asymptotically can, with the optimal code for $s = 2$ provided above.

To sense $\theta$, it is important to filter out all lower-order signals $\sum_{i=1}^{s-1} \zeta_i (a^\dagger a)^i$ using the QEC code (assuming we have no information of $\{\zeta_i, 0 \le i \le s-1\}$). We should use the following modified Lindblad span:

$$\mathcal{S} = \mathrm{span}_{\mathbb{H}}\{\mathbb{1}, a, a^\dagger, (a^\dagger a)^i,\ 1 \le i \le s-1\}. \tag{3.126}$$

Although the photon loss noise is not commuting because $[a, (a^\dagger a)^i] \neq 0$. this type of off-diagonal noise can be tackled by simply ensuring the distance of the supports (non-vanishing Fock states) of $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$ is at least 3.

To obtain the optimal code, we could solve the SDP in Eqs. (3.118)–(3.119). However, when $M$ is sufficiently large, we obtain a near-optimal solution analytically by observing that for large $M$, minimizing $\|(a^\dagger a)^s - \sum_{i=0}^{s-1} \chi_i (a^\dagger a)^i\|$ over all $\{\chi_i\}_{i=0}^{s-1}$ is equivalent to approximating a $s$-th degree polynomial using an $(s-1)$-degree polynomial. The optimal polynomial is the Chebyshev polynomial [Mason and Handscomb, 2002] and the near-optimal code, that we call the *$s$-th order Chebyshev code* [Layden et al., 2019], is supported by its max/min points:

$$|\mathfrak{c}_0\rangle = \sum_{k \text{ even}}^{[0,s]} \tilde{c}_k \left|\left\lfloor M \sin^2(k\pi/2s)\right\rfloor\right\rangle, \quad |\mathfrak{c}_1\rangle = \sum_{k \text{ odd}}^{[0,s]} \tilde{c}_k \left|\left\lfloor M \sin^2(k\pi/2s)\right\rfloor\right\rangle, \tag{3.127}$$

where $\lfloor x \rfloor$ denotes the largest integer $\le x$, and $|\tilde{c}_k|^2$ can be obtained from solving a linear system of equations of size $O(s^2)$. $|\tilde{c}_k|^2$ is approximately equal to $\frac{2}{s} - \frac{1}{s}\delta_{ks} - \frac{1}{s}\delta_{k0}$ for

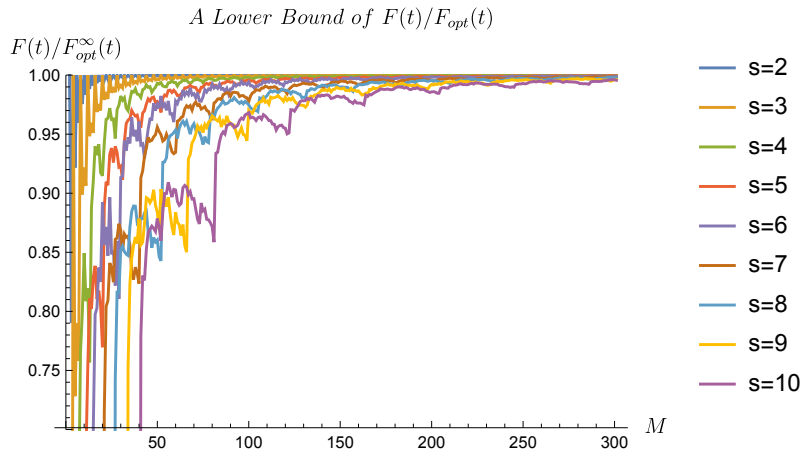


Figure 3.5: The near-optimality of the Chebyshev code ($s$ increases as the curve moves from left to right). $F^{\infty}_{\text{opt}}(T)$ is an upper bound of $F_{\text{opt}}(T)$ which is asymptotically tight as $M \to \infty$. The horizonal axis indicates $M$, the largest photon allowed in the bosonic channel and the vertical axis indicates $F(T)/F^{\infty}_{\text{opt}}(t)$, a lower bound of $F(T)/F_{\text{opt}}(T)$. When $M$ is sufficiently large, $F(T)$ is very close to its optimal value $F_{\text{opt}}(T)$; and when $s$ increases, we will need a larger $M$ to achieve the optimality. For example, when $s \leq 3$, $F(t)$ reaches 90% of the upper bound at $M = 10$.

sufficiently large $M$. Compared to binomial codes [Michael et al., 2016] another type of bosonic codes whose codewords are superpositions of Fock states and which coincide with Chebyshev codes at $s = 2$, Chebyshev codes have an almost uniform amplitude distribution but unevenly separated photon numbers for Fock basis; while binomial codes have a binomial amplitude distribution but evenly separated photon numbers for Fock basis.

In quantum sensing, the $s$-th order Chebyshev code corrects the Lindblad span (Eq. (3.126)) and provides a near optimal asymptotic QFI for $\theta$

$$F(\rho(T)) \approx F_{\text{opt}}(T) \approx 16T^2 \left(\frac{M}{4}\right)^{2s}, \tag{3.128}$$

for sufficiently large $M$, as illustrated in Figure 3.5. We provide the exact value of $\tilde{c}_k$ and prove the near-optimality of it Appendix D.

# Chapter 4

# Asymptotic Quantum Channel Estimation

In Section 2.3, we introduced the channel QFI $\mathfrak{F}_1(\mathcal{E}_\theta)$ as the QFI optimized over all possible input states that might be entangled over a probe and an ancilla, and the $N$-channel QFI $\mathfrak{F}_N(\mathcal{E}_\theta) := \mathfrak{F}_1(\mathcal{E}_\theta^{\otimes N})$ which is the channel QFI of $N$ copies of the channel. In the chapter, we study the behavior of $\mathfrak{F}_N(\mathcal{E}_\theta)$ asymptotically $N \to \infty$ and we will sometimes call $\mathfrak{F}_N(\mathcal{E}_\theta)$ the asymptotic QFI to represent the context where $N$ is large. In particular, the HL means the situation where the asymptotic QFI is $\Theta(N^2)$ and the SQL represents linear scaling.

In general, the asymptotic QFI of a quantum system, follows either the HL or the SQL and there was not a unified approach to determine the scaling. For quantum channels where the scalings are known, it is also crucial to understand how to achieve the asymptotic QFI. For example, for unitary channels, the HL is achievable and a GHZ state in the multipartite two-level systems consisting of the lowest and highest energy states is optimal [Giovannetti et al., 2006]. Under the effect of noise, a variety of quantum strategies were also proposed to enhance the QFI [Caves, 1981; Wineland et al., 1992; Huelga et al., 1997; Ulam-Orgikh and Kitagawa, 2001; Demkowicz-Dobrzański et al., 2013; Chaves et al., 2013; Gefen et al., 2016; Plenio and Huelga, 2016; Albarelli et al., 2017, 2018; Matsuzaki et al., 2011; Chin et al., 2012; Smirne et al., 2016; Liu and Yuan, 2017b; Xu et al., 2019; Chabuda et al., 2020; Zhou and Jiang, 2020b], but no conclusions for general quantum channels were drawn. One natural question to ask is whether entanglement between probes can improve the QFI. For

example, when estimating the noise parameter in the dissipative low-noise channels [Hotta et al., 2005, 2006] or teleportation-covariant channels [Pirandola et al., 2017; Pirandola and Lupo, 2017; Takeoka and Wilde, 2016; Laurenza et al., 2018] (e.g. Pauli or erasure channels), the asymptotic QFI follows the SQL and is achievable using only product states. However, when estimating the phase parameter in dephasing channels, although the HL is still not achievable, product states are no longer optimal and the asymptotic QFI is then achievable using spin-squeezed states [Huelga et al., 1997; Ulam-Orgikh and Kitagawa, 2001; Demkowicz-Dobrzański and Maccone, 2014].

Given a quantum channel, we aim to answer the following two important questions: how to determine whether the HL is achievable, and in both cases, how to find a metrological protocol achieving the asymptotic QFI? In this chapter, we answer these two open problems in the setting of ancilla-assisted channel estimation by providing an optimal QEC metrological protocol which entangles both the probe and a clean ancillary system. In Chapter 3, we introduced QEC protocols for Hamiltonian estimation under Markovian noise where we assumed fast and frequent quantum operations which have limited practical applications and the channel estimation framework partially solves this problem.

In this chapter, we construct a two-dimensional QEC protocol which reduces every quantum channel to a single-qubit dephasing channel where both the phase and the noise parameter could vary w.r.t. the unknown parameter. We first identify the asymptotic QFI for all single-qubit dephasing channels and then show that the asymptotic QFI of the logical dephasing channel is no smaller than the one of the original quantum channel after optimizing over the encoding and the recovery channel, proving the sufficiency of our QEC protocol. Using the above proof strategy, we obtain the asymptotic theory of quantum channel estimation, closing a long-standing open question in theoretical quantum metrology. We also push one step further towards achieving the ultimate estimation limit in practical quantum sensing experiments by providing efficiently solvable asymptotic QFIs and corresponding optimal estimation protocols.

## 4.1 "Hamiltonian-not-in-Kraus-span" condition

Recall from Section 2.3.2, we established the HNKS condition as a necessary and sufficient condition to achieve the HL for a given quantum channel using the following upper bounds for parallel and sequential strategies respectively:

$$\mathfrak{F}_N(\mathcal{E}_\theta) \leq 4 \min_h \left(N \|\alpha\| + N(N-1) \|\beta\|^2\right), \tag{4.1}$$

$$\mathfrak{F}_N^{(\text{seq})}(\mathcal{E}_\theta) \leq 4 \min_h \left(N \|\alpha\| + N(N-1) \|\beta\| \left(\|\beta\| + 2\sqrt{\|\alpha\|}\right)\right). \tag{4.2}$$

If there is an $h$ such that $\beta = 0$,

$$\mathfrak{F}_N(\mathcal{E}_\theta) \leq \mathfrak{F}_N^{(\text{seq})}(\mathcal{E}_\theta) \leq 4 \min_{h:\beta=0} N \|\alpha\|, \tag{4.3}$$

$\mathfrak{F}_N(\mathcal{E}_\theta)$ and $\mathfrak{F}_N^{(\text{seq})}(\mathcal{E}_\theta)$ follow the SQL asymptotically. Therefore, it is only possible to achievable the HL if the HNKS condition holds, i.e. $H \notin \mathcal{S}$, where

$$H = i\mathbf{K}^\dagger \dot{\mathbf{K}}, \quad \mathcal{S} = \text{span}_{\mathbb{H}}\{K_i^\dagger K_j, \forall i, j\}. \tag{4.4}$$

Note that the definitions of $H$ and $\mathcal{S}$ are not exactly the same as the ones in Chapter 3 (differ by constant factors) but we still use the same notations here because they are essentially equivalent and have the same physical meaning.

We will show in this section that HNKS is also a sufficient condition to achieve the HL for parallel strategies in Figure 2.1(b), and hence, sequential strategies in Figure 2.1(c) that contain the former. We summarize this by the following theorem:

**Theorem 4.1.** *$\mathfrak{F}_N(\mathcal{E}_\theta) = \Theta(N^2)$ if and only if $H \notin \mathcal{S}$. Otherwise, $\mathfrak{F}_N(\mathcal{E}_\theta) = \Theta(N)$. The statement is also true for $\mathfrak{F}_N^{(\text{seq})}(\mathcal{E}_\theta)$.*

Furthermore, the QFI upper bound in Eq. (4.3) is achievable asymptotically when $H \in \mathcal{S}$ for both parallel and sequential strategies:

**Theorem 4.2.** *When $H \in \mathcal{S}$,*

$$\mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta) := \lim_{N\to\infty} \mathfrak{F}_N(\mathcal{E}_\theta)/N = 4 \min_{h:\beta=0} \|\alpha\| . \tag{4.5}$$

*For any $\eta > 0$, there exists an input state $|\psi_{\eta,N}\rangle$ solvable via a SDP such that $\lim_{N\to\infty} F((\mathcal{E}_\theta^{\otimes N} \otimes \mathbb{1})(|\psi_{\eta,N}\rangle))/N > \mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta) - \eta$. Furthermore, $\mathfrak{F}_{\text{SQL}}^{(\text{seq})}(\mathcal{E}_\theta) = \mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta)$.*

Note that $\mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta)$ is named "asymptotic channel QFI" in [Kołodyński and Demkowicz-Dobrzański, 2013]. The quadratic term of the QFI upper bound in Eq. (4.1) is also achievable when $H \notin \mathcal{S}$ for parallel strategies:

**Theorem 4.3.** *When $H \notin \mathcal{S}$,*

$$\mathfrak{F}_{\text{HL}}(\mathcal{E}_\theta) := \lim_{N\to\infty} \mathfrak{F}_N(\mathcal{E}_\theta)/N^2 = 4 \min_{h} \|\beta\|^2 . \tag{4.6}$$

*There exists an input state $|\psi_N\rangle$ solvable via a SDP such that $F((\mathcal{E}_\theta^{\otimes N} \otimes \mathbb{1})(|\psi_N\rangle))/N^2 = \mathfrak{F}_{\text{HL}}(\mathcal{E}_\theta)$.*

Note that without the help of the ancilla system, the QFI upper bound in Eq. (4.3) may not be achievable asymptotically [Knysh et al., 2014; Layden et al., 2019]. For example, the upper bound in Eq. (4.3) for phase estimation in amplitude damping channels is reduced by a factor of four without ancilla [Knysh et al., 2014; Demkowicz-Dobrzański and Maccone, 2014].

Theorem 4.2 indicates that when HNKS is violated (which almost surely happens statistically), there is no advantage of sequential strategies over parallel strategies asymptotically, as conjectured in [Demkowicz-Dobrzański and Maccone, 2014]. Interestingly, similar results were discovered quantum channel discrimination, a related field [Hayashi, 2002; Yuan and Fung, 2017; Pirandola et al., 2019; Chen and Yuan, 2019; Katariya and Wilde, 2020a; Chiribella et al., 2008; Yang, 2019]. It was recently proven that sequential strategies cannot outperform parallel strategies asymptotically in asymmetric discrimination of two arbitrary quantum channels [Hayashi, 2009; Wilde et al., 2020; Wang and Wilde, 2019; Fang et al., 2020]. Our result is different, however, because the QFI cannot be characterized as the limit of quantum relative entropy [Hayashi, 2002] and it is also unclear how to interpret

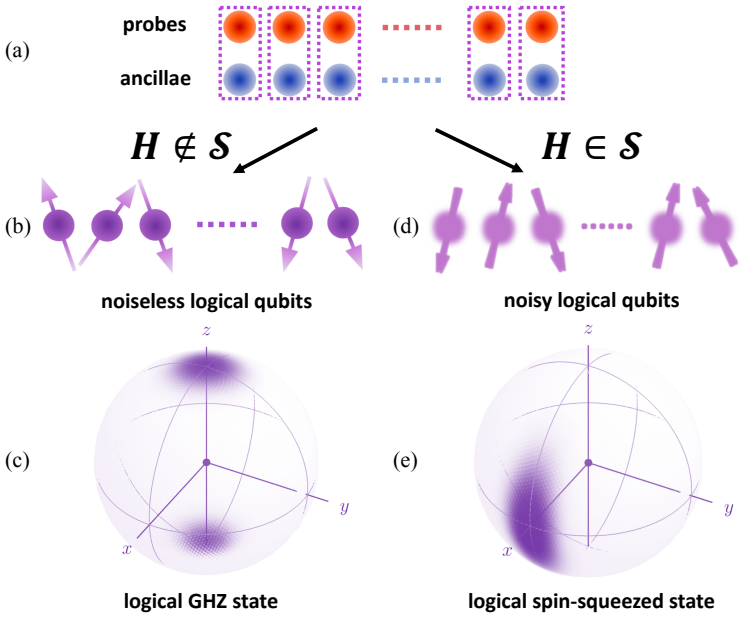


Figure 4.1: The optimal metrological protocol. (a) The original physical system where we have $N$ noisy probes and $N$ noiseless ancillas. Each pair of probe-ancilla subsystem (purple box) encodes a logical qubit (see Section 4.2). (b,c) When $H \notin \mathcal{S}$, the logical qubits are noiseless. We choose the GHZ state of $N$-logical qubits as the optimal input. (d,e) When $H \in S$, each logical qubit is subject to an effective dephasing noise. We choose the spin-squeezed state of the $N$-logical qubits with suitable parameters as the optimal input. We plot the quasiprobability distribution $Q(\theta,\varphi) = |\langle\theta,\varphi|\psi\rangle|^2$ on a sphere using coordinates $(x,y,z) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$ [Kitagawa and Ueda, 1993], where $|\theta,\varphi\rangle = (\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle)^{\otimes N}$ and $N = 50$. (Darker colors indicate larger values.)

the HNKS condition in terms of asymmetric channel discrimination. Moreover, we provide a constructive proof with explicit and efficiently computable QEC metrological protocols, which paves the way for practical implementation of error-corrected sensing schemes.

Based on the previous discussion, in order to prove the theorems, it is sufficient to provide a QEC protocol using parallel strategies which achieves the QFI upper bound (Eq. (4.1)) asymptotically both when $H \in \mathcal{S}$ or $H \notin \mathcal{S}$. Thus we will focus only on parallel strategies in the following. We first show Theorem 4.3 and Theorem 4.2 are true for the generalized single-qubit dephasing channels in Section 4.2 where both the phase and the noise parameter vary w.r.t. $\theta$. Then we will generalized the results to arbitrary quantum channels $\mathcal{E}_\theta$ using a QEC protocol in Sections 4.3-4.4. The two steps are summarized in Figure 4.1.

## 4.2 Reduction to dephasing channels

### 4.2.1 Asymptotic QFI of dephasing channels

According to Eq. (4.1), $\mathfrak{F}_{\text{HL}} \le \mathfrak{F}_{\text{HL}}^{(u)}$ and $\mathfrak{F}_{\text{SQL}} \le \mathfrak{F}_{\text{SQL}}^{(u)}$, where $\mathfrak{F}_{\text{HL}}^{(u)} := 4\min_h \|\beta\|^2$ and $\mathfrak{F}_{\text{SQL}}^{(u)} := 4\min_{h:\beta=0}\|\alpha\|$. $^{(u)}$ refers to the upper bounds here. In this section, we will show the above equalities hold for any single-qubit dephasing channel

$$\mathcal{D}_\theta(\rho) = (1-p)e^{-\frac{i\phi}{2}\sigma_z}\rho e^{\frac{i\phi}{2}\sigma_z} + p\sigma_z e^{-\frac{i\phi}{2}\sigma_z}\rho e^{\frac{i\phi}{2}\sigma_z}\sigma_z, \tag{4.7}$$

which is the composition of the conventional dephasing channel $\rho \mapsto (1-p)\rho + p\sigma_z\rho\sigma_z$ $(0 \le p < 1)$ and the rotation in the $z$-direction $\rho \mapsto e^{-\frac{i\phi}{2}\sigma_z}\rho e^{\frac{i\phi}{2}\sigma_z}$. Both $p$ and $\phi$ are functions of an unknown parameter $\theta$. The HNKS condition is equivalent to $p = 0$ and the QFI upper bounds for $\mathcal{D}_\theta$ are

$$\mathfrak{F}_{\text{HL}}^{(u)}(\mathcal{D}_\theta) = |\dot{\xi}|^2, \quad \mathfrak{F}_{\text{SQL}}^{(u)}(\mathcal{D}_\theta) = \frac{|\dot{\xi}|^2}{1-|\xi|^2}, \tag{4.8}$$

where $\xi = \langle 0|\,\mathcal{D}_\theta(|0\rangle\langle 1|)\,|1\rangle = (1-2p)e^{-i\phi}$.

To show Eq. (4.8), let $\mathcal{D}_\theta(\rho) = \sum_{i=1}^2 K_i\rho K_i^\dagger$, where $K_1 = \sqrt{1-p}e^{-\frac{i\phi}{2}\sigma_z}, K_2 = \sqrt{p}\sigma_z e^{-\frac{i\phi}{2}\sigma_z}$. Assume $p > 0$, then

$$\mathbf{K} = \begin{pmatrix} \sqrt{1-p}e^{-\frac{i\phi}{2}\sigma_z} \\ \sqrt{p}\sigma_z e^{-\frac{i\phi}{2}\sigma_z} \end{pmatrix}, \quad \dot{\mathbf{K}} = \begin{pmatrix} (\frac{-\dot{p}}{2\sqrt{1-p}} - \sqrt{1-p}\frac{i\dot{\phi}}{2}\sigma_z)e^{-\frac{i\phi}{2}\sigma_z} \\ (\frac{\dot{p}}{2\sqrt{p}} - \sqrt{p}\frac{i\dot{\phi}}{2}\sigma_z)e^{-\frac{i\phi}{2}\sigma_z}\sigma_z \end{pmatrix}, \tag{4.9}$$

$$\dot{\mathbf{K}} - ih\mathbf{K} = \begin{pmatrix} (\frac{-\dot{p}}{2\sqrt{1-p}} - ih_{11}\sqrt{1-p} - \sqrt{1-p}\frac{i\dot{\phi}}{2}\sigma_z - ih_{12}\sqrt{p}\sigma_z)e^{-\frac{i\phi}{2}\sigma_z} \\ (\frac{\dot{p}}{2\sqrt{p}}\sigma_z - ih_{22}\sqrt{p}\sigma_z - \sqrt{p}\frac{i\dot{\phi}}{2} - ih_{21}\sqrt{1-p})e^{-\frac{i\phi}{2}\sigma_z} \end{pmatrix}, \tag{4.10}$$

$$\beta = i\mathbf{K}^\dagger(\dot{\mathbf{K}} - ih\mathbf{K}) = \frac{\dot{\phi}}{2}\sigma_z + (1-p)h_{11} + ph_{22} + \sqrt{p(1-p)}(h_{12}+h_{21})\sigma_z, \tag{4.11}$$

$$\begin{aligned}\alpha = (\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K}) &= \frac{\dot{p}^2}{4p(1-p)} + h_{11}^2(1-p) + h_{22}^2 p + \frac{\dot{\phi}^2}{4} + |h_{12}|^2 + 2\sqrt{p(1-p)}\dot{\phi}\text{Re}[h_{12}] \\ &+ 2\text{Re}\Big[-\frac{\dot{p}\sqrt{p}}{\sqrt{1-p}}ih_{12} + ((1-p)h_{11} + h_{22}p)\frac{\dot{\phi}}{2} + (h_{11}h_{12} + h_{22}h_{21})\sqrt{p(1-p)} - i\frac{\dot{p}\sqrt{1-p}}{\sqrt{p}}h_{21}\Big]\sigma_z.\end{aligned} \tag{4.12}$$

$\beta = 0$ is equivalent to $(1-p)h_{11} + ph_{22} = 0$ and $\frac{\dot{\phi}}{2} + \sqrt{p(1-p)}(h_{12}+h_{21}) = 0$, which

is achievable for any $p > 0$. When $h_{11} = h_{22} = 0$ and $h_{12} = h_{21} = -\frac{\dot{\phi}}{4\sqrt{p(1-p)}}$, $\|\alpha\| = \min_{h:\beta=0} \|\alpha\| = \frac{(1-2p)^2\dot{\phi}^2}{16p(1-p)} + \frac{\dot{p}^2}{4(1-p)p}$. Then

$$\mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{D}_\theta) = 4 \min_{h:\beta=0} \|\alpha\| = \frac{(1-2p)^2\dot{\phi}^2}{4p(1-p)} + \frac{\dot{p}^2}{(1-p)p} = \frac{|\dot{\xi}|^2}{1-|\xi|^2}, \tag{4.13}$$

where $\xi = (1-2p)e^{-i\phi} = \langle 0| \mathcal{D}_\theta(|0\rangle\langle 1|)|1\rangle$ is a complex number completely determining the channel. When $p = 0$, we must also have $\dot{p} = 0$. Then $\beta = \frac{\dot{\phi}}{2}\sigma_z + h_{11}$ and

$$\mathfrak{F}^{(u)}_{\mathrm{HL}}(\mathcal{D}_\theta) = 4 \min_{h} \|\beta\|^2 = |\dot{\phi}|^2 = |\dot{\xi}|^2. \tag{4.14}$$

We can also calculate the channel QFI

$$\mathfrak{F}_1(\mathcal{D}_\theta) = 4 \min_{h} \|\alpha\| = \begin{cases} (1-2p)^2\dot{\phi}^2 + \frac{\dot{p}^2}{(1-p)p}, & p > 0, \\ (1-2p)^2\dot{\phi}^2, & p = 0. \end{cases} \tag{4.15}$$

It could be achieved using $|\psi_0\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$.

Now we show that $\mathfrak{F}_{\mathrm{HL,SQL}}(\mathcal{D}_\theta) = \mathfrak{F}^{(u)}_{\mathrm{HL,SQL}}(\mathcal{D}_\theta)$ and provide the optimal input states in both cases. When HNKS is satisfied ($p = 0$), $\mathcal{D}_\theta$ is unitary. Using the GHZ state $|\psi_0\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$ as the input state, we could achieve

$$F(\mathcal{D}_\theta^{\otimes N}(|\psi_0\rangle\langle\psi_0|)) = |\dot{\xi}|^2 N^2, \tag{4.16}$$

which implies $\mathfrak{F}_{\mathrm{HL}}(\mathcal{D}_\theta) = \mathfrak{F}^{(u)}_{\mathrm{HL}}(\mathcal{D}_\theta)$.

To calculate the optimal QFI when HNKS is violated ($p > 0$), we will use the following two useful formulas. For any pure state input $|\psi_0\rangle$ and output $\rho_\theta = \mathcal{D}_\theta^{\otimes N}(|\psi_0\rangle\langle\psi_0|)$, we have, for all $N$,

$$F(\rho_\theta) = F_p(\rho_\theta) + F_\phi(\rho_\theta), \tag{4.17}$$

where $F_p(\rho_\theta) = \mathrm{Tr}(L_p^2\rho_\theta)$ is the QFI w.r.t. $\theta$ when only the noise parameter $p$ varies w.r.t. $\theta$, where the SLD $L_p$ satisfies $\frac{1}{2}\frac{\partial\rho_\theta}{\partial p}\dot{p} = L_p\rho_\theta + \rho_\theta L_p$. Similarly, $F_\phi(\rho_\theta)$ is the QFI w.r.t. $\theta$ when only the phase parameter $\phi$ varies w.r.t. $\theta$.

To prove Eq. (4.17), $|\psi\rangle = e^{-i\phi J_z}|\psi_0\rangle$ and a subspace

$$\mathcal{Z} = \text{span}\Big\{\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}|\psi\rangle, (j_1,\ldots,j_N)\in\{0,1\}^N\Big\}. \tag{4.18}$$

Assume $\dim\mathcal{Z} = n$. $\mathcal{Z}$ must have an orthonormal basis $\{|e_\ell\rangle\}_{\ell=1}^n$ where $|e_\ell\rangle = \sum_{j_1,\ldots,j_N=0}^{1} r_{\ell,(j_1,\ldots,j_N)}\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}|\psi\rangle$ with real $r_{\ell,(j_1,\ldots,j_N)}$. For example, one can use the Gram-Schmidt procedure to find $\{|e_\ell\rangle\}_{\ell=1}^n$ because $\langle\psi|\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}|\psi\rangle \in \mathbb{R}$ for all $(j_1,\ldots,j_N)\in\{0,1\}^{\otimes N}$.

Then

$$\begin{aligned}\rho_\theta &= \mathcal{D}_\theta^{\otimes N}(|\psi_0\rangle\langle\psi_0|) = (\mathcal{D}_\theta|_{\phi=0})^{\otimes N}(|\psi\rangle\langle\psi|)\\ &= \sum_{j_1,\ldots,j_N=0}^{1}(1-p)^{(N-\sum_{k=1}^N j_k)}p^{(\sum_{k=1}^N j_k)}\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}|\psi\rangle\langle\psi|\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}\\ &= \sum_{\ell,\ell'=1}^{n}\chi_{\ell\ell'}|e_\ell\rangle\langle e_{\ell'}|\end{aligned} \tag{4.19}$$

where $\chi\in\mathbb{R}^{n\times n}$ is a symmetric matrix. $\chi = \sum_{i=1}^n \mu_i v_i v_i^T$ where $v_i$ are real orthonormal eigenvectors of $\chi$. Then we can write $\rho_\theta = \sum_{\ell=1}^n \mu_\ell|\psi_\ell\rangle\langle\psi_\ell|$ where $|\psi_\ell\rangle = \sum_{\ell'=1}^n v_{\ell\ell'}|e_{\ell'}\rangle$. Then according to the definition of QFI,

$$F(\rho_\theta) = 2\sum_{\ell\ell':\mu_\ell+\mu_{\ell'}\neq 0}\frac{|\langle\psi_\ell|\dot{\rho}_\theta|\psi_{\ell'}\rangle|^2}{\mu_\ell+\mu_{\ell'}}. \tag{4.20}$$

Note that in principle Eq. (4.20) only holds true when $\{|\psi_\ell\rangle\}$ is a complete basis of $\mathcal{H}_S^{\otimes N}$, that is, $\text{span}\{|\psi_\ell\rangle\} = \mathcal{H}_S^{\otimes N}$. However, we are allowed to restrict the summation in the RHS of Eq. (4.20) to states in the subspace $\mathcal{Z}$, i.e. $\text{span}\{|\psi_\ell\rangle\} = \mathcal{Z}$, because $\Pi_\mathcal{Z}\rho_\theta\Pi_\mathcal{Z} = \rho_\theta$ and $\Pi_\mathcal{Z}\dot{\rho}_\theta\Pi_\mathcal{Z} = \dot{\rho}_\theta$, i.e. any state perpendicular to $\mathcal{Z}$ does not contribute to the QFI.

The derivative of $\rho_\theta$ w.r.t. $\theta$ is

$$\begin{aligned}\dot{\rho}_\theta = \frac{\partial\rho_\theta}{\partial p}\dot{p} + \frac{\partial\rho_\theta}{\partial\phi}\dot{\phi} &= \sum_{j_1,\ldots,j_N=0}^{1}\frac{\partial(1-p)^{(N-\sum_{k=1}^N j_k)}p^{(\sum_{k=1}^N j_k)}}{\partial\theta}\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}|\psi\rangle\langle\psi|\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}\\ &+ \sum_{j_1,\ldots,j_N=0}^{1}(1-p)^{(N-\sum_{k=1}^N j_k)}p^{(\sum_{k=1}^N j_k)}\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}\frac{\partial|\psi\rangle\langle\psi|}{\partial\theta}\prod_{k=1}^{N}(\sigma_z^{(k)})^{j_k}.\end{aligned} \tag{4.21}$$

Then we have

$$\langle\psi_\ell|\,\dot{\rho}_\theta\,|\psi_{\ell'}\rangle = a_{\ell\ell'} + ib_{\ell\ell'}, \tag{4.22}$$

where $a_{\ell\ell'} = \langle\psi_\ell|\,\frac{\partial\rho_\theta}{\partial p}\dot{p}\,|\psi_{\ell'}\rangle \in \mathbb{R}$, $b_{\ell\ell'} = -i\,\langle\psi_\ell|\,\frac{\partial\rho_\theta}{\partial\phi}\dot{\phi}\,|\phi_{\ell'}\rangle \in \mathbb{R}$. Therefore,

$$F(\rho_\theta) = 2\sum_{\ell\ell':\mu_\ell+\mu_{\ell'}\neq 0}\frac{|a_{\ell\ell'}|^2+|b_{\ell\ell'}|^2}{\mu_\ell+\mu_{\ell'}} = F_p(\rho_\theta)+F_\phi(\rho_\theta), \tag{4.23}$$

Another useful formula is the error propagation formula [Pezzé and Smerzi, 2009],

$$F(\rho) \geq \frac{1}{\langle\Delta J^2\rangle_\rho}\left(\frac{\partial\langle J\rangle_\rho}{\partial\theta}\right)^2, \tag{4.24}$$

for arbitrary $\rho$ as a function of $\theta$ and arbitrary Hermitian operator $J$ where $\langle J\rangle_\rho = \mathrm{Tr}(J\rho)$ and $\langle\Delta J^2\rangle_\rho = \langle J^2\rangle_\rho - \langle J\rangle_\rho^2$. The equality holds when $J$ is equal to the SLD operator of $\rho$.

Consider an $N$-qubit spin-squeezed state [Kitagawa and Ueda, 1993; Ulam-Orgikh and Kitagawa, 2001]:

$$|\psi_{\mu,\nu}\rangle = e^{-i\nu J_x}e^{-\frac{i\mu}{2}J_z^2}e^{-i\frac{\pi}{2}J_y}\,|0\rangle^{\otimes N}, \tag{4.25}$$

where $J_{x,y,z} = \frac{1}{2}\sum_{k=1}^N\sigma_{x,y,z}^{(k)}$ with $^{(k)}$ denote operators on the $k$-th qubit. Let $|\psi_0\rangle = e^{i\phi J_z}\,|\psi_{\mu,\nu}\rangle$. Using Eq. (4.17) and Eq. (4.24), we have for $\rho_\theta = \mathcal{D}_\theta^{\otimes N}(|\psi_0\rangle\,\langle\psi_0|)$,

$$F(\rho_\theta) \geq \frac{1}{\langle\Delta J_x^2\rangle_{\rho_\theta}}\left(\frac{\partial\langle J_x\rangle_{\rho_\theta}}{\partial p}\dot{p}\right)^2 + \frac{1}{\langle\Delta J_y^2\rangle_{\rho_\theta}}\left(\frac{\partial\langle J_y\rangle_{\rho_\theta}}{\partial\phi}\dot{\phi}\right)^2, \tag{4.26}$$

where

$$\langle J_{x,y}\rangle_{\rho_\theta} = (1-2p)\,\langle J_{x,y}\rangle_{|\psi_{\mu,\nu}\rangle}, \tag{4.27}$$

$$\langle J_{x,y}^2\rangle_{\rho_\theta} = \frac{N}{4} + (1-2p)^2\Big(\langle J_{x,y}^2\rangle_{|\psi_{\mu,\nu}\rangle} - \frac{N}{4}\Big), \tag{4.28}$$

$$\frac{\partial\langle J_x\rangle_{\rho_\theta}}{\partial p}\dot{p} = -2\dot{p}\,\langle J_x\rangle_{|\psi_{\mu,\nu}\rangle}, \quad \frac{\partial\langle J_y\rangle_{\rho_\theta}}{\partial\phi}\dot{\phi} = (1-2p)\dot{\phi}\,\langle J_x\rangle_{|\psi_{\mu,\nu}\rangle}. \tag{4.29}$$

It was shown in [Kitagawa and Ueda, 1993] that choosing $\nu = \frac{\pi}{2} - \frac{1}{2}\arctan\frac{b}{a}$,

$$\langle J_x\rangle_{|\psi_{\mu,\nu}\rangle} = \frac{N}{2}\cos(\mu/2)^{N-1}, \quad \langle J_y\rangle_{|\psi_{\mu,\nu}\rangle} = 0, \tag{4.30}$$

$$\langle \Delta J_x^2 \rangle_{|\psi_{\mu,\nu}\rangle} = \frac{N}{4}\left(N\left(1-\cos^{2(N-1)}\frac{\mu}{2}\right)-\left(\frac{N-1}{2}\right)a\right), \tag{4.31}$$

$$\langle \Delta J_y^2 \rangle_{|\psi_{\mu,\nu}\rangle} = \frac{N}{4}\left(1+\frac{N-1}{4}\left(a-\sqrt{a^2+b^2}\right)\right), \tag{4.32}$$

where $a = 1-\cos^{N-2}\mu$, $b = 4\sin\frac{\mu}{2}\cos^{N-2}\frac{\mu}{2}$. Let $N \gg 1$, $\mu = \Theta(N^{-5/6})$, then

$$\langle J_x \rangle_{|\psi_{\mu,\nu}\rangle} \approx \frac{N}{2}, \quad \langle \Delta J_x^2 \rangle_{|\psi_{\mu,\nu}\rangle} \approx O(N^{2/3}), \quad \langle \Delta J_y^2 \rangle_{|\psi_{\mu,\nu}\rangle} \approx O(N^{2/3}), \tag{4.33}$$

and $\langle \Delta J_x^2 \rangle_{\rho_\theta} \approx \langle \Delta J_y^2 \rangle_{\rho_\theta} \approx p(1-p)N$, $\frac{\partial \langle J_x \rangle_{\rho_\theta}}{\partial p}\dot{p} \approx -\dot{p}N$ and $\frac{\partial \langle J_y \rangle_{\rho_\theta}}{\partial \phi}\dot{\phi} \approx (1-2p)\dot{\phi}N/2$. Therefore,

$$F(\rho_\theta) \geq \frac{|\dot{\xi}|^2}{1-|\xi|^2}N + o(N), \tag{4.34}$$

which implies $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_\theta) = \mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{D}_\theta)$. Compared with $\mathfrak{F}_1(\mathcal{D}_\theta)$ (eq-4:single-dephasing), $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_\theta)$ has a factor of $1/(4p(1-p))$ enhancement when we estimate the phase parameter ($\dot{p}=0$). When we estimate the noise parameter ($\dot{\phi}=0$), however, $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_\theta) = \mathfrak{F}_1(\mathcal{D}_\theta)$. In general, $\mathfrak{F}_{\mathrm{SQL}}/\mathfrak{F}_1$ is between 1 and $1/(4p(1-p))$.

To sum up, we proved Theorem 4.3 and Theorem 4.2 are true for dephasing channels. The ancilla is not required here. When the noise is non-zero, the QFI must follow the SQL and there exists a spin-squeezed state achieving the QFI asymptotically. In particular, the squeezing parameter should be tuned carefully such that both the $J_x$ and $J_y$ variance are small such that both the noise and the phase parameter are estimated with the optimal precision.

### 4.2.2 QEC protocol

Now we introduce a QEC protocol such that every quantum channel simulates the dephasing channel introduced above. To be specific, we find the encoding channel $\mathcal{E}_{\mathrm{enc}}$ and the recovery channel $\mathcal{R}$ such that

$$\mathcal{R} \circ \mathcal{E}_\theta \circ \mathcal{E}_{\mathrm{enc}} = \mathcal{D}_{L,\theta}. \tag{4.35}$$

The construction fully utilizes the advantage of the ancilla, which has the same dimension as the probe with an extra qubit. Let $\dim \mathcal{H}_S = d$ and $\dim \mathcal{H}_A = 2d$. We pick a QEC

code

$$|\mathfrak{c}_0\rangle = \sum_{i,j=1}^{d} A_{0,ij} |i\rangle_S |j,0\rangle_A , \ |\mathfrak{c}_1\rangle = \sum_{i,j=1}^{d} A_{1,ij} |i\rangle_S |j,1\rangle_A , \tag{4.36}$$

with the encoding channel is $\mathcal{E}_{\text{enc}}(\cdot) = V(\cdot)V^\dagger$ where $V = |\mathfrak{c}_0\rangle \langle 0_L| + |\mathfrak{c}_1\rangle \langle 1_L|$, and a recovery channel

$$\mathcal{R}(\cdot) = \sum_{m=1}^{M} \left(|0_L\rangle \langle R_m, 0| + |1_L\rangle \langle Q_m, 1|\right)(\cdot)\left(|R_m, 0\rangle \langle 0_L| + |Q_m, 1\rangle \langle 1_L|\right). \tag{4.37}$$

Here $A_{0,1}$ are matrices in $\mathbb{C}^{d\times d}$ satisfying $\text{Tr}(A_{0,1}^\dagger A_{0,1}) = 1$, $R = (|R_1\rangle \cdots |R_M\rangle)$ and $Q = (|Q_1\rangle \cdots |Q_M\rangle)$ are matrices satisfying $RR^\dagger = QQ^\dagger = I$. The last ancillary qubit in $\mathcal{H}_A$ guarantees the logical channel to be dephasing, which satisfies

$$\xi = \sum_{i,m} \langle R_m, 0| K_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| K_i^\dagger |Q_m, 1\rangle , \tag{4.38}$$

and $\mathfrak{F}_{\text{HL,SQL}}(\mathcal{D}_{L,\theta})$ could then be directly calculated using Eq. (4.8). Note that in the equation above and in what follows we use $K_i$ as a substitute for $K_i \otimes I$ for the simplicity of notations. Below, we will show that by optimizing $\mathfrak{F}_{\text{HL,SQL}}(\mathcal{D}_{L,\theta})$ over both the recovery channel ($R$,$Q$) and the QEC code ($A_{0,1}$), the QFI upper bounds $\mathfrak{F}^{(u)}_{\text{HL,SQL}}(\mathcal{E}_\theta)$ are achievable.

## 4.3 Asymptotic channel QFI: HNKS

When $H \notin \mathcal{S}$, we construct a QEC code such that the HL upper bound $\mathfrak{F}^{(u)}_{\text{HL}}(\mathcal{E}_\theta)$ is achieved. For dephasing channels, the HL is achievable only if $|\xi| = 1$. Since any transformation $R \leftarrow e^{i\varphi}R$ does not affect the QFI, without loss of generality (WLOG), we assume $\xi = 1$. It means that the QEC has to be perfect, i.e. satisfies the Knill-Laflamme condition [Knill and Laflamme, 1997]

$$\Pi_{\mathfrak{c}} K_i^\dagger K_j \Pi_{\mathfrak{c}} \propto \Pi_{\mathfrak{c}}, \quad \forall i, j, \tag{4.39}$$

where $\Pi_{\mathfrak{c}} = |\mathfrak{c}_0\rangle \langle \mathfrak{c}_0| + |\mathfrak{c}_1\rangle \langle \mathfrak{c}_1|$. Moreover, there exists a Kraus representation $\{K'_i\}_{i=1}^{r'}$ such that $\Pi_{\mathfrak{c}} K_i'^\dagger K'_j \Pi_{\mathfrak{c}} = \mu_i \delta_{ij} \Pi_{\mathfrak{c}}$ and $K'_i \Pi_{\mathfrak{c}} = U_i \sqrt{\mu_i} \Pi_{\mathfrak{c}}$. The unitary $U_i$ has the form

$$U_i = U_{0,i} \otimes |0\rangle \langle 0| + U_{1,i} \otimes |1\rangle \langle 1| , \tag{4.40}$$

where $U_{0,i}$ and $U_{1,i}$ are also unitary. Let

$$|R_i\rangle = \langle 0| U_i |\mathfrak{c}_0\rangle, \quad |Q_i\rangle = \langle 0| U_i |\mathfrak{c}_0\rangle, \tag{4.41}$$

for $1 \leq i \leq r'$. We could also add some additional $|R_i\rangle$ and $|Q_i\rangle$ to them to make sure they are two complete and orthonormal bases. Then one could verify that $\xi = 1$ and

$$\dot{\xi} = -i\mathrm{Tr}((H \otimes I)\sigma_{z,\mathfrak{c}}), \tag{4.42}$$

where $\sigma_{z,\mathfrak{c}} = |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| - |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|$. Let $\tilde{C} = A_0 A_0^\dagger - A_1 A_1^\dagger$, $\dot{\xi} = -i\mathrm{Tr}(H\tilde{C})$ and the Knill-Laflamme condition is equivalent to $\mathrm{Tr}(\tilde{C}S) = 0$, $\forall S \in \mathcal{S}$. The optimization of the QFI over the QEC code becomes

$$\text{maximize } |\dot{\xi}| = |\mathrm{Tr}(H\tilde{C})|, \tag{4.43}$$

$$\text{subject to } \|\tilde{C}\|_1 \leq 2,\ \mathrm{Tr}(\tilde{C}S) = 0,\ \forall \tilde{C} \in \mathbb{H}_d,\ S \in \mathcal{S}, \tag{4.44}$$

A similar SDP problem was considered in Section 3.4.1. The optimal $|\dot{\xi}|$ is equal to $2\min_{S\in\mathcal{S}} \|H - S\|$ and the optimal $\tilde{C}$ could be found via a SDP. Any $A_0, A_1$ such that $\tilde{C}$ is optimal would achieve the optimal QFI. It means there exists an encoding, and therefore an optimal input state $|\psi_N\rangle$ which is the logical GHZ state, such that

$$\lim_{N\to\infty} \frac{F((\mathcal{E}_\theta^{\otimes N} \otimes \mathbb{1})(|\psi_N\rangle))}{N^2} = 4 \min_{S\in\mathcal{S}} \|H - S\|^2. \tag{4.45}$$

Clearly, $4\min_{S\in\mathcal{S}} \|H - S\|^2 = 4\min_h \|\beta\|^2 = \mathfrak{F}_{\mathrm{HL}}^{(u)}(\mathcal{E}_\theta)$, where we used the fact that for any $S \in \mathcal{S}$ there exists an $h \in \mathbb{H}_r$ such that $S = \mathbf{K}^\dagger h \mathbf{K}$ and vice versa. Theorem 4.3 is then proven. Note that, given the optimal $\tilde{C}$, we can always choose $A_0 A_0^\dagger$ and $A_1 A_1^\dagger$ with orthogonal supports and the last ancillary qubit in $\mathcal{H}_A$ could be removed because $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$ in this case could be distinguished using projections onto the orthogonal supports in $\mathcal{H}_A$. Therefore a $d$-dimensional ancillary system is sufficient.

We have demonstrated the QEC code achieving the optimal HL for arbitrary quantum channels. The code is designed to satisfy the Knill-Laflamme condition and optimize the

QFI. The logical dephasing channel is exactly the identity channel at the true value of $\theta$ and any change in $\theta$ results in a detectable phase, allowing it to be estimated at the HL.

## 4.4 Asymptotic channel QFI: HKS

When $H \in \mathcal{S}$, the situation is much more complicated because when $|\xi| = 1$ we must also have $|\dot{\xi}| = 0$ and no signal could be detected. Therefore we must consider the trade-off between maximizing the signal and minimizing the noise. To be exact, we want to maximize

$$\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\theta}) = \frac{|\dot{\xi}|^2}{1 - |\xi|^2}. \tag{4.46}$$

We will show for any $\eta > 0$, there exists a near-optimal code and recovery such that $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\theta}) > \mathfrak{F}^{(u)}_{\text{SQL}}(\mathcal{E}_\theta) - \eta$, proving Theorem 4.2. We only consider the case where $\mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta) > \mathfrak{F}_1(\mathcal{E}_\theta) > 0$ because otherwise $\mathfrak{F}_1(\mathcal{E}_\theta) = \mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta)$ and product states are sufficient to achieve $\mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta)$.

To simplify the calculation, we consider a special type of code, the perturbation code, as introduced in Section 3.5, where

$$A_0 = \sqrt{1 - \varepsilon^2} C + \varepsilon D, \quad A_1 = \sqrt{1 - \varepsilon^2} C - \varepsilon D, \tag{4.47}$$

satisfying $\operatorname{Tr}(C^\dagger D) = 0$ and $\operatorname{Tr}(C^\dagger C) = \operatorname{Tr}(D^\dagger D) = 1$. In this section, we define $\tilde{C} = CD^\dagger + DC^\dagger$ (differed by a factor of $\varepsilon\sqrt{1-\varepsilon^2}$ from the $\tilde{C}$ defined in Section 4.3) and also assume $C$ is full rank so that $\tilde{C}$ could be an arbitrary Hermitian matrix. $\varepsilon$ is a small parameter and we will calculate $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\theta})$ up to the lowest order of $\varepsilon$. We adopt the small $\varepsilon$ treatment because it allows us to mathematically simplify the optimization of Eq. (4.46), though it is surprising that the optimal QFI is achievable in such a regime where both the signal and the noise are small. Heuristically, it comes from an observation that sometimes the absolute strengths of the signal and the noise are not important—they could cancel each other out in the numerator and the denominator and only the ratio between them matters. See [Zhou and Jiang, 2020b, Appx. G] for an example.

To proceed, we use the vectorization of matrices $|\star\rangle\!\rangle = \sum_{ij} \star_{ij} |i\rangle |j\rangle$ for all $\star \in \mathbb{C}^{d\times d}$ to

simplify the notations. We define $E_{0,1}, E, F \in \mathbb{C}^{d^2 \times r}$ in the following way:

$$E_{0,1} = (|K_1 A_{0,1}\rangle\rangle \cdots |K_r A_{0,1}\rangle\rangle), \tag{4.48}$$

$$E = (|K_1 C\rangle\rangle \cdots |K_r C\rangle\rangle),\ F = (|K_1 D\rangle\rangle \cdots |K_r D\rangle\rangle), \tag{4.49}$$

which satisfy $E_{0,1} = \sqrt{1-\varepsilon^2} \pm \varepsilon F$, $\mathrm{Tr}(E^\dagger F) = 0$ and $\mathrm{Tr}(E^\dagger E) = \mathrm{Tr}(F^\dagger F) = 1$. Let the recovery matrix $T = QR^\dagger \in \mathbb{C}^{d^2 \times d^2}$, then

$$\xi = \mathrm{Tr}(T E_0 E_1^\dagger), \quad \dot{\xi} = \mathrm{Tr}(T \dot{E}_0 E_1^\dagger) + \mathrm{Tr}(T E_0 \dot{E}_1^\dagger). \tag{4.50}$$

### 4.4.1 Optimizing the recovery channel

We consider the regime where both the signal and the noise are sufficiently small—both the denominator and the numerator in Eq. (4.46) will be $O(\varepsilon^2)$. The recovery matrix $T$ should also be close to the identity operator. We assume $T = e^{i\varepsilon G}$ where $G$ is Hermitian and let $\sigma = EE^\dagger$, $\tilde{\sigma} = i(FE^\dagger - EF^\dagger)$. Expanding $T, E_0, E_1$ around $\varepsilon = 0$, we first optimize $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\theta})$ over all possible $G$, which gives (up to the lowest order of $\varepsilon$),

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\omega}) \approx \max_G \frac{|\mathrm{Tr}(G\dot{\sigma})|^2}{4 - 2\mathrm{Tr}(G\tilde{\sigma}) + \mathrm{Tr}(G^2\sigma) - |\mathrm{Tr}(G\sigma)|^2}, \tag{4.51}$$

The maximization could be calculated by taking the derivative w.r.t. $G$. We can show that the optimal $G$ is

$$G_{\mathrm{opt}} = \frac{(4 - \mathrm{Tr}(L_\sigma[\tilde{\sigma}]\tilde{\sigma}))}{\mathrm{Tr}(L_\sigma[\dot{\sigma}]\tilde{\sigma})} L_\sigma[\dot{\sigma}] + L_\sigma[\tilde{\sigma}], \tag{4.52}$$

and the corresponding optimal QFI is

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\theta}) \approx \mathrm{Tr}(L_\sigma[\dot{\sigma}]\dot{\sigma}) + \frac{\mathrm{Tr}(L_\sigma[\dot{\sigma}]\tilde{\sigma})^2}{4 - \mathrm{Tr}(L_\sigma[\tilde{\sigma}]\tilde{\sigma})}. \tag{4.53}$$

Detailed derivations can be found in Appendix E.

### 4.4.2 Attaining the upper bound

Now $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\theta})$ is a function of the code ($C$ and $D$) only. We can further simplify it by writing it as a function of only $C$ and $\tilde{C}$.

We express $\text{Tr}(L_\sigma[\dot{\sigma}]\dot{\sigma})$, $\text{Tr}(L_\sigma[\dot{\sigma}]\tilde{\sigma})$ and $\text{Tr}(L_\sigma[\tilde{\sigma}]\tilde{\sigma})$ in Eq. (4.53) in terms of $C$ and $\tilde{C}$. Let $\tau = E^\dagger E$, $\tilde{\tau} = E^\dagger F + F^\dagger E$, $\tau' = iE^\dagger \dot{E} - i\dot{E}^\dagger E$ such that

$$\tau_{ij} = \text{Tr}(C^\dagger K_i^\dagger K_j C),\ \tilde{\tau}_{ij} = \text{Tr}(\tilde{C} K_i^\dagger K_j), \tag{4.54}$$

$$\tau'_{ij} = i\text{Tr}(C^\dagger K_i^\dagger \dot{K}_j C) - i\text{Tr}(C^\dagger \dot{K}_i^\dagger K_j C). \tag{4.55}$$

WLOG, assume $\tau_{ij} = \text{Tr}(C^\dagger K_i^\dagger K_j C) = \lambda_i \delta_{ij}$, which could always be achieved by performing a unitary transformation on $\mathbf{K}$. We also have $\lambda_i > 0$ for all $i$ because $C$ is full rank and $\{|K_i\rangle\rangle\}_{i=1}^r$ are linearly independent.

From the detailed calculations in Appendix F, we see that

$$\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\theta}) \approx f(C,\tilde{C}) = 4\text{Tr}(C^\dagger \dot{\mathbf{K}}^\dagger \dot{\mathbf{K}} C) - \text{Tr}(L_\tau[\tau']\tau') + \frac{(-2\text{Tr}(\tilde{C}H) + \text{Tr}(L_\tau[\tau']\tilde{\tau}))^2}{\text{Tr}(L_\tau[\tilde{\tau}]\tilde{\tau})}. \tag{4.56}$$

At this stage, it is not obvious why the maximization of $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\theta})$ over $C$ and $\tilde{C}$ is equal to $\mathfrak{F}^{(u)}_{\text{SQL}}(\mathcal{E}_\theta)$. To see that, we need to reformulate the SQL upper bound using its dual program. First we note that

$$\mathfrak{F}^{(u)}_{\text{SQL}}(\mathcal{E}_\theta) = \max_{C:\text{Tr}(C^\dagger C)=1} \min_{h:\beta=0} 4\text{Tr}(C^\dagger \alpha C), \tag{4.57}$$

where we are allowed to exchange the order of maximization and minimization thanks to Sion's minimax theorem [Komiya, 1988; do Rosário Grossinho and Tersian, 2001]. Fixing $C$, we consider the optimization problem $\min_{h:\beta=0} 4\text{Tr}(C^\dagger \alpha C)$. When $C$ is full rank, we can show that it is equivalent to $\max_{\tilde{C}\in\mathbb{H}_d} f(C,\tilde{C})$, where $\tilde{C}$ is introduced as the Lagrange multiplier associated with the constraint $\beta = 0$ [Boyd and Vandenberghe, 2004] and the optimal $\tilde{C}$ is traceless.

To be specific, we first show

$$\max_{\tilde{C}\in\mathbb{H}_d} f(C,\tilde{C}) = \min_{h:\beta=0} 4\mathrm{Tr}(C^\dagger \alpha C) \tag{4.58}$$

when $C$ is full rank. $\tilde{C}$ is a Hermitian matrix as a Lagrange multiplier of $\beta = 0$. The Lagrange function is

$$L(\tilde{C},h) = 4\mathrm{Tr}(C^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})C) + \mathrm{Tr}(\tilde{C}(H + \mathbf{K}^\dagger h\mathbf{K})), \tag{4.59}$$

then

$$\begin{aligned}
&\min_h L(\tilde{C},h) = \min_h 4\mathrm{Tr}(C^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K})C) + \mathrm{Tr}(\tilde{C}(H + \mathbf{K}^\dagger h\mathbf{K})) \\
&= \min_h 4\mathrm{Tr}(C^\dagger\dot{\mathbf{K}}^\dagger\dot{\mathbf{K}}C) + 4\mathrm{Tr}(\tau h^2) + 4\mathrm{Tr}(iC^\dagger\mathbf{K}^\dagger h\dot{\mathbf{K}}C - iC^\dagger\dot{\mathbf{K}}^\dagger h\mathbf{K}C) + \mathrm{Tr}(\tilde{C}(H + \mathbf{K}^\dagger h\mathbf{K})) \\
&= \min_h 4\mathrm{Tr}(C^\dagger\dot{\mathbf{K}}^\dagger\dot{\mathbf{K}}C) + 4\mathrm{Tr}(\tau h^2) + 4\mathrm{Tr}(h^T\tau') + \mathrm{Tr}(\tilde{C}H) + \mathrm{Tr}(h^T\tilde{\tau}) \\
&= 4\mathrm{Tr}(C^\dagger\dot{\mathbf{K}}^\dagger\dot{\mathbf{K}}C) + \mathrm{Tr}(\tilde{C}H) - \frac{1}{8}\sum_{i,j=1}^r \frac{|4\tau'_{ij} + \tilde{\tau}_{ij}|^2}{\lambda_i + \lambda_j}.
\end{aligned} \tag{4.60}$$

The dual program is

$$\begin{aligned}
&\max_{\tilde{C}}\min_h L(\tilde{C},h) = \max_{\tilde{C}} 4\mathrm{Tr}(C^\dagger\dot{\mathbf{K}}^\dagger\dot{\mathbf{K}}C) + \mathrm{Tr}(\tilde{C}H) - \frac{1}{8}\sum_{i,j=1}^r \frac{16|\tau'_{ij}|^2 + |\tilde{\tau}_{ij}|^2 + 4(\tilde{\tau}_{ij}\tau'_{ji} + \tilde{\tau}_{ji}\tau'_{ij})}{\lambda_i + \lambda_j} \\
&= \max_{\tilde{C},x} 4\mathrm{Tr}(C^\dagger\dot{\mathbf{K}}^\dagger\dot{\mathbf{K}}C) + x\mathrm{Tr}(\tilde{C}H) - \frac{1}{8}\sum_{i,j=1}^r \frac{16|\tau'_{ij}|^2 + x^2|\tilde{\tau}_{ij}|^2 + 8x\tilde{\tau}_{ij}\tau'_{ji}}{\lambda_i + \lambda_j} \\
&= \max_{\tilde{C}} 4\mathrm{Tr}(C^\dagger\dot{\mathbf{K}}^\dagger\dot{\mathbf{K}}C) - 2\sum_{i,j=1}^r \frac{|\tau'_{ij}|^2}{\lambda_i + \lambda_j} + \frac{\left(-\mathrm{Tr}(\tilde{C}H) + \sum_{i,j=1}^r \frac{\tilde{\tau}_{ij}\tau'_{ji}}{\lambda_i+\lambda_j}\right)^2}{\frac{1}{2}\sum_{i,j=1}^r \frac{|\tilde{\tau}_{ij}|^2}{\lambda_i+\lambda_j}} = \max_{\tilde{C}} f(C,\tilde{C}),
\end{aligned} \tag{4.61}$$

where we used the fact that $\tilde{C} \leftarrow x\tilde{C}$ does not change the result. Eq. (4.58) is then proved.

Moreover, the optimal $\tilde{C}$ in Eq. (4.58) must be traceless. Suppose $\tilde{C}$ is optimal in Eq. (4.58), we will prove that $\mathrm{Tr}(\tilde{C}) = 0$. Let $z$ be a real number,

$$q(z) := f(C,\tilde{C} + zCC^\dagger) = \frac{s(z)^2}{t(z)} + \text{const.} \tag{4.62}$$

Since $\max_z q(z) = q(0)$, we have $q'(0) = \frac{s(0)}{t(0)^2}(2s'(0)t(0) - s(0)t'(0)) = 0$.

$$s(z) = -\mathrm{Tr}((\tilde{C} + zCC^\dagger)H) + \sum_{i,j=1}^{r} \frac{(\tilde{\tau}_{ij} + z\lambda_i\delta_{ij})\tau'_{ij}}{\lambda_i + \lambda_j}, \tag{4.63}$$

$$s'(0) = -\mathrm{Tr}(CC^\dagger H) + \sum_{i=1}^{r} \frac{1}{2}\tau'_{ii} = 0, \tag{4.64}$$

$$t(z) = \frac{1}{2}\sum_{i,j=1}^{r} \frac{|\tilde{\tau}_{ij} + z\lambda_i\delta_{ij}|^2}{\lambda_i + \lambda_j} = \frac{1}{2}\sum_{i,j=1}^{r} \frac{|\tilde{\tau}_{ij}|^2 + z\lambda_i\delta_{ij}(\tilde{\tau}^*_{ij} + \tilde{\tau}_{ij}) + z^2\lambda_i^2\delta_{ij}}{\lambda_i + \lambda_j}, \tag{4.65}$$

$$t'(0) = \frac{1}{2}\sum_{i,j=1}^{r} \frac{\lambda_i\delta_{ij}(\tilde{\tau}^*_{ij} + \tilde{\tau}_{ij})}{\lambda_i + \lambda_j} = \frac{1}{2}\sum_{i=1}^{r} \tilde{\tau}_{ii} = \frac{1}{2}\mathrm{Tr}(\tilde{C}). \tag{4.66}$$

Then $q'(0) = 0$ implies $\mathrm{Tr}(\tilde{C}) = 0$.

### 4.4.3 Efficient numerical algorithm

The procedure to find a near-optimal code such that $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\theta}) > \mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\theta) - \eta$ for any $\eta > 0$ goes as follows:

(1) Find a full rank $C^\diamond$ such that $\mathrm{Tr}(C^{\diamond\dagger}C^\diamond) = 1$ and $\min_{h:\beta=0} 4\mathrm{Tr}(C^{\diamond\dagger}\alpha C^\diamond) > \mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\theta) - \eta/2$.

(2) Find a Hermitian $\tilde{C}^\diamond$ such that $f(C^\diamond, \tilde{C})$ is maximized and let $D^\diamond = \frac{1}{2}C^{\diamond -1}\tilde{C}^\diamond$. Rescale $D^\diamond$ such that $\mathrm{Tr}(D^{\diamond\dagger}D^\diamond) = 1$.

(3) Calculate $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\theta})|_{C=C^\diamond, D=D^\diamond}$ using Eqs. (4.47)-(4.50) and Eq. (4.52). Find a small $\varepsilon^\diamond > 0$ such that $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\theta}) > f(C^\diamond, \tilde{C}^\diamond) - \eta/2$.

The numerical algorithms for step (1) and (2) are provided below, where the most computationally intensive part is a SDP. Note that in contrast to the HL case, here we require $2d$-dimensional ancillas, twice as large as probes. In principle, however, $d$-dimensional ancillas are sufficient to achieve the asymptotic QFI, considering the Schmidt decomposition on the input state, though we no longer have explicit encoding and decoding protocols when using $d$-dimensional ancillas.

**Finding the optimal $C$**

We first describe a numerical algorithm finding a full rank $C^\diamond$ such that $\mathrm{Tr}(C^{\diamond\dagger}C^\diamond) = 1$ and

$$\min_{h:\beta=0} 4\mathrm{Tr}(C^{\diamond\dagger}\alpha C^\diamond) > \mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\theta) - \eta/2. \tag{4.67}$$

for any $\eta > 0$. We first note that $\mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\theta) = \min_{h:\beta=0} 4\left\|\alpha\right\|$ could be solved via the following SDP [Demkowicz-Dobrzański et al., 2012],

$$\min_h x, \quad \text{subject to} \begin{pmatrix} xI_d & \tilde{K}_1^\dagger & \cdots & \tilde{K}_r^\dagger \\ \tilde{K}_1 & I_{d'} & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ \tilde{K}_r & 0 & \cdots & I_{d'} \end{pmatrix} \succeq 0, \quad \beta = 0. \tag{4.68}$$

where $d$ and $d'$ are the input and output dimension of $\mathcal{E}_\theta$, $I_n$ is a $n \times n$ identity matrix and $\tilde{\mathbf{K}} = \dot{\mathbf{K}} - ih\mathbf{K}$.

To find the full rank $C^\diamond$, we first find a density matrix $\rho^\diamond$ such that

$$\min_{h:\beta=0} 4\mathrm{Tr}(\rho^\diamond\alpha) = \min_{h:\beta=0} 4\left\|\alpha\right\|. \tag{4.69}$$

It could be done via the following two-step algorithm-4:

1) Find an $h^\diamond$ using the SDP (Eq. (4.68)), such that $\alpha^\diamond = \alpha|_{h=h^\diamond}$ satisfies $\|\alpha^\diamond\| = \min_{h:\beta=0}\|\alpha\|$.

2) Let $\Pi^\diamond$ be the projection onto the subspace spanned by all eigenstates corresponding to the largest eigenvalue of $\alpha^\diamond$, we find an optimal density matrix $\rho^\diamond$ satisfying $\Pi^\diamond\rho^\diamond\Pi^\diamond = \rho^\diamond$ and

$$\mathrm{Re}[\mathrm{Tr}(\rho^\diamond(i\mathbf{K}^\dagger\delta h)(\dot{\mathbf{K}} - ih^\diamond\mathbf{K}))] = 0, \quad \forall\delta h \in \mathbb{H}_r, \text{ s.t. } \mathbf{K}^\dagger\delta h\mathbf{K} = 0. \tag{4.70}$$

Then $C^\diamond = \left((1-\eta')\rho^\diamond + \eta'\frac{I}{d}\right)^{1/2}$ where $\eta' = \eta/(2\mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\theta))$ is a full-rank matrix satisfying

$$\min_{h:\beta=0} 4\mathrm{Tr}(C^{\diamond\dagger}\alpha C^\diamond) \geq (1-\eta')\mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\theta) = \mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\theta) - \eta/2. \tag{4.71}$$

The two-step algorithm above could also be used to find $\rho^\blacklozenge$ whose purification is the optimal input state of a single quantum channel $\mathcal{E}_\theta$ achieving $\mathfrak{F}_1(\mathcal{E}_\theta)$:

1) Find an $h^\blacklozenge$ using the SDP in Eq. (4.68) without the requirement $\beta = 0$, such that $\alpha^\blacklozenge = \alpha|_{h=h^\blacklozenge}$ satisfies $\|\alpha^\blacklozenge\| = \min_h \|\alpha\|$.

2) Let $\Pi^\blacklozenge$ be the projection onto the subspace spanned by all eigenstates corresponding to the largest eigenvalue of $\alpha^\blacklozenge$, we find an optimal density matrix $\rho^\blacklozenge$ satisfying $\Pi^\blacklozenge\rho^\blacklozenge\Pi^\blacklozenge = \rho^\blacklozenge$ and

$$\mathrm{Re}[\mathrm{Tr}(\rho^\blacklozenge(i\mathbf{K}^\dagger\delta h)(\dot{\mathbf{K}} - ih^\blacklozenge\mathbf{K}))] = 0, \quad \forall\delta h \in \mathbb{H}_r. \tag{4.72}$$

**Validity of the algorithm to find the optimal $C$**

For completeness, we prove the validity of the above two-step algorithm. According to Sion's minimax theorem [Komiya, 1988; do Rosário Grossinho and Tersian, 2001], for convex compact sets $\mathfrak{P} \subset \mathbb{R}^m$ and $\mathfrak{Q} \subset \mathbb{R}^n$ and $g : P \times Q \to \mathbb{R}$ such that $g(x, y)$ is a continuous convex (concave) function in $x$ ($y$) for every fixed $y$ ($x$), then

$$\max_{y\in\mathfrak{Q}} \min_{x\in\mathfrak{P}} g(x, y) = \min_{x\in\mathfrak{P}} \max_{y\in\mathfrak{Q}} g(x, y). \tag{4.73}$$

In particular, if $(x^\blacktriangle, y^\blacklozenge)$ is a solution of $\max_{y\in\mathfrak{Q}} \min_{x\in\mathfrak{P}} g(x, y)$, then there must exists an $x^\blacklozenge$ such that $(x^\blacklozenge, y^\blacklozenge)$ is a saddle point. Let $(x^\blacklozenge, y^\blacktriangle)$ be a solution of $\min_{x\in\mathfrak{P}} \max_{y\in\mathfrak{Q}} g(x, y)$. Then we must have

$$g(x^\blacktriangle, y^\blacklozenge) \le g(x^\blacklozenge, y^\blacklozenge) \le g(x^\blacklozenge, y^\blacktriangle). \tag{4.74}$$

According to Eq. (4.73), $g(x^\blacktriangle, y^\blacklozenge) = g(x^\blacklozenge, y^\blacktriangle)$ and all equalities must hold for the above equation. Moreover,

$$g(x^\blacklozenge, y) \le g(x^\blacklozenge, y^\blacklozenge) \le g(x, y^\blacklozenge), \quad \forall(x, y) \in \mathfrak{P} \times \mathfrak{Q}, \tag{4.75}$$

which means $(x^\blacklozenge, y^\blacklozenge)$ is a saddle point. For example, we can take $x = h \in \mathbb{H}_r$, $y = CC^\dagger = \rho \in \mathscr{S}(\mathcal{H}_S)$ and $g(x, y) = 4\mathrm{Tr}(\rho\alpha)$. (We can also add the constraint $\beta = 0$ on $h$ which does not affect our discussion below). Then the solution of the above optimization problem is

$\mathfrak{F}_1(\mathcal{E}_\theta)$ (or $\mathfrak{F}_{\text{SQL}}(\mathcal{E}_\theta)$ with the constraint $\beta = 0$). Note that we can always confine $h$ in a compact set such that the solutions are not altered and the minimax theorem is applicable. Let $(h^\blacktriangle, \rho^\blacklozenge)$ be any solution of the optimization problem $\max_\rho \min_h 4\text{Tr}(\rho\alpha)$. Then there exists an $h^\blacklozenge$ such that $(h^\blacklozenge, \rho^\blacklozenge)$ is a saddle point. Similarly, if $g(x^\blacklozenge, y^\blacktriangle)$ is a solution of $\min_{x\in\mathfrak{P}} \max_{y\in\mathfrak{Q}} g(x,y)$, which in our case is a SDP (Eq. (4.68)). There must exists a $y^\blacklozenge$ such that $(x^\blacklozenge, y^\blacklozenge)$ is a saddle point. Let $(h^\blacklozenge, \rho^\blacktriangle)$ be any solution of the optimization problem $\min_h \max_\rho 4\text{Tr}(\rho\alpha)$. Then there exists an $\rho^\blacklozenge$ such that $(h^\blacklozenge, \rho^\blacklozenge)$ is a saddle point. Moreover, $(h^\blacklozenge, \rho^\blacklozenge)$ is a saddle point if and only if

(i) $\text{Tr}(\rho^\blacklozenge\alpha^\blacklozenge) = \|\alpha^\blacklozenge\|, \;\Leftrightarrow\; \text{Tr}(\rho^\blacklozenge\alpha^\blacklozenge) \geq \text{Tr}(\rho\alpha^\blacklozenge),\ \forall\rho.$

(ii) $\text{Re}[\text{Tr}(\rho^\blacklozenge(i\mathbf{K}^\dagger\delta h)(\dot{\mathbf{K}} - ih^\blacklozenge\mathbf{K}))] = 0,\ \forall\delta h\in\mathbb{H}_r, \;\Leftrightarrow\; \text{Tr}(\rho^\blacklozenge\alpha^\blacklozenge) \leq \text{Tr}(\rho^\blacklozenge\alpha),\ \forall h.$

It justifies the validity of the two-step algorithm we described above.

**Finding the optimal $\tilde{C}$**

Next, we describe how to find $\tilde{C}^\diamond$ such that $f(C^\diamond, \tilde{C}^\diamond) = \max_{\tilde{C}} f(C^\diamond, \tilde{C}) = \min_{h:\beta=0} 4\text{Tr}(C^{\diamond\dagger}\alpha C^\diamond)$. According to Eq. (4.56),

$$f(C,\tilde{C}) = 4\text{Tr}(C^\dagger\dot{\mathbf{K}}^\dagger\dot{\mathbf{K}}C) - 2\sum_{i,j=1}^{r}\frac{|\tau'_{ij}|^2}{\lambda_i+\lambda_j} + \frac{\left(-\text{Tr}(\tilde{C}H) + \sum_{i,j=1}^r \frac{\tilde{\tau}_{ij}\tau'_{ji}}{\lambda_i+\lambda_j}\right)^2}{\frac{1}{2}\sum_{i,j=1}^r \frac{|\tilde{\tau}_{ij}|^2}{\lambda_i+\lambda_j}}, \tag{4.76}$$

where we have assumed $\tau_{ij} = \text{Tr}(C^\dagger K_i^\dagger K_j C) = \lambda_i\delta_{ij}$. For a fixed $C$, $\tilde{\tau}$ is a linear function in $\tilde{C}$. We could always write

$$f(C,\tilde{C}) = f_1(C) + \frac{|\langle\!\langle\tilde{C}|f_2(C)\rangle\!\rangle|^2}{\langle\!\langle\tilde{C}|f_3(C)|\tilde{C}\rangle\!\rangle}, \tag{4.77}$$

where $f_1(C)\in\mathbb{R}$, $f_2(C)\in\mathbb{C}^{d\times d}$ is Hermitian and $f_3(C)\in\mathbb{C}^{d^2\times d^2}$ is positive semidefinite. Moreover, $|f_2(C)\rangle\!\rangle$ is in the support of $f_3(C)$. $f_{1,2,3}(C)$ are functions of $C$ only. According to Cauchy-Schwarz inequality,

$$\max_{\tilde{C}} f(C,\tilde{C}) = f_1(C) + \langle\!\langle f_2(C)|f_3(C)^{-1}|f_2(C)\rangle\!\rangle, \tag{4.78}$$

where the maximum is attained when $|\tilde{C}\rangle\rangle = f_3(C)^{-1}|f_2(C)\rangle\rangle$ and $^{-1}$ here means the Moore-Penrose pseudoinverse. Therefore, we take

$$|\tilde{C}^{\diamond}\rangle\rangle = f_3(C^{\diamond})^{-1}|f_2(C^{\diamond})\rangle\rangle. \tag{4.79}$$

To conclude, we proposed a perturbation code which could achieve the SQL upper bound with an arbitrarily small error. We take the limit where the parameter $\varepsilon$ which distinguishes the logical zero and one states is sufficiently small. Note that if we take $\varepsilon = 0$, the probe state will be a product state and we can only achieve $\mathfrak{F}_1(\mathcal{D}_{L,\theta})$. This discontinuity appears because we must first take the limit $N \to \infty$ before taking the limit $\varepsilon \to 0$ and the impact of a small $\varepsilon$ becomes significant in the asymptotic limit.

## 4.5 Examples

In this section, we provide three applications of our theorems. We first compute the asymptotic QFI of single-qubit depolarizing channels, which were not fully explored before. It is a case where $\mathfrak{F}_{\text{SQL}} > \mathfrak{F}_1$ whenever the HNKS condition is violated. Secondly, we consider amplitude damping channels and obtain an analytical solution of a near-optimal QEC protocol. We will directly see how the gap between the attainable QFI and $\mathfrak{F}_{\text{SQL}}$ shrinks when $\varepsilon$ approaches 0. In the third example, we consider a special type of channel which always satisfies $\mathfrak{F}_{\text{SQL}} = \mathfrak{F}_1$ and provide a new simple proof of it.

### 4.5.1 Qubit depolarizing channels

Here we calculate $\mathfrak{F}_1$, $\mathfrak{F}_{\text{SQL}}$ and $\mathfrak{F}_{\text{HL}}$ for depolarizing channels $\mathcal{N}_\theta(\rho) = \mathcal{N}(\mathcal{U}_\theta(\rho))$ where

$$\mathcal{N}(\rho) = (1-p)\rho + p_x\sigma_x\rho\sigma_x + p_y\sigma_y\rho\sigma_y + p_z\sigma_z\rho\sigma_z, \tag{4.80}$$

$p_{x,y,z} \geq 0$, $p = p_x + p_y + p_z < 1$ and $\mathcal{U}_\theta(\cdot) = e^{-\frac{i\theta}{2}\sigma_z}(\cdot)e^{\frac{i\theta}{2}\sigma_z}$.

First, we notice that HNKS is satisfied if and only if $p_x = p_z = 0$ or $p_y = p_z = 0$. When HNKS is satisfied, $\mathfrak{F}_{\text{HL}}(\mathcal{N}_\theta) = 1$. It is the same as the $\mathfrak{F}_{\text{HL}}$ when there is no noise ($p = 0$) because the Kraus operator ($\sigma_x$ or $\sigma_y$) is perpendicular to the Hamiltonian ($\sigma_z$) and

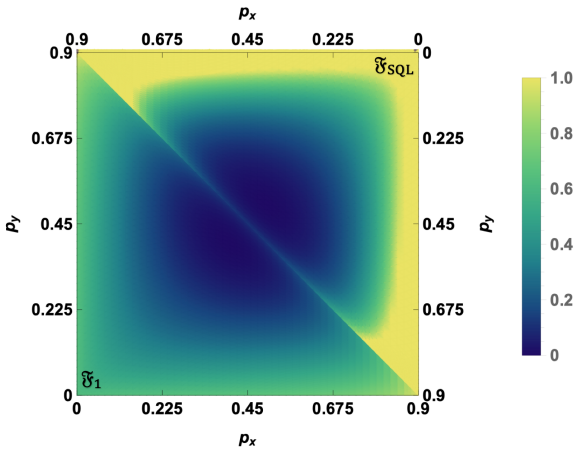


Figure 4.2: Plots of $\mathfrak{F}_1(\mathcal{N}_\theta)$ and $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta)$ as functions of $p_x$ and $p_y$ when $p_z = 0.1$. The lower left and upper right part are the plots of $\mathfrak{F}_1(\mathcal{N}_\theta)$ and $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta)$ respectively.

could be fully corrected. It is consistent with previous results for single-qubit Hamiltonian estimation that the HL is achievable if and only if the Markovian noise is rank-one and not parallel to the Hamiltonian [Kessler et al., 2014; Arrad et al., 2014; Dür et al., 2014; Ozeri, 2013; Unden et al., 2016; Reiter et al., 2017; Sekatski et al., 2017].

As calculated in Appendix G,

$$\mathfrak{F}_1(\mathcal{N}_\theta) = 1 - w, \tag{4.81}$$

where $w = 4\left(\frac{p_x p_y}{p_x+p_y} + \frac{(1-p)p_z}{1-p+p_z}\right) \leq 1$. When HNKS is violated,

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta) = (1-w)/w. \tag{4.82}$$

In the equations above, when $p_x = p_y = 0$, we take $\frac{p_x p_y}{p_x+p_y} = 0$, in which case $\mathcal{N}_\theta$ becomes the dephasing channel introduced in Section 4.2 where $\phi = \theta$ and $p$ is independent of $\theta$.

We observe that

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta) = \mathfrak{F}_1(\mathcal{N}_\theta)/w \geq \mathfrak{F}_1(\mathcal{N}_\theta), \tag{4.83}$$

and the equality ($w = 1$) holds if and only if $p_x = p_y$ and $p_z + p_x = 1/2$, in which case $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta) = \mathfrak{F}_1(\mathcal{N}_\theta) = 0$ and $\mathcal{N}_\theta = \mathcal{N}$ becomes a mixture of a completely dephasing channel and a completely depolarizing channel [Watrous, 2018] where $\theta$ cannot be detected.

$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta)$ is in general non-additive. In particular, when $p \ll 1$, we have $w \ll 1$

and $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\theta) \gg \mathfrak{F}_1(\mathcal{N}_\theta)$. We also illustrate the difference between $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\theta)$ and $\mathfrak{F}_1(\mathcal{N}_\theta)$ in Figure 4.2 by plotting $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\theta)$, $\mathfrak{F}_1(\mathcal{N}_\theta)$ as a function of $p_x$ and $p_y$ when $p_z = 0.1$. $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\theta) = \mathfrak{F}_1(\mathcal{N}_\theta) = 0$ at $(p_x, p_y, p_z) = (0.4, 0.4, 0.1)$. The ratio $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\theta)/\mathfrak{F}_1(\mathcal{N}_\theta)$ increases near the boundary of $p_x + p_y < 0.9$.

We remark here that when the dimension of the system is large, for example, a qudit depolarizing channel or a collective dephasing channel [Dorner, 2012], although $\mathfrak{F}_1$, $\mathfrak{F}_{\text{SQL}}$ and the optimal input states can be found numerically via SDPs, analytical solutions may not exist. In that case, it might be helpful to compute analytical upper bounds on the QFI (Section 5.3.3) or consider the limit of large ensembles and use variational methods to solve for the QFI [Knysh et al., 2014].

### 4.5.2 Amplitude damping channels

In the first example, we focus on computing the asymptotic QFIs for single-qubit depolarizing channels, but we do not provide explicit QEC protocols achieving the QFIs. Here we present a second example, where we obtain an analytical solution of the optimal QEC protocol and also analyze its performance when $\varepsilon$ is not a small constant.

Here we consider amplitude damping channels $\mathcal{N}_\omega^{\text{ad}}(\rho) = \mathcal{N}^{\text{ad}}(\mathcal{U}_\omega(\rho))$ defined by

$$\mathcal{N}^{\text{ad}}(\rho) = K_1^{\text{ad}} \rho K_1^{\text{ad}\dagger} + K_2^{\text{ad}} \rho K_2^{\text{ad}\dagger}, \tag{4.84}$$

where $K_1^{\text{ad}} = |0\rangle\langle 0| + \sqrt{1-p}\,|1\rangle\langle 1|$ and $K_2^{\text{ad}} = \sqrt{p}\,|0\rangle\langle 1|$ and $p$ represents the probability of a particle switching from $|1\rangle$ to $|0\rangle$ which is independent of $\omega$. $\mathcal{U}_\omega$ again is the Pauli-Z rotation $e^{-i\frac{\omega}{2}\sigma_z}$. We will assume $\omega = 0$ in this section for simplicity, because for non-zero $\omega$, the QFI is the same and we only need to rotate the code accordingly.

As before, amplitude damping channels follow the SQL as long as $p > 0$. Thus, we shall only focus on the situation where HNKS is violated. As shown in Appendix H, $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}}) = 4(1-p)/p$ [Demkowicz-Dobrzański et al., 2012] and the near-optimal QEC protocol can be obtained using our algorithm from Section 4.4.3. The QEC code is characterized by two small but non-zero constants $\delta$ and $\varepsilon$, where $\delta$ is to make sure $C^\diamond$ is full rank, originated from step (1) in our algorithm from Section 4.4.3 and $\varepsilon = o(\delta)$ is the small constant in the

perturbation code:

$$|\mathfrak{c}_0\rangle = \sin(\delta+\varepsilon)\,|0\rangle_S|00\rangle_A + \cos(\delta+\varepsilon)\,|1\rangle_S|10\rangle_A\,, \tag{4.85}$$

$$|\mathfrak{c}_1\rangle = \sin(\delta-\varepsilon)\,|0\rangle_S\,|01\rangle_A + \cos(\delta-\varepsilon)\,|1\rangle_S|11\rangle_A\,. \tag{4.86}$$

Note that we use trigonometric functions instead of $\varepsilon$ and $\sqrt{1-\varepsilon^2}$ as before just to simplify the notations. We also need the optimal recovery channel which is determined by

$$G_{\text{opt}} = \frac{2i}{\sqrt{1-p}}\,|00\rangle\langle 11| + \frac{-2i}{\sqrt{1-p}}\,|11\rangle\langle 00|\,. \tag{4.87}$$

The asymptotic channel QFI $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\omega})$ attainable using the QEC protocol above is $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\omega}) = |\dot{\xi}|^2/(1-|\xi|^2)$, where

$$\xi = 1 - \frac{2p\sin^2(\delta)}{1-p}\varepsilon^2 + O(\varepsilon^4), \tag{4.88}$$

$$\dot{\xi} = -2i\sin(2\delta)\varepsilon + O(\varepsilon^3), \tag{4.89}$$

and

$$\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\omega}) = \frac{4(1-p)\cos^2(\delta)}{p} + O(\varepsilon^2). \tag{4.90}$$

which approaches $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}})$ for small $\delta$. Note, however, that we cannot take $\delta = 0$ because then $\dot{\xi} = 0$. It means $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}})$ is achievable with an arbitrarily small but non-zero error. The exact values of $\xi$ and $\dot{\xi}$ as a function of $\delta$ and $\varepsilon$ can be found in Appendix H.

To visualize the gap between $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\omega})$ and $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}})$, we plot it in Figure 4.3. We take $p = 0.5, 0.001$ in Figure 4.3(a) and Figure 4.3(b), and $\varepsilon = 0.9\delta, 0.5\delta, 0.1\delta$, $\varepsilon \to 0$ in each figure, and plot the ratio between the attainable QFI $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\omega})$ and the optimal QFI $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}})$ as a function of $\delta$. Figure 4.3(a) and Figure 4.3(b) are almost identical, showing the robustness of our code against the change in noise rates. We also see that the curve from $\varepsilon = 0.1\delta$ almost overlaps with the limiting one (Eq. (4.90)) as $\varepsilon \to 0$. Moreover, we compare our ancilla-assisted QEC protocol with ancilla-free ones which achieve at most $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}})/4$ [Knysh et al., 2014; Demkowicz-Dobrzański and Maccone, 2014]. It outperforms the optimal ancilla-free ones in a large range (of $\delta$ and $\varepsilon$), showing the power of noise-

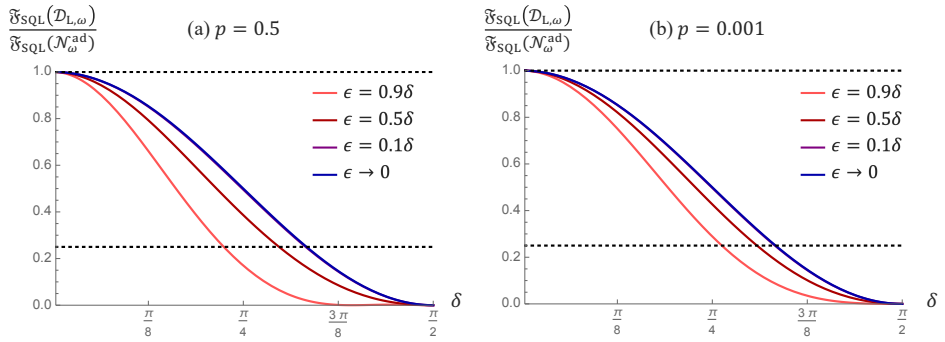


Figure 4.3: Plots of $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\omega})/\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\omega^{\mathrm{ad}})$ as a function of $\delta$. We take $p = 0.5, 0.001$ in (a) and (b) and $\varepsilon = 0.9\delta, 0.5\delta, 0.1\delta$, $\varepsilon \to 0$ in each figure. The curves from $\varepsilon = 0.1\delta$ and $\varepsilon \to 0$ are almost indistinguishable from each other. The dashed lines are at 1 and 1/4 where the former represents the upper bound $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\omega^{\mathrm{ad}})$ and the latter represents the optimal asymptotic QFI without the assistance of ancillas $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\omega^{\mathrm{ad}})/4$ [Knysh et al., 2014; Demkowicz-Dobrzański and Maccone, 2014]. Our QEC protocol outperforms the ancilla-free protocols even for large $\delta$ and $\varepsilon$.

less ancillas in phase estimation under amplitude damping noise. This type of phenomenon does not occur in dephasing channels where ancilla-free protocols are optimal [Ulam-Orgikh and Kitagawa, 2001; Knysh et al., 2014; Demkowicz-Dobrzański and Maccone, 2014].

### 4.5.3 $\mathbb{U}$-covariant channels

Let $\mathbb{U} = \{U_i\}_{i=1}^n \subset \mathbb{C}^{d\times d}$ be a set of unitary operators such that for some probability distribution $\{p_i\}_{i=1}^n$, $\{(p_i, U_i)\}_{i=1}^n$ is a unitary 1-design [Dankert, 2005], satisfying

$$\sum_{i=1}^n p_i U_i A U_i^\dagger = \mathrm{Tr}(A)\frac{I}{d},\ \forall A \in \mathbb{C}^{d\times d}. \tag{4.91}$$

For example, when $\mathbb{U}$ is a unitary orthonormal basis of $\mathbb{C}^{d\times d}$, $\{(\frac{1}{d^2}, U_i)\}_{i=1}^{d^2}$ is a unitary 1-design. In general, a unitary $t$-design is a set of unitary operators whose first $t$ moments are indistinguishable from the Haar random unitaries. Given a quantum channel $\mathcal{T}_\theta(\cdot) = \sum_{i=1}^r K_i(\cdot)K_i^\dagger$, we call it $\mathbb{U}$-covariant if for all $U \in \mathbb{U}$, there is a unitary $V$ such that

$$\mathcal{T}_\theta(U\rho U^\dagger) = V\mathcal{T}_\theta(\rho)V^\dagger. \tag{4.92}$$

Note that here it is important that $U$ and $V$ are independent of $\theta$, a feature called joint covariance [Laurenza et al., 2018]. It could be shown that $\mathfrak{F}_1(\mathcal{T}_\theta) = \mathfrak{F}_{\mathrm{SQL}}(\mathcal{T}_\theta)$ when $\mathcal{T}_\theta$ is $\mathbb{U}$-covariant, using the teleportation simulation technique [Pirandola and Lupo, 2017;

Pirandola et al., 2017; Chiribella et al., 2009; Wilde et al., 2017]. Here we provide an alternative proof using only the definitions of $\mathfrak{F}_1$ and $\mathfrak{F}_{\text{SQL}}$ in the minimax formulation.

Let $h^\blacklozenge$ be a solution of $\min_h \max_\rho 4\text{Tr}(\rho\alpha)$. As explained in Section 4.4.3, for every $\rho^\blacklozenge$ which is a solution of $\max_\rho \min_h 4\text{Tr}(\rho\alpha)$, $(h^\blacklozenge, \rho^\blacklozenge)$ is a saddle point, i.e.

$$4\text{Tr}(\rho\alpha^\blacklozenge) \leq 4\text{Tr}(\rho^\blacklozenge\alpha^\blacklozenge) \leq 4\text{Tr}(\rho^\blacklozenge\alpha), \tag{4.93}$$

for all $\rho$ and $h$, where $\alpha^\blacklozenge = \alpha|_{h=h^\blacklozenge}$. Then $|C^\blacklozenge\rangle\!\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A$ is an optimal input state of a single quantum channel $\mathcal{T}_\theta$, if and only if $\rho^\blacklozenge = C^\blacklozenge C^{\blacklozenge\dagger}$ satisfies Eq. (4.93). According to Eq. (4.92), if $|C^\blacklozenge\rangle\!\rangle$ is an optimal input, $|UC^\blacklozenge\rangle\!\rangle = (U \otimes I)|C^\blacklozenge\rangle\!\rangle$ is also an optimal input for all $U \in \mathbb{U}$ and satisfies Eq. (4.93). Then $\sum_{i=1}^n p_i U_i \rho^\blacklozenge U_i^\dagger = \frac{I}{d}$ also satisfies Eq. (4.93), implying the maximally entangled state $|\frac{I}{d}\rangle\!\rangle$ is an optimal input for $\mathcal{T}_\theta$. The discussion above also works for $\mathcal{T}_\theta^{\otimes N}$ because $\mathcal{T}_\theta^{\otimes N}$ is $\mathbb{U}^{\otimes N}$-covariant and $\{(\Pi_k p_{i_k}, \otimes_k U_{i_k})\}$ is a unitary 1-design on $\mathbb{C}^{Nd\times Nd}$. Therefore $|\frac{I}{d^N}\rangle\!\rangle$ is an optimal input for $\mathcal{T}_\theta^{\otimes N}$, which implies $\mathfrak{F}_N(\mathcal{T}_\theta) = N\mathfrak{F}_1(\mathcal{T}_\theta)$.

### 4.5.4 Mach-Zehnder interferometer

Here we consider a two-arm Mach-Zehnder interferometer with one noisy arm and one noiseless arm, where the input state is an $M$-photon state

$$|\psi_0\rangle = \sum_{m=0}^{M} \gamma_m |m\rangle |M-m\rangle . \tag{4.94}$$

Here $|m, M-m\rangle$ represents a two-mode Fock state where $m$ is the number of photons in the first mode and $M-m$ is the number of photons in the second. The noisy quantum channel $\mathcal{M}_\theta(\cdot)$ acting on the first mode is described by the Kraus operators

$$K_i = \sqrt{\frac{(1-\eta)^i}{i!}} e^{-i\theta a^\dagger a} \eta^{\frac{1}{2}a^\dagger a} a^i, \quad i = 0, 1, \ldots, M, \tag{4.95}$$

where $\theta$ is the unknown phase to be estimated, $a$ is the photon annihilation operator, $0 < \eta < 1$ is the loss rate and $K_i$ is associated with losing $i$ photons. Note that we are allowed to truncate the maximum photon number at $M$ because of the restriction on the input state (Eq. (4.94)).

We will show that the algorithm described in Section 4.4.3 naturally gives a SDP solving the optimal $\{\gamma_m\}_{m=0}^M$. We emphasize here that it was already shown in [Demkowicz-Dobrzański et al., 2009] that solving the optimal input state in an interferometer with two noisy arms is a convex optimization problem. Here we provide an alternative algorithm as a demonstration of our approach which also contains a proof that states of the form Eq. (4.94) are optimal for $\mathcal{M}_\theta$.

Recall that given $\mathcal{M}_\theta$, we can find an optimal input state achieving $\mathfrak{F}_1(\mathcal{M}_\theta)$ by purifying $\rho^\blacklozenge$ which is a solution of

$$\mathfrak{F}_1(\mathcal{M}_\theta) = \max_\rho \min_h 4\mathrm{Tr}(\rho\alpha) = \min_h 4\left\|\alpha\right\|. \tag{4.96}$$

We show that the optimization problem above has a diagonal solution of $\rho$. Note that

$$\begin{aligned}\alpha &= \sum_{i=0}^M \left(\dot{K}_i - i\sum_{i,j=0}^M h_{ij}K_j\right)^\dagger \left(\dot{K}_i - i\sum_{j'=0}^M h_{ij'}K_{j'}\right) \\ &= \sum_{i=0}^M \left(\dot{K}_i - ih_{ii}K_i\right)^\dagger \left(\dot{K}_i - ih_{ii}K_i\right) + \sum_{i=0}^M\sum_{j\neq i} K_j^\dagger h_{ij}^* h_{ij}K_j + \text{off-diagonal terms},\end{aligned} \tag{4.97}$$

where we divided $\alpha$ into diagonal terms and off-diagonal terms (in the Fork basis). The second term is always non-negative and the off-diagonal terms will only increase $\|\alpha\|$. It is then clear that we can always assume the optimal $h^\blacklozenge$ and the corresponding $\alpha^\blacklozenge$ are diagonal because

$$\|\alpha\| \geq \left\|\sum_{i=0}^M (\dot{K}_i - ih_{ii}K_i)^\dagger(\dot{K}_i - ih_{ii}K_j)\right\|. \tag{4.98}$$

Choose a diagonal $h^\blacklozenge$ and let $\Pi^\blacklozenge$ be the projection onto the subspace spanned by all eigenstates corresponding to the largest eigenvalue of $\alpha^\blacklozenge$, $\rho^\blacklozenge$ is optimal if it satisfies $\Pi^\blacklozenge\rho^\blacklozenge\Pi^\blacklozenge = \rho^\blacklozenge$ and

$$\mathrm{Re}[\mathrm{Tr}(\rho^\blacklozenge(i\mathbf{K}^\dagger\delta h)(\dot{\mathbf{K}} - ih^\blacklozenge\mathbf{K}))] = 0, \quad \forall\delta h \in \mathbb{H}_{M+1}. \tag{4.99}$$

We observe that when $\rho^\blacklozenge$ is optimal, the equation above still holds by replacing $\rho^\blacklozenge$ with its diagonal part. Then any diagonal $\rho^\blacklozenge$ which satisfies

$$\mathrm{Re}[\mathrm{Tr}(\rho^\blacklozenge iK_i^\dagger\delta h_{ii}(\dot{K}_i - ih_{ii}^\blacklozenge K_i))] = 0, \quad \forall\{\delta h_{ii}\}_{i=0}^M \in \mathbb{R}^{M+1} \tag{4.100}$$

is optimal. Therefore, we can always assume the input state has the form Eq. (4.94). Moreover, by assuming $h$ and $\rho$ are diagonal, we only need to deal with diagonal operators in this algorithm and the problem is essentially a quadratically constrained quadratic program, which might admit more efficient numerical methods than the SDP formulation.

# Chapter 5

# Application: Covariant Quantum Error Correction

The Eastin–Knill theorem [Eastin and Knill, 2009] (see also [Bravyi and König, 2013; Pastawski and Yoshida, 2015; Jochym-O'Connor et al., 2018; Wang et al., 2020]) states that any non-trivial local-error-correcting quantum code does not admit transversal implementations of a universal set of logical gates, ruling out the possibility of realizing fault-tolerant quantum computation using only transversal gates. In particular, any finite-dimensional local-error-correcting quantum code only admits a finite number of transversal logical operations, which forbids the existence of codes covariant with continuous symmetries (discrete symmetries are allowed though [Hayden et al., 2017; Faist et al., 2020]). More generally, quantum codes under symmetry constraints, namely covariant codes, are of great practical and theoretical interest. In general, a quantum code is covariant with respect to a logical Hamiltonian $H_L$ and a physical Hamiltonian $H_S$ if any symmetry transformation $e^{-iH_L\theta}$ is encoded into a symmetry transformation $e^{-iH_S\theta}$ in the physical system.

Although covariant codes cannot be perfectly local-error-correcting, they can still approximately correct errors with infidelity depending on the number of subsystems, the dimension of each subsystem, etc. The quantifications of such infidelity in covariant QEC were explored recently, leading to an approximate, or robust, version of the Eastin–Knill theorem [Faist et al., 2020; Woods and Alhambra, 2020], using complementary channel techniques [Bény and Oreshkov, 2010; Hayden et al., 2008; Bény et al., 2018]. Note that

these existing results only apply to erasure errors.

In this chapter, we investigate covariant QEC using quantum channel estimation theory leading to a series of improved understandings and bounds on the performance of covariant QEC. Covariant QEC is naturally a quantum metrological protocol—estimating the angle of any rotation of the physical system is equivalent to estimating that of the logical system with protection against noise. We already knew from Chapter 4 that perfectly error-correcting codes admitting a non-trivial logical Hamiltonian does not exist if the physical Hamiltonian fall into the Kraus span of the noise channel, also known as the HKS condition. It is also a sufficient condition of the non-existence of perfectly covariant QEC codes. When the HKS condition is satisfied, we establish a connection between the asymptotic channel QFI and the performance (or infidelity) of covariant QEC, which gives rise to the desired lower bound.

Our approach to covariant QEC is innovative and also advantageous compared to previous ones in many ways. The bounds generalize the no-go theorems of covariant QEC from local Hamiltonians with erasure errors to generic Hamiltonian and noise structures. In the special case of erasure noise, our lower bound improves the previous results in the small infidelity limit [Faist et al., 2020]. Furthermore, we shall demonstrate that there is an example of covariant codes called thermodynamic codes [Faist et al., 2020; Brandão et al., 2019] that saturates the lower bound for erasure noise and matches the scaling of the lower bound for depolarizing noise, while previous bounds only apply to the erasure noise setting and were not known to be saturable [Faist et al., 2020].

## 5.1 Covariant codes

A quantum code is a subspace of a physical system $S$, usually defined by the image of an (usually isometric) encoding channel $\mathcal{E}_{S\leftarrow L}$ from a logical system $L$. We call a code $\mathcal{E}_{S\leftarrow L}$ *covariant* if there exists a logical Hamiltonian $H_L$ and a physical Hamiltonian $H_S$ such that

$$\mathcal{E}_{S\leftarrow L}\circ\mathcal{U}_{L,\theta}=\mathcal{U}_{S,\theta}\circ\mathcal{E}_{S\leftarrow L},\ \forall\theta\in\mathbb{R}, \tag{5.1}$$

where $\mathcal{U}_{L,\theta}(\rho_L) = e^{-iH_L\theta}\rho_L e^{iH_L\theta}$ and $\mathcal{U}_{S,\theta}(\rho_S) = e^{-iH_S\theta}\rho_S e^{iH_S\theta}$ are the symmetry transformations on the logical and physical systems, respectively. We assume the dimensions of the physical and logical systems $d_S$ and $d_L$ are both finite and $H_L$ is non-trivial ($H_L \propto \mathbb{1}$). For simplicity, we also assume all Hamiltonians in this paper are traceless and we use $\Delta H_L$ and $\Delta H_S$ to denote the difference between the maximum and minimum eigenvalues of the operators.

As described in Section 2.4.2, we use the worse-case entanglement fidelity to quantify the code infidelity

$$\varepsilon(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L}) = 1 - \max_{\mathcal{R}_{L\leftarrow S}} f_B^2(\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L}, \mathbb{1}_L). \tag{5.2}$$

We call a code $\mathcal{E}_{S\leftarrow L}$ *$\varepsilon$-correctable* under $\mathcal{N}_S$, if $\varepsilon \geq \varepsilon(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L})$. We will use $\mathcal{R}^{\text{opt}}_{L\leftarrow S}$ to represent the optimal recovery channel and $\mathcal{I}_L$ to denote the effective noise channel $\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S \circ \mathcal{E}_{S\leftarrow L}$ in the logical system. We ignore highly inaccurate codes and will always assume $\varepsilon < 1/2$ in this paper.

## 5.2 Lower bound on the code infidelity

A good approximately error-correcting covariant code naturally provides a good quantum sensor to estimate an unknown parameter $\theta$ in the symmetry transformation $e^{-iH_S\theta}$. Consider a quantum signal $e^{-iH_S\theta}$ in a physical system, for example, the magnetic field in a spin system with $H_S$ being the angular momentum. The optimal sensitivity is usually limited by the strength of noise in the system. Instead of using the entire system to probe the signal, one could prepare an encoded probe state using covariant codes where $H_S$ is mapped into $H_L$ in the logical system. For covariant codes with low infidelity, the noise will be significantly reduced in the logical system and therefore provide a good sensitivity of the signal.

Theorem 4.1 prevents the existence of perfectly error-correcting covariant codes in the above scenario. In particular, it was known that given a noise channel $\mathcal{N}_S(\cdot) = \sum_{i=1}^r K_{S,i}(\cdot)K_{S,i}^\dagger$ and a physical Hamiltonian $e^{-iH_S\theta}$, there exist an encoding channel $\mathcal{E}_{S\leftarrow L}$

and a recovery channel $\mathcal{R}_{L\leftarrow S}$ such that

$$\mathcal{R}_{L\leftarrow S}\circ\mathcal{N}_S\circ\mathcal{U}_{S,\theta}\circ\mathcal{E}_{S\leftarrow L} \tag{5.3}$$

is a non-trivial unitary channel only if $H_S \notin \mathrm{span}_{\mathbb{H}}\{K_{S,i}^\dagger K_{S,j}, \forall i,j\}$. However, the above channel (Eq. (5.3)) with respect to any perfectly error-correcting covariant code is simply $\mathcal{U}_{L,\theta}$. Therefore, we conclude that perfectly error-correcting covariant codes does not exist when

$$H_S \in \mathrm{span}_{\mathbb{H}}\{K_{S,i}^\dagger K_{S,j}, \forall i,j\}, \tag{5.4}$$

which we call the "Hamiltonian-in-Kraus-span" (HKS) condition. One could check that local Hamiltonians with non-trivial local errors is a special case of the HKS condition. Note that the no-go result might be circumvented when the system dimension is infinite [Hayden et al., 2017; Faist et al., 2020].

To obtain a lower bound of the code infidelity using quantum channel estimation theory, we will use Theorem 4.2, which provides a single-letter expression for $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta)$:

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta) = \lim_{N\to\infty}\frac{F(\mathcal{N}_\theta^{\otimes N})}{N} = \begin{cases} 4\min_{h:\beta_\theta=0}\|\alpha_\theta\| & \text{(S)}, \\ +\infty & \text{otherwise}, \end{cases} \tag{5.5}$$
$$\text{(S):}\quad i\sum_{i=1}^r K_{i,\theta}^\dagger\partial_\theta K_{i,\theta} \in \mathrm{span}\{K_{i,\theta}^\dagger K_{j,\theta}, \forall i,j\},$$

where $\mathcal{N}_\theta(\cdot) = \sum_{i=1}^r K_{i,\theta}(\cdot)K_{i,\theta}^\dagger$, $h\in\mathbb{H}_r$ and

$$\alpha_\theta = (\partial_\theta\mathbf{K}_\theta + ih\mathbf{K}_\theta)^\dagger(\partial_\theta\mathbf{K}_\theta + ih\mathbf{K}_\theta), \tag{5.6}$$

$$\beta_\theta = \mathbf{K}_\theta^\dagger h\mathbf{K}_\theta - i\mathbf{K}_\theta^\dagger\partial_\theta\mathbf{K}_\theta. \tag{5.7}$$

Note that when (S) is violated, $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta)=\infty$ because we will have $F(\mathcal{N}_\theta^{\otimes N})\propto N^2$. We will call $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta)$ the regularized SLD QFI of $\mathcal{N}_\theta$ in this chapter. It is by definition monotonic, satisfying $\mathfrak{F}_{\mathrm{SQL}}(\Phi_1\circ(\mathcal{N}_\theta\otimes\mathbb{1})\circ\Phi_2)\le\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta)$ where $\Phi_{1,2}$ are any parameter-independent channels, due to the monotonicity of the state QFI.

In order to derive a lower bound on the infidelity of covariant codes using the channel QFI, we note that the channel QFI provides upper limit to the sensitivity of $\theta$ for $\mathcal{N}_{S,\theta} = \mathcal{N}_S \circ \mathcal{U}_{S,\theta}$, which cannot be broke using covariant QEC. For example, consider $N$ logical qubits each under a unitary evolution $e^{-i\theta H_L}$ with a noise rate $\varepsilon$. It is known that the SLD QFI of a noiseless $N$-qubit GHZ state is $(\Delta H_L)^2 N^2$ [Giovannetti et al., 2006]. Taking $N = \Theta(1/\varepsilon)$, the total noise can be bounded by a small constant, and the state SLD QFI per qubit is still roughly $\Theta((\Delta H_L)^2 N) = \Theta((\Delta H_L)^2/\varepsilon)$, which is always no greater than the regularized channel SLD QFI $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_{S,\theta})$ before QEC. Thus, $\varepsilon$ must be lower bounded by $\Theta((\Delta H_L)^2/\mathfrak{F}_{\text{SQL}}(\mathcal{N}_{S,\theta}))$. In fact, using the regularized SLD QFI, we can prove the following theorem:

**Theorem 5.1.** *Suppose a covariant code $\mathcal{E}_{S\leftarrow L}$ is $\varepsilon$-correctable under a noise channel $\mathcal{N}_S(\cdot) = \sum_{i=1}^r K_{S,i}(\cdot)K_{S,i}^\dagger$. If the HKS condition is satisfied, i.e.*

$$H_S \in \text{span}\{K_{S,i}^\dagger K_{S,j}, \forall i,j\}, \tag{5.8}$$

*then $\varepsilon$ is lower bounded as follows,*

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{(\Delta H_L)^2}{4\mathfrak{F}_{\text{SQL}}(\mathcal{N}_S, H_S)}, \tag{5.9}$$

*where $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_S, H_S)$ is the regularized SLD QFI of $\mathcal{N}_{S,\theta}$.*

*Specifically, $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_S, H_S) = 4\min_{h:\beta_S=0} \|\alpha_S\|$, $h$ is a Hermitian operator in $\mathbb{C}^{r\times r}$. $\alpha_S$ and $\beta_S$ are Hermitian operators acting on $S$ defined by*

$$\alpha_S = \mathbf{K}_S^\dagger h^2 \mathbf{K}_S - H_S^2, \quad \beta_S = \mathbf{K}_S^\dagger h \mathbf{K}_S - H_S. \tag{5.10}$$

We remark that Theorem 5.1 holds for non-isometric encoding channels, widening the scope of Theorem 1 in [Faist et al., 2020].

### 5.2.1 Proof of Theorem 5.1

The main obstacle to proving Theorem 5.1 is to relate the infidelity of covariant codes to the QFI of the effective quantum channel in the logical system. Here we overcome this obstacle

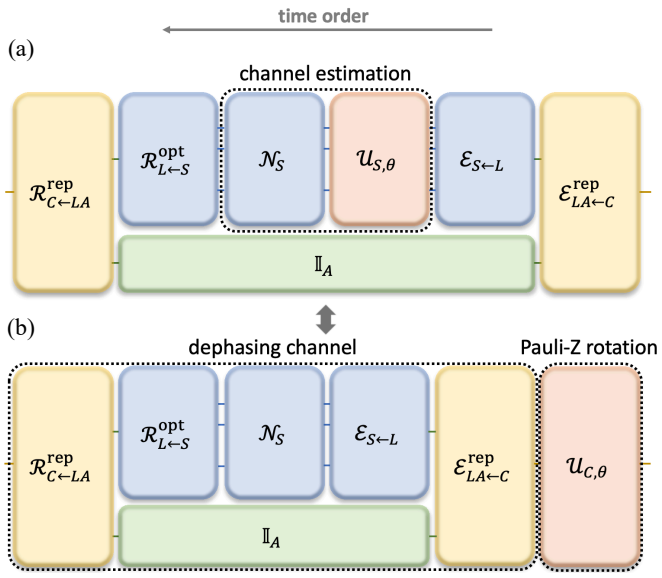


Figure 5.1: Reduction of $\mathcal{N}_{S,\theta} = \mathcal{N}_S \circ \mathcal{U}_{S,\theta}$ to dephasing channels using ancilla-assisted QEC. (a) represents the quantum channel $\mathcal{R}_{B\leftarrow SA} \circ (\mathcal{N}_{S,\theta} \otimes \mathbb{1}_A) \circ \mathcal{E}_{SA\leftarrow B}$ with a channel QFI no larger than $F(\mathcal{N}_{S,\theta})$. Because of the covariance of the code, (a) is equivalent to (b) which consists of a Pauli-Z rotation $\mathcal{U}_{B,\theta}$ and a $\theta$-independent dephasing channel $\mathcal{I}_B$ whose noise rate is smaller than $\varepsilon(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L})$ (see Lemma 5.1).

and provide a proof of Theorem 5.1 by employing ancilla-assisted QEC to reduce $\mathcal{N}_{S,\theta}$ to dephasing channels whose QFI has simple mathematical forms and then connecting the noise rate of the dephasing channels to the infidelity of the covariant codes (see Figure 5.1).

We define single-qubit dephasing channels to be

$$\mathcal{D}_{p,\phi}(\rho) = (1-p)e^{-i\frac{\phi}{2}Z}\rho e^{i\frac{\phi}{2}Z} + pe^{-i\frac{\phi}{2}Z}Z\rho Z e^{i\frac{\phi}{2}Z}, \tag{5.11}$$

where $Z$ is the Pauli-Z operator, $0 < p < 1/2$ and $\phi$ is real. When $\phi$ is a function of $\theta$, we could calculate the regularized SLD QFI of $\mathcal{D}_{p,\phi_\theta}$ (Section 4.2):

$$\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{p,\phi_\theta}) = \frac{(1-2p)^2(\partial_\theta\phi_\theta)^2}{4p(1-p)}, \tag{5.12}$$

which are both inversely proportional to the noise rate $p$ when $p$ is small—a crucial feature in deriving the lower bounds.

Next, we present an ancilla-assisted QEC protocol to reduce $\mathcal{N}_S$ to dephasing channels with a noise rate lower than $\varepsilon(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L})$. Let $|0_L\rangle$ and $|1_L\rangle$ be eigenstates respectively corresponding to the largest and the smallest eigenvalues of $H_L$. Consider the following

two-dimensional ancilla-assisted code

$$\mathcal{E}^{\text{rep}}_{LA\leftarrow B}(|0_B\rangle) = |0_L 0_A\rangle\,, \quad \mathcal{E}^{\text{rep}}_{LA\leftarrow B}(|1_B\rangle) = |1_L 1_A\rangle\,, \tag{5.13}$$

where $A$ is a noiseless ancillary qubit and the superscript rep means "repetition". The encoding channel from the two-level system $C$ to $SA$ will simply be $\mathcal{E}_{SA\leftarrow B} = (\mathcal{E}_{S\leftarrow L} \otimes \mathbb{1}_A) \circ \mathcal{E}^{\text{rep}}_{LA\leftarrow B}$. $\mathcal{E}_{SA\leftarrow B}$ is still a covariant code whose the logical and physical Hamiltonians are

$$H_B = \frac{\Delta H_L}{2} \cdot Z_B, \quad H_{SA} = H_S \otimes \mathbb{1}_A. \tag{5.14}$$

The noiseless ancillary qubit will help us suppress off-diagonal noises in the system because any single qubit bit-flip noise on $L$ could be fully corrected by mapping $|i_L j_A\rangle$ to $|j_B\rangle$ for all $i, j$. In fact, $\mathcal{N}_S$ will be reduced to a dephasing channel, as shown in the following lemma:

**Lemma 5.1.** *Consider a noise channel $\mathcal{N}_{SA} = \mathcal{N}_S \otimes \mathbb{1}_A$. There exists a recovery channel $\mathcal{R}_{B\leftarrow SA}$ such that the effective noise channel $\mathcal{I}_B = \mathcal{R}_{B\leftarrow SA} \circ \mathcal{N}_{SA} \circ \mathcal{E}_{SA\leftarrow B}$ is a dephasing channel, satisfying*

$$\mathcal{I}_B = \mathcal{D}_{B,\varepsilon',\phi'}, \tag{5.15}$$

*where $\varepsilon' \leq \varepsilon(\mathcal{N}_S, \mathcal{E}_{S\leftarrow L})$.*

*Proof.* To prove the lemma, we first calculate the worst-case entanglement fidelity for dephasing channels (Eq. (5.11)). We use the following formula for the worst-case entanglement fidelity [Schumacher, 1996]:

$$f^2(\mathcal{D}_{p,\phi}, \mathbb{1}) = \min_{|\psi\rangle} \langle\psi| \left(\mathcal{D}_{p,\phi} \otimes \mathbb{1}\right)(|\psi\rangle\langle\psi|)\,|\psi\rangle\,. \tag{5.16}$$

Let $|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$, then

$$(\mathcal{D}_{p,\phi}\otimes\mathbb{1})(|\psi\rangle\langle\psi|) = \begin{pmatrix} \alpha_{00}\alpha_{00}^* & \alpha_{00}\alpha_{01}^* & (1-2p)e^{-i\phi}\alpha_{00}\alpha_{10}^* & (1-2p)e^{-i\phi}\alpha_{00}\alpha_{11}^* \\ \alpha_{00}\alpha_{01}^* & \alpha_{01}\alpha_{01}^* & (1-2p)e^{-i\phi}\alpha_{01}\alpha_{10}^* & (1-2p)e^{-i\phi}\alpha_{01}\alpha_{11}^* \\ (1-2p)e^{i\phi}\alpha_{10}\alpha_{00}^* & (1-2p)e^{i\phi}\alpha_{10}\alpha_{01}^* & \alpha_{10}\alpha_{10}^* & \alpha_{10}\alpha_{11}^* \\ (1-2p)e^{i\phi}\alpha_{11}\alpha_{00}^* & (1-2p)e^{i\phi}\alpha_{11}\alpha_{01}^* & \alpha_{11}\alpha_{10}^* & \alpha_{11}\alpha_{11}^* \end{pmatrix}. \quad (5.17)$$

Then

$$\begin{aligned} 1-f^2(\mathcal{D}_{p,\phi},\mathbb{1}) &= \max_{\alpha_{00,01,10,11}} 2\mathrm{Re}[(1-(1-2p)e^{-i\phi})](|\alpha_{00}|^2+|\alpha_{01}|^2)(|\alpha_{10}|^2+|\alpha_{11}|^2) \\ &= \frac{1}{2}(1-(1-2p)\cos\phi) \geq p. \end{aligned} \quad (5.18)$$

Consider the following recovery channel

$$\mathcal{R}_{B\leftarrow SA} = \mathcal{R}^{\mathrm{rep}}_{B\leftarrow LA}\circ(\mathcal{R}^{\mathrm{opt}}_{L\leftarrow S}\otimes\mathbb{1}_A), \quad (5.19)$$

where $\mathcal{R}^{\mathrm{rep}}_{B\leftarrow LA}(\rho_{LA}) = \sum_{i=0}^{d_L-1}\sum_{j=0}^{1} R_{ij}\rho_{LA}R_{ij}^\dagger$, where $R_{ij} = |j_B\rangle\langle i_L j_A|$. One could check that

$$\mathcal{I}_B(|k_B\rangle\langle j_B|) = \begin{cases} |k_B\rangle\langle j_B|, & k=j, \\ (1-2\varepsilon')e^{i\phi'(k-j)}|k_B\rangle\langle j_B|, & k\neq j, \end{cases} \quad (5.20)$$

which indicates that $\mathcal{I}_B = \mathcal{D}_{B,\varepsilon',\phi'}$ (Eq. (5.15)). Here,

$$\varepsilon' \leq 1-f^2(\mathcal{I}_B,\mathbb{1}_B) \leq 1-f^2(\mathcal{I}_L^{\mathrm{opt}},\mathbb{1}_L) = \varepsilon(\mathcal{N}_S,\mathcal{E}_{S\leftarrow L}). \quad (5.21)$$

where the first inequality follows from the worst-case entanglement fidelity for dephasing channels, and the the second inequality follows from $\mathbb{1}_B = \mathcal{R}^{\mathrm{rep}}_{B\leftarrow LA}\circ\mathcal{E}^{\mathrm{rep}}_{LA\leftarrow B}$ and the monotonicity of the fidelity [Nielsen and Chuang, 2010].

□

Lemma 5.1 shows that $\mathcal{N}_S$ could be reduced to a dephasing channel $\mathcal{I}_B$ through ancilla-assisted QEC. Consider parameter estimation of $\theta$ in the quantum channel $\mathcal{N}_{S,\theta} = \mathcal{N}_S\circ\mathcal{U}_{S,\theta}$.

We have the error-corrected quantum channel

$$\mathcal{N}_{B,\theta} = \mathcal{R}_{B\leftarrow SA} \circ (\mathcal{N}_{S,\theta} \otimes \mathbb{1}_A) \circ \mathcal{E}_{SA\leftarrow B} = \mathcal{I}_B \circ \mathcal{U}_{B,\theta}, \tag{5.22}$$

equal to a dephasing channel with noise rate $\varepsilon'$ and phase $\phi_\theta = \phi' + \Delta H_L \theta$. The monotonicity of the regularized channel SLD QFI implies that

$$\mathfrak{F}_{\text{SQL}}(\mathcal{N}_{S,\theta}) \geq \mathfrak{F}_{\text{SQL}}(\mathcal{N}_{B,\theta}), \tag{5.23}$$

where

$$\mathfrak{F}_{\text{SQL}}(\mathcal{N}_{S,\theta}) = \begin{cases} \mathfrak{F}_{\text{SQL}}(\mathcal{N}_S, H_S) & H_S \in \text{span}\{K_{S,i}^\dagger K_{S,j}, \forall i,j\}, \\ +\infty & \text{otherwise}, \end{cases} \tag{5.24}$$

and

$$\mathfrak{F}_{\text{SQL}}(\mathcal{N}_{B,\theta}) = \frac{(1-2\varepsilon')^2 (\Delta H_L)^2}{4\varepsilon'(1-\varepsilon')}. \tag{5.25}$$

Theorem 5.1 then follows from Eq. (5.23) and $\varepsilon' \leq \varepsilon < 1/2$.

## 5.3 Local Hamiltonian and local noise

One of the most common scenarios where covariant codes is considered is when $S$ is an $n$-partite system, consisting of subsystems $S_1, S_2, \ldots, S_n$. The physical Hamiltonian and the noise channel are both local, given by

$$H_S = \sum_{k=1}^n H_{S_k}, \quad \mathcal{N}_S = \bigotimes_{k=1}^n \mathcal{N}_{S_k}, \quad \mathcal{N}_{S_k}(\cdot) = \sum_{i=1}^{r_k} K_{S_k,i}(\cdot) K_{S_k,i}^\dagger. \tag{5.26}$$

In general, it takes time exponential in the number of subsystems to solve our lower bounds on the code infidelity. However, when the Hamiltonians and the noises are local, using the additivity of $\mathfrak{F}_{\text{SQL}}$ (proven later), we could directly calculate the lower bounds, requiring only computation of the subsystem QFI. To be specific, for $\varepsilon$-correctable codes under $\mathcal{N}_S$, Theorem 5.1 indicates that when

$$H_{S_k} \in \mathrm{span}\{K_{S_k,i}^\dagger K_{S_k,j}, \forall i,j\}, \quad \forall k, \tag{5.27}$$

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{(\Delta H_L)^2}{4\sum_{k=1}^n \mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_{S_k}, H_{S_k})}. \tag{5.28}$$

Instead of finding bounds for local noise channels $\mathcal{N}_S$ with certain noise rates, we sometimes are more interested the capability of a code to correct single errors (each described by $\mathcal{M}_{S_k}$). Consider the single-error noise channel

$$\mathcal{M}_S = \sum_{k=1}^n q_k \mathcal{M}_{S_k}, \quad \sum_{k=1}^n q_k = 1, \tag{5.29}$$

where $q_k$ is the probability that an error $\mathcal{M}_{S_k}$ occurs on the $k$-th subsystem. In order to obtain lower bounds on the code infidelity under noise channels $\mathcal{M}_S$, we use the following local noise channel

$$\mathcal{N}_S(\delta) = \bigotimes_{k=1}^n \mathcal{N}_{S_k}(\delta) = \bigotimes_{k=1}^n \left((1-\delta q_k)\mathbb{1} + \delta q_k \mathcal{M}_{S_k}\right) = (1-\delta)\mathbb{1} + \delta \sum_{k=1}^n q_k \mathcal{M}_{S_k} + O(\delta^2), \tag{5.30}$$

whose local noise rates are proportional to a small positive parameter $\delta$. Using the concavity of $f^2(\Phi, \mathbb{1})$, we have

$$f^2(\mathcal{R}_{L\leftarrow S} \circ \mathcal{N}_S(\delta) \circ \mathcal{E}_{S\leftarrow L}, \mathbb{1}_L) \geq (1-\delta) + \delta f^2(\mathcal{R}_{L\leftarrow S} \circ \mathcal{M}_S \circ \mathcal{E}_{S\leftarrow L}, \mathbb{1}_L) + O(\delta^2). \tag{5.31}$$

Taking the limit $\delta \to 0^+$, we must have $\varepsilon(\mathcal{M}_S, \mathcal{E}_{S\leftarrow L}) \geq \liminf_{\delta\to 0^+} \frac{1}{\delta} \cdot \varepsilon(\mathcal{N}_S(\delta), \mathcal{E}_{S\leftarrow L})$. Therefore, for $\varepsilon$-correctable codes under single-error noise channels $\mathcal{M}_S$, Theorem 5.1 indicates that when Eq. (5.27) is satisfied,

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \liminf_{\delta\to 0^+} \frac{(\Delta H_L)^2}{4\delta\sum_{k=1}^n \mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_{S_k}(\delta), H_{S_k})}. \tag{5.32}$$

### 5.3.1 Additivity of $\mathfrak{F}_{\mathrm{sql}}$

Here we prove the additivity of the regularized SLD QFI.

$$F^{\mathrm{reg}}(\mathcal{N}_\theta \otimes \tilde{\mathcal{N}}_\theta) = F^{\mathrm{reg}}(\mathcal{N}_\theta) + F^{\mathrm{reg}}(\tilde{\mathcal{N}}_\theta), \tag{5.33}$$

for arbitrary quantum channels $\mathcal{N}_\theta$ and $\tilde{\mathcal{N}}_\theta$.

First, according to the additivity of the state QFI, we must have

$$F^{\mathrm{reg}}(\mathcal{N}_\theta \otimes \tilde{\mathcal{N}}_\theta) \geq F^{\mathrm{reg}}(\mathcal{N}_\theta) + F^{\mathrm{reg}}(\tilde{\mathcal{N}}_\theta). \tag{5.34}$$

Thus, we only need to prove

$$F^{\mathrm{reg}}(\mathcal{N}_\theta \otimes \tilde{\mathcal{N}}_\theta) \leq F^{\mathrm{reg}}(\mathcal{N}_\theta) + F^{\mathrm{reg}}(\tilde{\mathcal{N}}_\theta). \tag{5.35}$$

We use the following definition of the regularized SLD QFI:

$$F^{\mathrm{reg}}(\mathcal{N}_\theta) = \begin{cases} 4 \min_{\mathbf{K}':\beta=0} \|\alpha\|, & i\sum_{i=1}^r (\partial_\theta K_i)^\dagger K_i \in \mathrm{span}\{K_i^\dagger K_j, \forall i,j\}, \\ +\infty & \text{otherwise}, \end{cases} \tag{5.36}$$

where $\mathbf{K}'$ is any set of Kraus operators representing $\mathcal{N}_\theta$, $\alpha = \sum_{i=1}^r (\partial_\theta K_i')^\dagger (\partial_\theta K_i')$ and $\beta = i\sum_{i=1}^r (\partial_\theta K_i')^\dagger K_i'$. Without loss of generality, assume both $F^{\mathrm{reg}}(\mathcal{N}_\theta)$ and $F^{\mathrm{reg}}(\tilde{\mathcal{N}}_\theta)$ are finite, i.e. $i\sum_{i=1}^r (\partial_\theta K_i)^\dagger K_i \in \mathrm{span}\{K_i^\dagger K_j, \forall i,j\}$ and $i\sum_{i=1}^{\tilde{r}} (\partial_\theta \tilde{K}_i)^\dagger \tilde{K}_i \in \mathrm{span}\{\tilde{K}_i^\dagger \tilde{K}_j, \forall i,j\}$. We first note that $F^{\mathrm{reg}}(\mathcal{N}_\theta \otimes \tilde{\mathcal{N}}_\theta)$ is also finite, because

$$i\sum_{i=1}^r \sum_{j=1}^{\tilde{r}} (\partial_\theta (K_i \otimes \tilde{K}_j))^\dagger (K_i \otimes \tilde{K}_j) = i\sum_{i=1}^r (\partial_\theta K_i)^\dagger K_i \otimes \mathbb{1} + i\sum_{j=1}^{\tilde{r}} \mathbb{1} \otimes (\partial_\theta \tilde{K}_j)^\dagger \tilde{K}_j \tag{5.37}$$

$$\in \mathrm{span}\{\mathbb{1} \otimes K_i^\dagger K_j, \tilde{K}_i^\dagger \tilde{K}_j \otimes \mathbb{1}, \forall i,j\}. \tag{5.38}$$

According to Eq. (5.36), there exists $\mathbf{K}'$ and $\tilde{\mathbf{K}}'$ such that $\beta = \tilde{\beta} = 0$ and

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta) = 4\|\alpha\|, \quad \mathfrak{F}_{\mathrm{SQL}}(\tilde{\mathcal{N}}_\theta) = 4\|\tilde{\alpha}\|. \tag{5.39}$$

Then $\tilde{\tilde{K}}_{ij}' = K_i' \otimes \tilde{K}_j'$ is a set of Kraus operators representing $\mathcal{N}_\theta \otimes \tilde{\mathcal{N}}_\theta$.

$$\tilde{\tilde{\alpha}} = \sum_{i=1}^{r}\sum_{j=1}^{\tilde{r}} \partial_\theta(\tilde{\tilde{K}}_{ij})^\dagger \partial_\theta(\tilde{\tilde{K}}_{ij}) = \alpha\otimes\mathbb{1} + \mathbb{1}\otimes\tilde{\alpha} + 2\beta\otimes\tilde{\beta} = \alpha\otimes\mathbb{1} + \mathbb{1}\otimes\tilde{\alpha}, \tag{5.40}$$

$$\tilde{\tilde{\beta}} = \beta\otimes\mathbb{1} + \mathbb{1}\otimes\tilde{\beta} = 0. \tag{5.41}$$

Therefore $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\theta\otimes\tilde{\mathcal{N}}_\theta) \le 4\left\|\tilde{\tilde{\alpha}}\right\| = 4\left\|\alpha\right\| + 4\left\|\tilde{\alpha}\right\| = F^{\mathrm{reg}}(\mathcal{N}_\theta) + F^{\mathrm{reg}}(\tilde{\mathcal{N}}_\theta)$.

### 5.3.2 Erasure noise

Now we present our bounds for the local erasure noise channel $\mathcal{N}^{\mathrm{e}}(\rho) = (1-p)\rho + p\left|\mathrm{vac}\right\rangle\left\langle\mathrm{vac}\right|$ on each subsystem. Here $p$ is the noise rate and we use the vacuum state $\left|\mathrm{vac}\right\rangle$ to represent the state of the erased subsystems. The Kraus operators for $\mathcal{N}^{\mathrm{e}}$ are

$$K_1 = \sqrt{1-p}\mathbb{1}, \quad K_{i+1} = \sqrt{p}\left|\mathrm{vac}\right\rangle\left\langle i\right|, \quad \forall 1 \le i \le d. \tag{5.42}$$

Different subsystems can have different noise rates $p_k$ and dimensions $d_k$.

We first calculate $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}^{\mathrm{e}}, H)$ where $\mathcal{N}^{\mathrm{e}} = (1-p)\rho + p\left|\mathrm{vac}\right\rangle\left\langle\mathrm{vac}\right|$. Using the Kraus operators in Eq. (5.42),

$$\beta = \mathbf{K}^\dagger h\mathbf{K} - H \quad\Leftrightarrow\quad h = \begin{pmatrix} \frac{h_{11}}{1-p} & 0 \\ 0 & \frac{H-h_{11}\mathbb{1}}{p} \end{pmatrix}. \tag{5.43}$$

Then

$$\alpha = \mathbf{K}^\dagger h^2\mathbf{K} - H^2 = \frac{h_{11}^2}{1-p} + \frac{(H-h_{11}\mathbb{1})^2}{p} - H^2 = \frac{1-p}{p}H^2 - \frac{2h_{11}}{p}H + \frac{h_{11}^2}{p(1-p)}, \tag{5.44}$$

$$\begin{aligned} F^{\mathrm{reg}}_{\mathscr{I}}(\mathcal{N}^{\mathrm{e}}, H) &= 4\min_{h_{11}}\left\|\alpha\right\| = 4\max_{\rho}\min_{h_{11}}\mathrm{Tr}(\rho\alpha) \\ &= 4\max_{\rho}\frac{1-p}{p}\left(\mathrm{Tr}(H^2\rho) - \mathrm{Tr}(\rho H)^2\right) = \frac{1-p}{p}(\Delta H)^2, \end{aligned} \tag{5.45}$$

where we use the minimax theorem [Komiya, 1988; do Rosário Grossinho and Tersian, 2001] in the second step.

Therefore, the regularized SLD QFI for erasure noise is

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}^{\mathrm{e}}, H) = (\Delta H)^2 \frac{1-p}{p}. \tag{5.46}$$

For $\varepsilon$-correctable codes under local erasure noise channel $\mathcal{N}_S^{\mathrm{e}} = \bigotimes_{k=1}^n \mathcal{N}_{S_k}^{\mathrm{e}}$, we have

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{(\Delta H_L)^2}{4\sum_{k=1}^n \frac{1-p_k}{p_k}(\Delta H_{S_k})^2}, \tag{5.47}$$

using Eq. (5.28). For $\varepsilon$-correctable codes under single-error erasure noise channel $\mathcal{M}_S^{\mathrm{e}} = \sum_{k=1}^n q_k \mathcal{M}_{S_k}^{\mathrm{e}}$ where $\mathcal{M}_{S_k}^{\mathrm{e}}(\rho_{S_k}) = |\mathrm{vac}\rangle \langle \mathrm{vac}|_{S_k}$,

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{(\Delta H_L)^2}{4\sum_{k=1}^n \frac{1}{q_k}(\Delta H_{S_k})^2}, \tag{5.48}$$

using Eq. (5.32).

In particular, when the probability of erasure is uniform on each subsystem, i.e. $q_k = \frac{1}{n}$, we have

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{(\Delta H_L)^2}{4n\sum_{k=1}^n (\Delta H_{S_k})^2}. \tag{5.49}$$

As a comparison, Theorem 1 in [Faist et al., 2020] showed that

$$\varepsilon \geq \frac{(\Delta H_L)^2}{4n^2 \max_k (\Delta H_{S_k})^2}. \tag{5.50}$$

Our bound Eq. (5.49) has a clear advantage in the small infidelity limit by improving the maximum of $\Delta H_{S_k}$ to their quadratic mean. A direct implication of Eq. (5.49) is an improved approximate Eastin–Knill theorem which establishes the infidelity lower bound for covariant codes with respect to special unitary groups. $SU(d_L)$-covariant codes in an $n$-partite system $S$ are defined by the encoding channels $\mathcal{E}_{S\leftarrow L}$ which satisfy

$$\mathcal{E}_{S\leftarrow L}(U_L(g)(\cdot)U_L^\dagger(g)) = \left(\bigotimes_{k=1}^n U_{S_k}(g)\right)\mathcal{E}_{S\leftarrow L}(\cdot)\left(\bigotimes_{k=1}^n U_{S_k}^\dagger(g)\right),\ \forall g \in SU(d_L), \tag{5.51}$$

where $U_{S_k}(g)$ and $U_L(g)$ are unitary representations of $SU(d_L)$. It was shown in Theorem 18 in the Supplemental Material of [Faist et al., 2020] that fixing $H_L = \mathrm{diag}(1, 0, \ldots, -1)$ and letting $H_{S_k}$ be the corresponding generator acting on the subsystem $k$, we have

$$d_k \geq \binom{d_L - 1 + \lceil \|H_{S_k}\| \rceil}{d_L - 1}, \tag{5.52}$$

where $\lceil \|H_{S_k}\| \rceil$ denotes the closest integer no smaller than $\|H_{S_k}\|$. Using the inequality $\binom{a+b}{a} \geq (1 + \frac{b}{a})^a$,

$$d_k \geq \left( \frac{d_L - 1 + \lceil \|H_{S_k}\| \rceil}{d_L - 1} \right)^{d_L - 1}, \quad \Rightarrow \quad \left( \exp\left( \frac{\ln d_k}{d_L - 1} \right) - 1 \right) (d_L - 1) \geq \|H_{S_k}\|, \tag{5.53}$$

$$\Rightarrow \quad \sum_{k=1}^{n} \left( \exp\left( \frac{\ln d_k}{d_L - 1} \right) - 1 \right)^2 (d_L - 1)^2 \geq \frac{1}{4} \sum_k (\Delta H_{S_k})^2. \tag{5.54}$$

Then using Eq. (5.49), we have for any $\varepsilon \geq \varepsilon(\mathcal{M}_S, \mathcal{E}_{S \leftarrow L})$,

$$\varepsilon \cdot \frac{1 - \varepsilon}{(1 - 2\varepsilon)^2} \geq \frac{1}{4n \sum_{k=1}^{n} \left( \exp\left( \frac{\ln d_k}{d_L - 1} \right) - 1 \right)^2 (d_L - 1)^2}. \tag{5.55}$$

For large $d_L$,

$$\varepsilon \cdot \frac{1 - \varepsilon}{(1 - 2\varepsilon)^2} \geq \frac{1}{4n \sum_{k=1}^{n} (\ln d_k)^2} + O\left( \frac{1}{n^2 d_L} \right). \tag{5.56}$$

Compared to Theorem 4 in [Faist et al., 2020]:

$$\varepsilon \geq \left( \frac{1}{2n \max_k \ln d_k} + O\left( \frac{1}{n d_L} \right) \right)^2, \tag{5.57}$$

our bound improves the maximum of $\ln d_k$ in the denominator to their quadratic mean. Moreover, it works for not only single-error erasure noise channel $\mathcal{M}_S = \sum_{k=1}^{n} \frac{1}{n} \mathcal{M}_{S_k}$ where $\mathcal{M}_{S_k}(\cdot) = |\text{vac}\rangle \langle \text{vac}|_{S_k}$, but also single-error depolarizing noise channel $\mathcal{M}_S = \sum_{k=1}^{n} \frac{1}{n} \mathcal{M}_{S_k}$ where $\mathcal{M}_{S_k}(\cdot) = \frac{\mathbb{1}}{d_k}$.

### 5.3.3 Depolarizing noise

Next, we present our bounds for local depolarizing noise channel $\mathcal{N}^{\text{d}}(\rho) = (1 - p)\rho + p\frac{\mathbb{1}}{d}$ on each subsystem, which has not been studied before. Again, we assume different subsystems can have different noise rates $p_k$ and dimensions $d_k$. The Kraus operators for $\mathcal{N}^{\text{d}}$ are

$$K_1 = \sqrt{1 - \frac{d^2 - 1}{d^2} p}\, \mathbb{1}, \quad K_i = \sqrt{\frac{p}{d^2}} U_{i-1}, \quad \forall 2 \leq i \leq d^2, \tag{5.58}$$

where $\{U_0 = \mathbb{1}, U_1, \ldots, U_{d_k^2-1}\}$ is a unitary orthonormal basis in $\mathbb{C}^{d\times d}$.

In order to apply Theorem 5.1, we need to solve the following SDP

$$\mathfrak{F}_{\text{SQL}}(\mathcal{N}^{\text{d}}, H) = \min_{h:\beta=0} 4\,\|\alpha\|\,, \tag{5.59}$$

where $\beta = \mathbf{K}^\dagger h\mathbf{K} - H$ and $\alpha = \mathbf{K}^\dagger h^2\mathbf{K} - H^2$.

When $d = 2$, using Section 4.5.1, we have $\mathfrak{F}_{\text{SQL}}(\mathcal{N}^{\text{d}}, H) = (\Delta H)^2 \frac{2(1-p)^2}{p(3-2p)}$. When all subsystems are qubits, for $\varepsilon$-correctable codes under local depolarizing noise channels $\mathcal{N}_S^{\text{d}} = \bigotimes_{k=1}^n \mathcal{N}_{S_k}^{\text{d}}$,

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{(\Delta H_L)^2}{4\sum_{k=1}^n \frac{2(1-p_k)^2}{p_k(3-2p_k)}(\Delta H_{S_k})^2}, \tag{5.60}$$

using Eq. (5.28) and for $\varepsilon$-correctable codes under single-error depolarizing noise channels $\mathcal{M}_S^{\text{d}} = \sum_{k=1}^n q_k \mathcal{M}_{S_k}^{\text{d}}$ where $\mathcal{M}_{S_k}^{\text{d}}(\rho_{S_k}) = \frac{\mathbb{1}}{2}$,

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{3(\Delta H_L)^2}{8\sum_{k=1}^n \frac{1}{q_k}(\Delta H_{S_k})^2}, \tag{5.61}$$

using Eq. (5.32).

The situation is more complicated when $d > 2$, because the regularized SLD QFI may not have a closed-form expression. Instead, we can show that

$$\mathfrak{F}_{\text{SQL}}(\mathcal{N}^{\text{d}}, H) \leq (\Delta H)^2 \frac{(1-p)^2}{p(1+\frac{2}{d^2}-p)} \leq (\Delta H)^2 \frac{1-p}{p}, \tag{5.62}$$

by choosing a special $h$ which satisfies $\beta = 0$ to calculate an upper bound on $4\min_{h:\beta=0}\|\alpha\|$.

To prove Eq. (5.62), note that general depolarizing channels $\mathcal{N}^{\text{d}}(\rho) = (1-p)\rho + p\frac{\mathbb{1}}{d}$ have the Kraus operators

$$K_1 = \sqrt{x}\mathbb{1}, \quad K_i = \sqrt{y}U_{i-1}, \forall 2 \leq i \leq d^2, \tag{5.63}$$

where we define $x = 1 - \frac{d^2-1}{d^2}p$, $y = \frac{1}{d^2}p$. Any $\tilde{h}$ satisfying $\tilde{\beta} = \mathbf{K}^\dagger \tilde{h}\mathbf{K} - H = 0$ provides an upper bound on $\mathfrak{F}_{\text{SQL}}(\mathcal{N}^{\text{d}}, H)$ through

$$\mathfrak{F}_{\text{SQL}}(\mathcal{N}^{\text{d}}, H) = 4\min_{h:\beta=0}\|\alpha\| \leq 4\,\|\alpha\|\,|_{h=\tilde{h}}. \tag{5.64}$$

To find a suitable $\tilde{h}$ which provides a good upper bound on $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}^{\mathrm{d}}, H)$, we use $\tilde{h}$ which is the solution of

$$4 \min_{h:\beta=0} \mathrm{Tr}(\alpha). \tag{5.65}$$

The solution of Eq. (5.65) is

$$\tilde{h} = \frac{1}{2zd} \begin{pmatrix} 0 & \frac{\sqrt{xy}}{x+y}\mathrm{Tr}(HU_1^\dagger U_0) & \cdots & \frac{\sqrt{xy}}{x+y}\mathrm{Tr}(HU_{d^2-1}^\dagger U_0) \\ \frac{\sqrt{xy}}{x+y}\mathrm{Tr}(HU_0^\dagger U_1) & 0 & \cdots & \frac{1}{2}\mathrm{Tr}(HU_{d^2-1}^\dagger U_1) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sqrt{xy}}{x+y}\mathrm{Tr}(HU_0^\dagger U_{d^2-1}) & \frac{1}{2}\mathrm{Tr}(HU_1^\dagger U_{d^2-1}) & \cdots & 0 \end{pmatrix}, \tag{5.66}$$

where $z = \frac{xy}{x+y} + \frac{y(d^2-2)}{4}$ and we used the assumption $\mathrm{Tr}(H) = 0$ and

$$\mathbf{K}^\dagger \tilde{h}^2 \mathbf{K} = \left(\frac{1}{4z} - \frac{y}{4z^2}\left(\frac{1}{4} - \frac{xy}{(x+y)^2}\right) - 1\right) H^2 + \frac{y}{4z^2 d}\left(\frac{x}{x+y} - \frac{1}{2}\right)^2 \mathrm{Tr}(H^2)\mathbb{1}. \tag{5.67}$$

Using $\|H^2\| = \frac{(\Delta H)^2}{4}$ and $\mathrm{Tr}(H^2) \leq \frac{d}{4}(\Delta H)^2$,

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}^{\mathrm{d}}, H) \leq 4\,\|\alpha\| \leq (\Delta H)^2\left(\frac{1}{4z} - 1\right) = (\Delta H)^2 \frac{d^2(1-p)^2}{p(d^2(1-p)+2)}, \tag{5.68}$$

proving Eq. (5.62).

Note that the right-hand side of Eq. (5.62) is equal to the regularized SLD QFI for erasure channels Eq. (5.46). We conclude that Eqs. (5.47)-(5.49) hold true for general depolarizing channels as well, regardless of the dimensions of subsystems. We also remark that the upper bound on the regularized SLD QFI for depolarizing channels we derived here might be of independent interest in quantum metrology.

### 5.3.4 Example: Thermodynamic codes

Finally, we provide an example saturating the lower bound for single-error erasure noise channels in the small infidelity limit and matching the scaling of the lower bound for single-error depolarizing noise channels, while previously only the scaling optimality for erasure

channels was demonstrated [Faist et al., 2020].

We consider the following two-dimensional thermodynamic code [Brandão et al., 2019; Faist et al., 2020; Ouyang et al., 2019]

$$\mathcal{E}_{S\leftarrow L}(|0_L\rangle) = |\mathfrak{c}_0\rangle = |m_n\rangle\,, \quad \mathcal{E}_{S\leftarrow L}(|1_L\rangle) = |\mathfrak{c}_1\rangle = |(-m)_n\rangle\,, \tag{5.69}$$

where

$$|(\pm m)_n\rangle = \binom{n}{\frac{n\pm m}{2}}^{-\frac{1}{2}} \sum_{\boldsymbol{j}:\sum_k j_k=\pm m} |\boldsymbol{j}\rangle\,, \tag{5.70}$$

and $\boldsymbol{j} = (j_1, j_2, \ldots, j_n) \in \{-1, 1\}^n$. The logical subspace is spanned by two Dicke states with different values of the total angular momentum along the z-axis. We also assume $n+m$ is an even number and $3 \leq m \ll N$. It is a covariant code whose physical and logical Hamiltonians are

$$H_S = \sum_{k=1}^{n} (\sigma_z)_{S_k}, \quad H_L = mZ_L, \tag{5.71}$$

where $\sigma_z = |1\rangle\langle 1| - |-1\rangle\langle -1|$.

Let $|\mathfrak{c}^{(k)}_{0,\pm1}\rangle = |(m \pm 1)_{n-1}\rangle_{S\setminus S_k} |\mathrm{vac}\rangle_{S_k}$, $|\mathfrak{c}^{(k)}_{1,\pm1}\rangle = |(-m \pm 1)_{n-1}\rangle_{S\setminus S_k} |\mathrm{vac}\rangle_{S_k}$, which represent the logical states after an erasure error occurs on $S_k$, and $\Pi^{\perp}$ be the projector onto the orthogonal subspace of $\mathrm{span}\{|\mathfrak{c}^{(k)}_{0,\pm1}\rangle, |\mathfrak{c}^{(k)}_{1,\pm1}\rangle, \forall k\}$. Consider the erasure noise channel $\mathcal{M}_S = \frac{1}{n}\sum_{k=1}^{n} \mathcal{M}_{S_k}$ where $\mathcal{M}_{S_k}(\rho_{S_k}) = |\mathrm{vac}\rangle\langle\mathrm{vac}|_{S_k}$ and the recovery channel

$$\mathcal{R}_{L\leftarrow S}(\rho_S) = \sum_{k=1}^{n} \sum_{i,i'=0}^{1} \sum_{j=\pm1} |\mathfrak{c}_i\rangle\langle\mathfrak{c}^{(k)}_{i,j}| \rho_S |\mathfrak{c}^{(k)}_{i',j}\rangle\langle\mathfrak{c}_{i'}| + \mathrm{Tr}(\Pi^{\perp}\rho_S\Pi^{\perp}) |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0|\,, \tag{5.72}$$

which maps the state $|\mathfrak{c}^{(k)}_{i,\pm1}\rangle$ to $|\mathfrak{c}_i\rangle$ for all $k$. Then we could verify that $\mathcal{R}_{L\leftarrow S}\circ\mathcal{M}_S\circ\mathcal{E}_{S\leftarrow L} = \mathcal{D}_{p,0}$ with $p = \frac{1}{2}(1 - \sqrt{1 - \frac{m^2}{n^2}})$. Using the relation between the noise rate $p$ and the worst-case entanglement fidelity of a dephasing channel, we must have

$$\varepsilon(\mathcal{M}_S, \mathcal{E}_{S\leftarrow L}) \leq 1 - f^2(\mathcal{R}_{L\leftarrow S}\circ\mathcal{M}_S\circ\mathcal{E}_{S\leftarrow L}, \mathbb{1}_L) \tag{5.73}$$

$$= \frac{1}{2}\left(1 - \sqrt{1 - \frac{m^2}{n^2}}\right) = \frac{m^2}{4n^2} + O\left(\frac{m^4}{n^4}\right). \tag{5.74}$$

On the other hand, the lower bound (Eq. (5.49)) for $\varepsilon = \varepsilon(\mathcal{M}_S, \mathcal{E}_{S\leftarrow L})$ is given by

$$\varepsilon \cdot \frac{1-\varepsilon}{(1-2\varepsilon)^2} \geq \frac{m^2}{4n^2}, \tag{5.75}$$

which is saturated asymptotically when $m/N \to 0$.

Next, we consider the single-error depolarizing noise channel $\mathcal{M}_S = \frac{1}{n}\sum_{k=1}^{n} \mathcal{M}_{S_k}$ where $\mathcal{M}_{S_k}(\rho_{S_k}) = \frac{\mathbb{1}}{2}$. It is in general difficult to write down the optimal recovery map explicitly. Instead, in order to calculate $\varepsilon(\mathcal{M}_S, \mathcal{E}_{S\leftarrow L})$, we apply Lemma 2.2 to calculate an upper bound on the infidelity of thermodynamic codes in the limit $m/N \to 0$.

Let $\Pi = |\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| + |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|$, $\mathcal{M} = \mathcal{M}_S$ with Kraus operators

$$E_{k,i} = \frac{1}{2\sqrt{n}}(U_i)_{S_k}, \quad i = 0,1,2,3, \tag{5.76}$$

where $U_0, U_1, U_2, U_3$ are respectively $\mathbb{1}, \sigma_x = |1\rangle\langle -1| + |-1\rangle\langle 1|, \sigma_y = -i\,|1\rangle\langle -1| + i\,|-1\rangle\langle 1|$, and $\sigma_z = |1\rangle\langle 1| - |-1\rangle\langle -1|$.

For $m \geq 3$, $\langle\mathfrak{c}_0|E|\mathfrak{c}_1\rangle = 0$ for any operator $E$ acting on at most two qubits. Here we consider $\delta A_{ij} \propto (|\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| - |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|)$. That is, let $\delta A_{ij} = B_{ij}(|\mathfrak{c}_0\rangle\langle\mathfrak{c}_0| - |\mathfrak{c}_1\rangle\langle\mathfrak{c}_1|)$. $A$ and $B$ are $4n \times 4n$ matrices

$$A = \begin{pmatrix} A^{(0,0)} & A^{(0,1)} & A^{(0,2)} & A^{(0,3)} \\ A^{(1,0)} & A^{(1,1)} & A^{(1,2)} & A^{(1,3)} \\ A^{(2,0)} & A^{(2,1)} & A^{(2,2)} & A^{(2,3)} \\ A^{(3,0)} & A^{(3,1)} & A^{(3,2)} & A^{(3,3)} \end{pmatrix}, \quad B = \begin{pmatrix} B^{(0,0)} & B^{(0,1)} & B^{(0,2)} & B^{(0,3)} \\ B^{(1,0)} & B^{(1,1)} & B^{(1,2)} & B^{(1,3)} \\ B^{(2,0)} & B^{(2,1)} & B^{(2,2)} & B^{(2,3)} \\ B^{(3,0)} & B^{(3,1)} & B^{(3,2)} & B^{(3,3)} \end{pmatrix}, \tag{5.77}$$

where $A^{(i,j)}_{kk'} = \frac{1}{2}(\langle\mathfrak{c}_0|E^\dagger_{k,i}E_{k',j}|\mathfrak{c}_0\rangle + \langle\mathfrak{c}_1|E^\dagger_{k,i}E_{k',j}|\mathfrak{c}_1\rangle)$, $B^{(i,j)}_{kk'} = \frac{1}{2}(\langle\mathfrak{c}_0|E^\dagger_{k,i}E_{k',j}|\mathfrak{c}_0\rangle - \langle\mathfrak{c}_1|E^\dagger_{k,i}E_{k',j}|\mathfrak{c}_1\rangle)$, so that $\Pi E_i^\dagger E_j \Pi = A_{ij}\Pi + \Pi\delta A_{ij}\Pi$ holds. A detailed calculation shows that $A^{(i,j)} = 0$ when $i \neq j$, $B^{(i,j)} = 0$ when $i + j \neq 3$, and

$$A^{(0,0)} = \frac{1}{4n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \tag{5.78}$$

$$A^{(1,1)} = A^{(2,2)} = \frac{1}{4n}\begin{pmatrix} 1 & \frac{n^2-m^2}{2n(n-1)} & \cdots & \frac{n^2-m^2}{2n(n-1)} \\ \frac{n^2-m^2}{2n(n-1)} & 1 & \cdots & \frac{n^2-m^2}{2n(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n^2-m^2}{2n(n-1)} & \frac{n^2-m^2}{2n(n-1)} & \cdots & 1 \end{pmatrix}, \tag{5.79}$$

$$A^{(3,3)} = \frac{1}{4n}\begin{pmatrix} 1 & \frac{m^2-n}{n(n-1)} & \cdots & \frac{m^2-n}{n(n-1)} \\ \frac{m^2-n}{n(n-1)} & 1 & \cdots & \frac{m^2-n}{n(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{m^2-n}{n(n-1)} & \frac{m^2-n}{n(n-1)} & \cdots & 1 \end{pmatrix}, \tag{5.80}$$

$$B^{(0,3)} = B^{(3,0)} = \frac{m}{4n^2}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \quad B^{(1,2)} = -B^{(2,1)} = i\frac{m}{4n^2}\mathbb{1}. \tag{5.81}$$

Next we note that

$$\begin{aligned} f_B(\mathcal{A}, \mathcal{A}+\delta\mathcal{A}) &= \min_{|\psi\rangle} f_B((\mathcal{A}\otimes \mathbb{1}_R)(|\psi\rangle\langle\psi|), ((\mathcal{A}+\delta\mathcal{A})\otimes \mathbb{1}_R)(|\psi\rangle\langle\psi|)) \\ &= \min_{p_i,\rho_i,i=0,1} f_B(A\otimes(p_0\rho_0+p_1\rho_1), p_0(A+B)\otimes\rho_0 + p_1(A-B)\otimes\rho_1) \\ &\geq \min_{p_i,\rho_i,i=0,1} p_0 f_B(A, A+B) + p_1 f_B(A, A-B) = f_B(A, A+B), \end{aligned} \tag{5.82}$$

where in the second step we define $\langle \mathfrak{c}_i|\psi\rangle\langle\psi|\mathfrak{c}_i\rangle = p_i\rho_i$ for $i = 0, 1$, and in the third step we use the joint concavity of fidelity and in the last step we use $f_B(A+B) = f_B(A-B)$. Therefore we must have

$$f_B(\mathcal{A}, \mathcal{A}+\delta\mathcal{A}) = f_B(A, A+B), \tag{5.83}$$

by noticing that $f_B(\mathcal{A}(|\mathfrak{c}_0\rangle\langle\mathfrak{c}_0|), (\mathcal{A}+\delta\mathcal{A})(|\mathfrak{c}_0\rangle\langle\mathfrak{c}_0|)) = f_B(A, A+B)$. First note that $A^{(i,i)}$ and $B^{(i,j)}$ could be diagonalized in the following way:

$$A^{(0,0)} = \frac{1}{4n}(n|\psi_1\rangle\langle\psi_1|), \quad B^{(0,3)} = B^{(3,0)} = \frac{m}{4n}|\psi_1\rangle\langle\psi_1|, \tag{5.84}$$

$$A^{(1,1)} = A^{(2,2)} = \frac{1}{4n}\Big(\frac{n^2+2n-m^2}{2n}\left|\psi_1\right\rangle\left\langle\psi_1\right| + \frac{n^2-2n+m^2}{2n(n-1)}\sum_{k=2}^{n}\left|\psi_k\right\rangle\left\langle\psi_k\right|\Big), \tag{5.85}$$

$$A^{(3,3)} = \frac{1}{4n}\Big(\frac{m^2}{n}\left|\psi_1\right\rangle\left\langle\psi_1\right| + \frac{n^2-m^2}{n(n-1)}\sum_{k=2}^{n}\left|\psi_k\right\rangle\left\langle\psi_k\right|\Big), \tag{5.86}$$

where $|\psi_1\rangle = \frac{1}{\sqrt{n}}(1\ 1\ \cdots\ 1)$ and $\{|\psi_k\rangle\}_{k>1}$ is an arbitrary orthonormal basis of the orthogonal subspace of $|\psi_1\rangle$. Since $A^{(i,j)} = A^{(j,i)} = B^{(i,j)} = B^{(j,i)} = 0$ when $i \in \{1,2\}$ and $j \in \{0,3\}$, we have

$$f_B(A, A+B) = f_B(A^{(0)}, A^{(0)}+B^{(0)}) + f_B(A^{(1)}, A^{(1)}+B^{(1)}), \tag{5.87}$$

where

$$(\cdot)^{(0)} = \begin{pmatrix} (\cdot)^{(0,0)} & (\cdot)^{(0,3)} \\ (\cdot)^{(3,0)} & (\cdot)^{(3,3)} \end{pmatrix}, \quad (\cdot)^{(1)} = \begin{pmatrix} (\cdot)^{(1,1)} & (\cdot)^{(1,2)} \\ (\cdot)^{(2,1)} & (\cdot)^{(2,2)} \end{pmatrix}. \tag{5.88}$$

We first calculate $f_B(A^{(0)}, A^{(0)}+B^{(0)})$. We have

$$(A^{(0)})^{1/2}(A^{(0)}+B^{(0)})(A^{(0)})^{1/2} = \begin{pmatrix} \frac{1}{4} \\ \frac{m^2}{4n^2} \end{pmatrix}\begin{pmatrix} \frac{1}{4} & \frac{m^2}{4n^2} \end{pmatrix} \otimes \left|\psi_1\right\rangle\left\langle\psi_1\right| + \begin{pmatrix} 0 & 0 \\ 0 & \left(\frac{n^2-m^2}{4n^2(n-1)}\right)^2 \end{pmatrix} \otimes \sum_{k=2}^{n}\left|\psi_k\right\rangle\left\langle\psi_k\right|. \tag{5.89}$$

Then

$$\begin{aligned} f_B(A^{(0)}, A^{(0)}+B^{(0)}) &= \mathrm{Tr}\Big(\big((A^{(0)})^{1/2}(A^{(0)}+B^{(0)})(A^{(0)})^{1/2}\big)^{1/2}\Big) \\ &= \sqrt{\frac{1}{4^2} + \Big(\frac{m^2}{4n^2}\Big)^2} + \frac{n^2-m^2}{4n^2} = \frac{1}{2} - \frac{m^2}{4n^2} + O\Big(\frac{m^4}{n^4}\Big). \end{aligned} \tag{5.90}$$

In order to calculate $f_B(A^{(0)}, A^{(0)}+B^{(0)})$, we first note that

$$(A^{(1)})^{1/2}(A^{(1)}+B^{(1)})(A^{(1)})^{1/2} = \begin{pmatrix} (A^{(1,1)})^2 & 0 \\ 0 & (A^{(1,1)})^2 \end{pmatrix} + \begin{pmatrix} 0 & i\frac{m}{4n^2}A^{(1,1)} \\ -i\frac{m}{4n^2}A^{(1,1)} & 0 \end{pmatrix}. \tag{5.91}$$

Then we use the Taylor expansion formula for square root of positive matrices: $\sqrt{\Lambda^2+Y} = \Lambda + \chi[Y] - \chi(\chi[Y]^2) + O(Y^3)$ for any positive diagonal matrix $\Lambda$ and small $Y$ [Del Moral

and Niclas, 2018], where $\chi[(\cdot)]_{ij} = \frac{(\cdot)_{ij}}{\Lambda_i+\Lambda_j}$. Let $A^{(1)} = \Lambda$ such that $\Lambda_1 = \frac{n^2+2n-m^2}{8n^2}$ and $\Lambda_k = \frac{n^2-2n+m^2}{8n^2(n-1)}$ for $k > 1$, we find that

$$f_B(A^{(1)}, A^{(1)} + B^{(1)}) = \frac{1}{2} - \Big(\frac{m}{4n^2}\Big)^2 \sum_{k=1}^{n} \frac{1}{4\Lambda_k} + O\Big(\frac{m^3}{n^3}\Big) = \frac{1}{2} - \frac{m^2}{8n^2} + O\Big(\frac{m^3}{n^3}\Big). \tag{5.92}$$

Therefore

$$1 - f_B(\mathcal{A}, \mathcal{A} + \delta\mathcal{A})^2 = 1 - f_B(A, A + B)^2 = \frac{3m^2}{4n^2} + O\Big(\frac{m^3}{n^3}\Big), \tag{5.93}$$

which serves as an upper bound on the infidelity of thermodynamic codes under depolarizing noise due to Lemma 2.2. We obtain

$$\varepsilon(\mathcal{M}_S, \mathcal{E}_{S\leftarrow L}) \leq \frac{3m^2}{4n^2} + o\Big(\frac{m^2}{n^2}\Big), \tag{5.94}$$

which also matches the scaling of our lower bound for depolarizing noise channels (Eq. (5.61)), i.e. $\varepsilon(\mathcal{M}_S, \mathcal{E}_{S\leftarrow L}) \geq \frac{3m^2}{8n^2}$.

# Chapter 6

# Summary and Outlook

## 6.1 Summary

Noise limits the precision of quantum metrology. QEC can suppress the damaging effects of noise, but whether QEC improves the efficacy of quantum metrology depends on the structure of the Hamiltonian and the noise. Unless suitable conditions are met, the QEC code that tames the noise might obscure the signal as well, nullifying the advantages of QEC. In this thesis, we studied the interplay between quantum metrology and QEC, namely, how quantum metrology is enhanced by QEC and how QEC is limited by metrological limits.

In Chapter 3, we focused on one-parameter Hamiltonian estimation under Markovian noise where the experimentalist is assumed to have access to noiseless ancillas and fast and frequent quantum controls. We found a necessary and sufficient condition, the HNLS condition, for achieving the HL in terms of the probing time. When HNLS is satisfied, we constructed a two-dimensional QEC code that achieves the HL and presented a geometrical interpretation of the optimal estimation precision, and when HNLS is violated, then we proved that the SQL cannot be surpassed.

We then generalized the HNLS condition to the multi-parameter regime. In scenarios where multi-parameter HNLS is satisfied, we developed an efficient numerical algorithm (SDP) to find the optimal QEC protocol, including the optimal input states, QEC codes and measurements. In contrast to the one-parameter case, the code space has dimension $P+1$, where $P$ is the number of parameter. Our algorithm is applicable to arbitrary system

dynamics (including noiseless cases), which goes beyond previous works focusing on specific quantum dynamics or quantum state estimation.

So far, the error-corrected metrological protocol has focused on optimizing the estimation precision when the HNLS condition is satisfied. When it is slightly violated, it is still possible to enhance metrology significantly in the finite time regime (see the discussion on approximate QEC in [Zhou et al., 2018]). But little was known in general situations. To address this issue systematically, we studied approximate QEC protocols and proposed a new coding technique called the perturbation coding such that the optimal SQL coefficient in one-parameter Hamiltonian estimation under Markovian noise could be achieved asymptotically. Instead of fully correcting noises in the HNLS case, the optimal code in the HLS case achieves a balance between preserving the signal and suppressing the noise.

We also discussed the possibility of removing the noiseless ancilla assumption which is a stringent requirement in experiments and also of theoretical interest. We found that when the Hamiltonian and the noise commutes, it is possible to construct optimal ancilla-free codes which not only recovers the HL but also reaches the optimal estimation precision in one-parameter Hamiltonian estimation under Markovian noise. We provided an example of sensing in lossy bosonic channels where a family of closed-form ancilla-free codes, called Chebyshev codes, was proposed to optimizing the sensitivity.

Hamiltonian estimation under Markovian noise is an important sensing scenario, but the fundamental question is whether the results above are applicable to general quantum channel estimation. We studied the asymptotic theory of quantum channel estimation in Chapter 4, aiming at identifying the asymptotic scaling of the QFI and achieving the optimal QFI coefficients. The key challenges were to define the Hamiltonian and the noise span for arbitrarily parametrized quantum channels and to devise QEC protocols to attain the QFI upper bounds. For single-parameter estimation, we obtained a necessary and sufficient condition, the HNKS condition, for achieving the HL in terms of the number of channels and we also devised a three-step constructive proof to achieve the asymptotic QFI coefficients: (1) we proved the asymptotic QFI for qubit dephasing channels with an arbitrarily encoded parameter are achievable using spin-squeezed states when HNKS fails (It was already known GHZ-type states are optimal when HNKS holds); (2) we showed that

using a two-dimensional QEC protocol, every channel can be reduced to qubit dephasing channels; (3) we proved that by optimizing the encoding and recovery channels, the QFI upper bounds are attainable. The QEC protocols are solvable using SDP. Furthermore, our results implied that sequential strategies provide no asymptotic advantages over parallel strategies in the HKS case, answering another open problem in quantum metrology.

Finally, we established a close connection between covariant QEC and quantum metrology in Chapter 5. We presented covariant QEC as a special type of metrological protocol where parameter estimation limit is linked to the code infidelity. The HNKS condition, as a necessary condition to achieve the HL, is therefore also a necessary condition of the existence of exactly error-correcting covariant codes. Moreover, when HNKS fails, we lower bounded the code infidelity using the inverse of the regularized channel QFI. The lower bound we derived not only has a broader range of applications, but also improves compared to previous lower bounds, leading to an improved approximate Eastin-Knill theorem. It also opens doors to future applications of quantum metrology in other areas of quantum information science.

## 6.2 Outlook

A lot of discoveries were made on error-corrected quantum metrology over the years and here we comment on some related aspects which merit further exploration.

First, recall that in Chapter 4 we showed that sequential strategies cannot outperform parallel strategies asymptotically when the HNKS condition is violated. It is left open, however, whether the statement is still true when HNKS is satisfied. It was known to be true only for unitary channels [Giovannetti et al., 2006], but there is still a gap between the HL QFI coefficient $\mathfrak{F}_{\text{HL}}(\mathcal{E}_\omega)$ for parallel strategies and the state-of-the-art upper bounds on $\mathfrak{F}_{\text{HL}}^{(\text{seq})}(\mathcal{E}_\omega)$ for sequential strategies [Demkowicz-Dobrzański and Maccone, 2014; Sekatski et al., 2017; Yuan and Fung, 2017; Katariya and Wilde, 2020a]. If $\mathfrak{F}_{\text{HL}} < \mathfrak{F}_{\text{HL}}^{(\text{seq})}$, one consequence is that for sensing in open quantum systems with the help of ancillas and fast and frequent quantum controls, e.g. Hamiltonian estimation under Markovian noise, the current QEC sensing protocols introduced in Chapter 3 which only reaches $\mathfrak{F}_{\text{HL}}$ might be

further improved using new types of (possibly time-dependent) quantum controls.

Another important question of interest is multi-parameter quantum channel estimation. In Chapter 4, we consider optimization of QEC protocols for multi-parameter Hamiltonian estimation under Markovian noise when HNLS holds. However, it is not clear yet if the optimal precision obtained from the QEC protocol is in fact optimal among all possible sequential strategies. Moreover, no results on error-corrected metrology were developed in the case when HNLS fails, or in general quantum channel estimation. The study of multi-parameter channel estimation is largely unexplored—for example, unlike the one-parameter case where the expression for the one-shot QFI was founded more than a decade ago, it was not clear how to minimize the weight MSE for an arbitrary quantum channel, let alone the asymptotic case. Note, however, that a multi-parameter bound based on the RLD channel QFI was recently derived [Katariya and Wilde, 2020b]. The extension from one-parameter to multi-parameter estimation is usually highly non-trivial. For example, for multi-parameter estimation, a gap between parallel strategies and sequential strategies exists even for unitary channels [Yuan, 2016]. Another interesting feature in multi-parameter estimation to take into account is the scaling of the MSE with respect to the number of parameters [Imai and Fujiwara, 2007; Humphreys et al., 2013; Imai and Fujiwara, 2007; Yuan, 2016; Górecki et al., 2020].

Lastly, we list some practical concerns on error-corrected quantum metrology. First, similar to QEC in quantum information processing, in quantum metrology, noise in the QEC procedure should not be ignored in practical applications, in which case fault-tolerant QEC protocols [Kapourniotis and Datta, 2019] must be devised in order to improve the estimation precision as expected. However, the error threshold which determines the point where quantum strategies beats classical strategies in terms of estimation precision shall depend on the structure of the signal and the noise as well. Second, the current optimal QEC protocol suffers from to drawbacks: the requirement of noiseless ancilla and the perturbation nature in the HKS case. One may consider removing the noiseless ancilla assumption by consider special types of noise [Layden et al., 2019]; on the other hand, solving the code optimization problem in the ancilla-free case is also of theoretical interest. The perturbation coding grants us mathematical simplification in terms of proving the attainability of the

SQL upper bounds, but requires a very long probing time until it reaches the asymptotic limit (because the signal and the noise are both weak for perturbation codes). Therefore, for practical applications one may need to consider other more resource efficient coding probably through numerical optimization [Liu and Yuan, 2017b,a; Chabuda et al., 2020; Koczor et al., 2020; Meyer et al., 2020; Beckey et al., 2020]. Finally, the two asymptotic limits should be treated carefully in real-world quantum sensing: the limit where the number of channels is infinitely large and the limit where the number of repeated experiments is infinitely large. Both limits may not be reachable in practice and the QFI may not be entirely meaningful in those cases. To tackle this problem, one may need to consider the second-order asymptotics [Tomamichel and Hayashi, 2013; Li et al., 2014] or simply the non-asymptotic sample complexity of quantum states [Haah et al., 2017; O'Donnell and Wright, 2016; Aaronson, 2019; Huang et al., 2020].

# Appendix A

# Perturbative expansion of the noise rate

To obtain Eq. (3.100), we expand the minimum noise rate $\gamma$ around $\varepsilon = 0$ using the perturbation code. For simplicity, we ignoring all $o(\varepsilon^2)$ terms in the following equations in this subsection and the *equal sign* "=" means approximate equality up to the second order of $\varepsilon$. We will also use the following lemma:

**Lemma A.1** ([Mirsky, 1960]). *$\|X + \varepsilon Y\|_1 = \|X\|_1 + O(\varepsilon)$ for arbitrary $X$ and $Y$.*

To calculate Eq. (3.98), we first compute the second term,

$$-\operatorname{Re}\left[\sum_{i=1}^{r} \langle \mathfrak{c}_0| \left(\mathcal{P}(J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger) - \frac{1}{2}\{\mathcal{P}(J_i^\dagger J_i), |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1|\}\right) |\mathfrak{c}_1\rangle\right] = \sum_i \lambda_i + \varepsilon^2 \left|\operatorname{Tr}(\tilde{C} J_i)\right|^2 + \varepsilon^2 \operatorname{Tr}(DD^\dagger J_i^\dagger J_i). \quad \text{(A.1)}$$

The remaining first term is equal to (thanks to Lemma 1) minus

$$\left\| \sum_i \Pi_{\mathfrak{c}}^{\perp} J_i |\mathfrak{c}_0\rangle \langle \mathfrak{c}_1| J_i^\dagger \Pi_{\mathfrak{c}}^{\perp} \right\|_1 = \left\| \begin{pmatrix} \sqrt{\tilde{\Lambda}^{-1}}(\Lambda + \varepsilon X_1 + \varepsilon^2 X_1') \\ \varepsilon X_2 + \varepsilon^2 X_2' \end{pmatrix} \begin{pmatrix} (\Lambda - \varepsilon X_1 + \varepsilon^2 X_1')^\dagger \sqrt{\tilde{\Lambda}^{-1}} & -\varepsilon X_2^\dagger + \varepsilon^2 X_2'^\dagger \end{pmatrix} \right\|_1, \quad \text{(A.2)}$$

where $\Lambda \in \mathbb{R}^{r\times r}$ is a diagonal matrix whose $k$-th diagonal element is $\lambda_k$ and $\tilde{\Lambda} \in \mathbb{R}^{r\times r}$ is

a diagonal matrix whose $k$-th diagonal element is $\lambda_k$ if $\lambda_k > 0$ and 1 if $\lambda_k = 0$. Assume $\{\lambda_k\}_{k=1}^r$ is arranged in a non-ascending order and $r_0$ is the largest integer such that $\lambda_{r_0}$ is positive. $X_1, X_1' \in \mathbb{C}^{r\times r}$ satisfy

$$\begin{aligned}(\Lambda + \varepsilon X_1 + \varepsilon^2 X_1')_{ji} &= \sqrt{\lambda_j}\,\langle \tilde{J}_{j,0}|\,\Pi_{\mathfrak{c}}^{\perp} J_i\,|\mathfrak{c}_0\rangle = \mathrm{Tr}(C^\dagger J_j^\dagger J_i A_0) - \mathrm{Tr}(C^\dagger J_j^\dagger A_0)\mathrm{Tr}(A_0^\dagger J_i A_0)\\ &= \lambda_i\delta_{ij} + \varepsilon\mathrm{Tr}(C^\dagger J_j^\dagger J_i D) - \varepsilon^2\mathrm{Tr}(C^\dagger J_j^\dagger D)\mathrm{Tr}(\tilde{C}J_i),\end{aligned} \tag{A.3}$$

for $1 \le j \le r_0$ and

$$\begin{aligned}(\Lambda + \varepsilon X_1 + \varepsilon^2 X_1')_{ji} &= \langle \tilde{J}_{j,0}|\,\Pi_{\mathfrak{c}}^{\perp} J_i\,|\mathfrak{c}_0\rangle = \mathrm{Tr}(\tilde{J}_j^\dagger J_i A_0) - \mathrm{Tr}(\tilde{J}_j^\dagger A_0)\mathrm{Tr}(A_0^\dagger J_i A_0)\\ &= \varepsilon\mathrm{Tr}(\tilde{J}_j^\dagger J_i D) - \varepsilon^2\mathrm{Tr}(\tilde{J}_j^\dagger D)\mathrm{Tr}(\tilde{C}J_i),\end{aligned} \tag{A.4}$$

for $r_0 + 1 \le j \le r$. $X_2, X_2' \in \mathbb{C}^{(d^2-r)\times r}$ satisfy

$$\begin{aligned}(\Lambda + \varepsilon X_2 + \varepsilon^2 X_2')_{(j-r)i} &= \langle \tilde{J}_{j,0}|\,\Pi_{\mathfrak{c}}^{\perp} J_i\,|\mathfrak{c}_0\rangle = \mathrm{Tr}(\tilde{J}_j^\dagger J_i A_0) - \mathrm{Tr}(\tilde{J}_j^\dagger A_0)\mathrm{Tr}(A_0^\dagger J_i A_0)\\ &= \varepsilon\mathrm{Tr}(\tilde{J}_j^\dagger J_i D) - \varepsilon^2\mathrm{Tr}(\tilde{J}_j^\dagger D)\mathrm{Tr}(\tilde{C}J_i),\end{aligned} \tag{A.5}$$

for $r + 1 \le j \le d^2 - 1$ and

$$\begin{aligned}(\Lambda + \varepsilon X_2 + \varepsilon^2 X_2')_{(d^2-r)i} &= \langle \tilde{J}_{d^2,0}|\,\Pi_{\mathfrak{c}}^{\perp} J_i\,|\mathfrak{c}_0\rangle = \mathrm{Tr}(\tilde{J}_{d^2}^\dagger J_i A_0) - \mathrm{Tr}(\tilde{J}_{d^2}^\dagger A_0)\mathrm{Tr}(A_0^\dagger J_i A_0)\\ &= \varepsilon\mathrm{Tr}(C^\dagger J_i D) - \varepsilon\mathrm{Tr}(\tilde{C}J_i).\end{aligned} \tag{A.6}$$

Here

$$|\tilde{J}_{j,0/1}\rangle = \begin{cases} \frac{1}{\sqrt{\lambda_j}}\sum_{ik} C_{ik} J_j\,|i\rangle\,|k,0/1\rangle\,, & j \le r_0,\\ \sum_{ik}(\tilde{J}_j)_{ik}\,|i\rangle\,|k,0/1\rangle\,, & r_0 < j \le d^2,\\ \sum_{ik}\frac{C_{ik}}{\sqrt{\mathrm{Tr}(C^\dagger C)}}\,|i\rangle\,|k,0/1\rangle\,, & j = d^2,\end{cases} \tag{A.7}$$

are two sets of orthonormal basis of $\mathcal{H}_S \otimes \mathcal{H}_A$.

To calculate the first and second order expansion of Eq. (A.2), we consider the singular

value decompositions

$$\begin{aligned}\begin{pmatrix}\sqrt{\tilde{\Lambda}^{-1}}(\Lambda+\varepsilon X_1+\varepsilon^2 X_1') \\ \varepsilon X_2+\varepsilon^2 X_2'\end{pmatrix} &= U(\varepsilon)\begin{pmatrix}\Sigma(\varepsilon) \\ 0\end{pmatrix}V(\varepsilon)^\dagger, \\ \begin{pmatrix}(\Lambda-\varepsilon X_1+\varepsilon^2 X_1')^\dagger\sqrt{\tilde{\Lambda}^{-1}} & -\varepsilon X_2^\dagger+\varepsilon^2 X_2'^\dagger\end{pmatrix} &= V(-\varepsilon)\begin{pmatrix}\Sigma(-\varepsilon) & 0\end{pmatrix}U(-\varepsilon)^\dagger,\end{aligned} \tag{A.8}$$

Then

$$\begin{aligned}\text{Eq. (A.2)} &= \left\| U(\varepsilon)\begin{pmatrix}\Sigma(\varepsilon)V(\varepsilon)^\dagger V(-\varepsilon)\Sigma(-\varepsilon) & 0 \\ 0 & 0\end{pmatrix}U(-\varepsilon)^\dagger\right\|_1 \\ &= \mathrm{Tr}\left(\sqrt{\sqrt{Y(\varepsilon)}Y(-\varepsilon)\sqrt{Y(\varepsilon)}}\right),\end{aligned} \tag{A.9}$$

where

$$\begin{aligned}Y(\varepsilon) &= V(\varepsilon)\Sigma(\varepsilon)^2 V(\varepsilon)^\dagger \\ &= \Lambda+\varepsilon(X_1^\dagger\Pi_\Lambda+\Pi_\Lambda X_1)+\varepsilon^2(X_1^\dagger\tilde{\Lambda}^{-1}X_1+X_1'\Pi_\Lambda+\Pi_\Lambda X_1'^\dagger+X_2^\dagger X_2),\end{aligned} \tag{A.10}$$

and $\Pi_\Lambda$ is the projector onto the support of $\Lambda$.

Using Theorem 2 in [Zhou and Jiang, 2019], we have

$$\begin{aligned}\mathrm{Tr}\left(\sqrt{\sqrt{Y(\varepsilon)}Y(-\varepsilon)\sqrt{Y(\varepsilon)}}\right) &= \mathrm{Tr}(\Lambda)+ \\ \varepsilon^2\mathrm{Tr}(X_1^\dagger\tilde{\Lambda}^{-1}X_1+X_1'\Pi_\Lambda+\Pi_\Lambda X_1'^\dagger+X_2^\dagger X_2) &- \varepsilon^2\sum_{i,j:\lambda_i+\lambda_j\neq 0}^{r}\frac{|X_{1,ij}+X_{1,ji}^*|^2}{\lambda_i+\lambda_j}.\end{aligned} \tag{A.11}$$

A few lines of calculation reveals:

$$\begin{aligned}\gamma &= \text{Eq. (A.1)} - \text{Eq. (A.2)} \\ &= 2\varepsilon^2\sum_i\left|\mathrm{Tr}(J_i\tilde{C})\right|^2+\varepsilon^2\sum_{ij:\lambda_i+\lambda_j\neq 0}\frac{|\mathrm{Tr}(J_i^\dagger J_j\tilde{C})|^2}{(\lambda_i+\lambda_j)},\end{aligned} \tag{A.12}$$

where $\tilde{C}=CD^\dagger+DC^\dagger$.

# Appendix B

# Validity of the numerical algorithm in Section 3.5.2

In this appendix, we show $(C^\diamond, \tilde{C}^\diamond)$ is indeed a solution of the optimal code and the algorithm in Section 3.5.2 is valid. Let $(\mathbf{g}^\Delta, \mathfrak{g}^\Delta, C^\Delta)$ be the saddle point of Eq. (3.102). Then

$$\mathrm{Tr}(C^{\Delta\dagger}\alpha^\Delta C^\Delta) = \|\alpha^\Delta\| = \min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\|\,, \tag{B.1}$$

which means that $\Pi^\Delta C^\Delta = C^\Delta$ where $\Pi^\Delta$ is the projection onto the subspace spanned by all eigenstates corresponding to the largest eigenvalue of $\alpha^\Delta$.

Now assume we have a solution $(\mathbf{g}^\diamond, \mathfrak{g}^\diamond)$ of Eq. (3.81) such that $\alpha^\diamond = (\mathbf{g}^\diamond \mathbb{1} + \mathfrak{g}^\diamond \mathbf{L})^\dagger (\mathbf{g}^\diamond \mathbb{1} + \mathfrak{g}^\diamond \mathbf{L})$ satisfies $\|\alpha^\diamond\| = \min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\|$. We prove that $(\mathbf{g}^\diamond, \mathfrak{g}^\diamond, C^\Delta)$ is also a saddle point. Choose $p \in (0,1)$ and let $(\mathbf{g}, \mathfrak{g}) = (p\mathbf{g}^\diamond + (1-p)\mathbf{g}^\Delta, p\mathfrak{g}^\diamond + (1-p)\mathfrak{g}^\Delta)$. Then

$$\begin{aligned}
\mathrm{Tr}(C^{\Delta\dagger}\alpha C^\Delta) &= p^2 \mathrm{Tr}(C^{\Delta\dagger}\alpha^\diamond C^\Delta) + (1-p)^2 \mathrm{Tr}(C^{\Delta\dagger}\alpha^\Delta C^\Delta) \\
&\quad + 2p(1-p)\mathrm{Re}[\mathrm{Tr}(C^{\Delta\dagger}(\mathbf{g}^\diamond \mathbb{1} + \mathfrak{g}^\diamond \mathbf{L})^\dagger (\mathbf{g}^\Delta \mathbb{1} + \mathfrak{g}^\Delta \mathbf{L}) C^\Delta)] \\
&\leq p^2 \mathrm{Tr}(C^{\Delta\dagger}\alpha^\diamond C^\Delta) + (1-p)^2 \mathrm{Tr}(C^{\Delta\dagger}\alpha^\Delta C^\Delta) \\
&\quad + 2p(1-p)\sqrt{\mathrm{Tr}(C^{\Delta\dagger}\alpha^\diamond C^\Delta)\mathrm{Tr}(C^{\Delta\dagger}\alpha^\Delta C^\Delta)} \leq \|\alpha^\Delta\|\,.
\end{aligned} \tag{B.2}$$

On the other hand, we know $\mathrm{Tr}(C^{\Delta\dagger}\alpha C^\Delta) \geq \|\alpha^\Delta\|$. Therefore the equality in Eq. (B.2) must

hold, which means

$$\mathrm{Tr}(C^{\Delta\dagger}\alpha^{\diamond}C^{\Delta}) = \|\alpha^{\diamond}\|\,, \quad (\mathbf{g}^{\Delta}\mathbb{1} + \mathfrak{g}^{\Delta}\mathbf{L})C^{\Delta} = (\mathbf{g}^{\diamond}\mathbb{1} + \mathfrak{g}^{\diamond}\mathbf{L})C^{\Delta}. \tag{B.3}$$

As a result, we have $\mathrm{Tr}(C^{\dagger}\alpha^{\diamond}C) \leq \mathrm{Tr}(C^{\Delta\dagger}\alpha^{\diamond}C^{\Delta})$ for arbitrary $C$ satisfying $\mathrm{Tr}(C^{\dagger}C) = 1$. Moreover, $\mathrm{Re}[\mathrm{Tr}(C^{\Delta\dagger}(\Delta\mathbf{g}\mathbb{1} + \Delta\mathfrak{g}\mathbf{L})^{\dagger}(\mathbf{g}^{\diamond}\mathbb{1} + \mathfrak{g}^{\diamond}\mathbf{L})C^{\Delta})] = \mathrm{Re}[\mathrm{Tr}(C^{\Delta\dagger}(\Delta\mathbf{g}\mathbb{1} + \Delta\mathfrak{g}\mathbf{L})^{\dagger}(\mathbf{g}^{\Delta}\mathbb{1} + \mathfrak{g}^{\Delta}\mathbf{L})C^{\Delta})] = 0$, and $\mathrm{Tr}(C^{\Delta\dagger}\alpha^{\diamond}C^{\Delta}) \leq \mathrm{Tr}(C^{\Delta\dagger}\alpha C^{\Delta})$, proving $(\mathbf{g}^{\diamond}, \mathfrak{g}^{\diamond}, C^{\Delta})$ is also a saddle point. Hence, step (2) in our algorithm will at least have one solution $C^{\Delta}$, and the solution of step (2) $(\mathbf{g}^{\diamond}, \mathfrak{g}^{\diamond}, C^{\diamond})$ is also a saddle point satisfying

$$\mathrm{Tr}(C^{\diamond\dagger}\alpha C^{\diamond}) \leq \mathrm{Tr}(C^{\diamond\dagger}\alpha^{\diamond}C^{\diamond}) \leq \mathrm{Tr}(C^{\diamond\dagger}\alpha C^{\diamond}), \tag{B.4}$$

for all $(g, \mathbf{g}, \mathfrak{g}, C)$ satisfying $\beta = 0$ and $\mathrm{Tr}(C^{\dagger}C) = 1$. Strong duality [Boyd and Vandenberghe, 2004] implies the optimal value of

$$\begin{aligned} &\max_{\tilde{C}} F_{\mathrm{SQL}}(C^{\diamond}, \tilde{C}), \\ &\text{subject to } \mathrm{Tr}(C^{\dagger}C) = 1,\ \mathrm{Tr}(\tilde{C}) = 0 \text{ and } \mathrm{Tr}(J_i^{\dagger}J_j\tilde{C}) = 0, \forall i, j \in \mathfrak{n}, \end{aligned} \tag{B.5}$$

is equal to that of $\min_{g,\mathbf{g},\mathfrak{g}|\beta=0}\ 4\mathrm{Tr}(C^{\diamond\dagger}\alpha C^{\diamond}) = \min_{g,\mathbf{g},\mathfrak{g}|\beta=0} \|\alpha\|$, proving the optimality of $(C^{\diamond}, \tilde{C}^{\diamond})$.

# Appendix C

# An example where noiseless ancilla is necessary

Consider Gell-Mann matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{C.1}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \tag{C.2}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{C.3}$$

and $\lambda_0$ is the identity matrix. The Hilbert space $\mathcal{H}_S = \mathcal{H}_3 \oplus \mathcal{H}_{d-3}$ is the direct sum of a 3-dimensional and a $(d-3)$-dimensional Hilbert space. Let $H = \lambda_5 \oplus 0_{d-3}$ and $L_i = \lambda_i \oplus 0_{d-3}$

with $i = 1, 2, 4$ where $0_i$ means a $i$-dimensional zero matrix. One can check that

$$\mathcal{S} = \text{span}_{\mathbb{H}}\{\mathbb{1}, \lambda_i \oplus 0_{d-3},\ i = 0, 1, 2, 3, 4, 6, 7, 8\} \tag{C.4}$$

then the HNLS condition $H \notin \mathcal{S}$ is satisfied. Suppose we have a two-dimensional QEC sensing code

$$|\mathfrak{c}_0\rangle = \alpha_3^0 |\mathfrak{c}_0^3\rangle + \alpha_{d-3}^0 |\mathfrak{c}_0^{d-3}\rangle, \quad |\mathfrak{c}_1\rangle = \alpha_3^1 |\mathfrak{c}_1^3\rangle + \alpha_{d-3}^1 |\mathfrak{c}_1^{d-3}\rangle. \tag{C.5}$$

where $|\mathfrak{c}_{0(1)}^3\rangle \in \mathcal{H}_3$ and $|\mathfrak{c}_{0(1)}^{d-3}\rangle \in \mathcal{H}_{d-3}$. First of all, we note that $\alpha_3^0$ and $\alpha_3^1$ are not both zero because $\Pi_\mathfrak{c} H \Pi_\mathfrak{c} \not\propto \Pi_\mathfrak{c}$. If $\alpha_3^0 = 0$, due to the error correction condition $\Pi_\mathfrak{c} L_i \Pi_\mathfrak{c} \propto \Pi_\mathfrak{c}$ and $\Pi_\mathfrak{c} L_i^\dagger L_j \Pi_\mathfrak{c} \propto \Pi_\mathfrak{c}$, we must have

$$\langle 1_3 | \lambda_i | 1_3 \rangle = 0, \quad i = 1, 2, 3, 4, 6, 7, 8, \tag{C.6}$$

leading to $|\mathfrak{c}_1^3\rangle = 0$. Therefore, we conclude that $\alpha_3^0$ and $\alpha_3^1$ are both non-zero. In this case we must have

$$\langle 0_3 | \lambda_i | 1_3 \rangle = 0, \quad i = 1, 2, 3, 4, 6, 7, 8, \tag{C.7}$$

which again could not be satisfied for non-zero $|\mathfrak{c}_{0(1)}^3\rangle$. Therefore we conclude that a valid QEC sensing that satisfies Eq. (3.116) and Eq. (3.117) does not exist, without noiseless ancilla. The dimension $d$ of the Hilbert space $\mathcal{H}$ could be arbitrary large compared to the number of noise operators $\dim \mathcal{S} = 9$, yet there is no valid QEC code correcting noise and preserving signal simultaneously.

# Appendix D

# Exact coefficients and near-optimality of Chebyshev codes

In this appendix, we provide the exact value of $\tilde{c}_k$ and prove that the Chebyshev code Eq. (3.127) indeed corrects the Lindblad span Eq. (3.126) and leads to a near-optimal QFI. To do so, we will use the following Lemma:

**Lemma D.1.** *Suppose $s$ is an integer larger than one. Then we have*

$$\sum_{k=0}^{s}(-1)^k |c_k|^2 \left(\sin\frac{k\pi}{2s}\right)^{2i} = 0, \quad \forall 1 \le i \le s-1, \tag{D.1}$$

*and*

$$\sum_{k=0}^{s}(-1)^k |c_k|^2 \left(\sin\frac{k\pi}{2s}\right)^{2s} = \frac{(-1)^s}{2^{2s-2}}, \tag{D.2}$$

*where $|c_k|^2 = \frac{2}{s} - \frac{1}{s}\delta_{k0} - \frac{1}{s}\delta_{ks}$.*

*Proof.* We first notice that for all $0 \le \ell \le s-1$,

$$\sum_{k=0}^{s-1}(-1)^k \cos\frac{k\ell\pi}{s} = \mathrm{Re}\left[\sum_{k=0}^{s-1} e^{i\left(\frac{k\ell\pi}{s}+k\pi\right)}\right] = \mathrm{Re}\left[\frac{1-(-1)^{s+\ell}}{1+e^{i\frac{\ell\pi}{s}}}\right] = \frac{1+(-1)^{s+\ell+1}}{2}, \tag{D.3}$$

which only depends on the parity of $\ell$. Then we have

$$\sum_{k=0}^{s-1}(-1)^k\left(\cos\frac{k\pi}{s}\right)^\ell=\frac{1}{2^\ell}\sum_{k=0}^{s-1}(-1)^k\sum_{j=0}^{\ell}\binom{\ell}{j}\cos\frac{(2j-\ell)k\pi}{s}=\frac{1+(-1)^{s+\ell+1}}{2}. \tag{D.4}$$

When $\ell=s$,

$$\sum_{k=0}^{s-1}(-1)^k\left(\cos\frac{k\pi}{s}\right)^s=\frac{1}{2^\ell}\sum_{k=0}^{s-1}(-1)^k\sum_{j=0}^{s}\binom{s}{j}\cos\frac{(2j-s)k\pi}{s}=\frac{s}{2^{2s-1}}. \tag{D.5}$$

Therefore,

$$\begin{aligned}\frac{1}{2^i}\sum_{k=0}^{s-1}(-1)^k\left(1-\cos\frac{k\pi}{s}\right)^i&=\frac{1}{2^i}\sum_{k=0}^{s-1}(-1)^k\sum_{\ell=0}^{i}\binom{i}{\ell}(-1)^\ell\left(\cos\frac{k\pi}{s}\right)^\ell\\&=\frac{1}{2^i}\sum_{\ell=0}^{i}\binom{i}{\ell}(-1)^\ell\left(\frac{1+(-1)^{s+\ell+1}}{2}\right)=\frac{(-1)^{s+1}}{2},\end{aligned} \tag{D.6}$$

and

$$\begin{aligned}\frac{1}{2^s}\sum_{k=0}^{s-1}(-1)^k\left(1-\cos\frac{k\pi}{s}\right)^s&=\frac{1}{2^s}\sum_{k=0}^{s-1}(-1)^k\sum_{\ell=0}^{s}\binom{s}{\ell}(-1)^\ell\left(\cos\frac{k\pi}{s}\right)^\ell\\&=\frac{(-1)^{s+1}}{2}+\frac{s(-1)^s}{2^{2s-1}}.\end{aligned} \tag{D.7}$$

As a result, when $1\le i\le s-1$

$$\sum_{k=0}^{s}(-1)^k|c_k|^2\left(\sin\frac{k\pi}{2s}\right)^{2i}=\frac{2}{s}\left(\sum_{k=0}^{s-1}(-1)^k\left(\sin\frac{k\pi}{2s}\right)^{2i}+\frac{(-1)^s}{2}\right)=0; \tag{D.8}$$

and when $i=s$,

$$\sum_{k=0}^{s}(-1)^k|c_k|^2\left(\sin\frac{k\pi}{2s}\right)^{2s}=\frac{2}{s}\left(\sum_{k=0}^{s-1}(-1)^k\left(\sin\frac{k\pi}{2s}\right)^{2s}+\frac{(-1)^s}{2}\right)=4\left(\frac{-1}{4}\right)^s. \tag{D.9}$$

□

The $s$-th order Chebyshev code should be capable of correcting the Lindblad span $\mathcal{S}=\mathrm{span}_{\mathbb{H}}\{\mathbb{1},a,a^\dagger,(a^\dagger a)^i,\forall 1\le i\le s-1\}$. To correct the off-diagonal noise $a^i$ and $(a^\dagger)^i$ up to

$1 \leq i \leq t$, we simply need the distance between $|\mathfrak{c}_0\rangle$ and $|\mathfrak{c}_1\rangle$ defined by

$$\operatorname{dist}(|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle) = \min_{\substack{m_0,m_1\in\mathbb{N},\\ \langle m_0|\mathfrak{c}_0\rangle^2>0,\, \langle m_1|\mathfrak{c}_1\rangle^2>0}} |m_0 - m_1| \tag{D.10}$$

is larger than $2t+1$. Note that

$$\operatorname{dist}(|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle) \geq M \sin^2\left(\frac{\pi}{2s}\right) - 1, \tag{D.11}$$

so the off-diagonal noise can be corrected as long as $M \sin^2 \frac{\pi}{s} \geq 2(t+1)$. Particularly, when $t = 1$, we only need $\operatorname{dist}(|\mathfrak{c}_0\rangle, |\mathfrak{c}_1\rangle) \geq 3$, or $M \sin^2\left(\frac{\pi}{s}\right) \geq 4$. In fact, from the point of view of quantum memories [Michael et al., 2016], the $s$-th order Chebyshev code could correct $s-1$ dephasing events, $L$ photon losses and $G$ gains, when $L + G = \left\lfloor \frac{M}{2} \sin^2\left(\frac{\pi}{s}\right) - 1 \right\rfloor$.

To correct the diagonal noise $(a^\dagger a)^i$ for $1 \leq i \leq s-1$, we simply need to choose a suitbale $\{|c_k|^2\}_{k=0}^s$ to satisfy the following $s-1$ equations

$$\langle \mathfrak{c}_0|(a^\dagger a)^i|\mathfrak{c}_0\rangle - \langle \mathfrak{c}_1|(a^\dagger a)^i|\mathfrak{c}_1\rangle = \sum_{k=0}^{s} (-1)^k \tilde{c}_k^2 \left\lfloor M \sin^2\left(\frac{k\pi}{2s}\right)\right\rfloor^{2i} = 0, \tag{D.12}$$

and

$$\sum_{k=0}^{s} (-1)^k |\tilde{c}_k|^2 = 0, \quad \sum_{k=0}^{s} |\tilde{c}_k|^2 = 2. \tag{D.13}$$

The linear system of equations could be written as $\tilde{A}\,\tilde{\boldsymbol{c}} = \boldsymbol{e}$, where $\tilde{\boldsymbol{c}} = (|\tilde{c}_0|^2\ |\tilde{c}_1|^2 \cdots |\tilde{c}_s|^2)^T$, $\boldsymbol{e} = (0 \cdots 0\, 1)^T$, $\tilde{A}$ is a $s+1$ by $s+1$ matrix $\tilde{A}_{ik} = (-1)^k \left\lfloor M \sin^2 \frac{k\pi}{2s} \right\rfloor^i / M^i$ when $0 \leq i \leq s-1$ (we assume $0^0 = 1$) and $\tilde{A}_{sk} = 1$. The linear equations are solvable since $A$ is invertible, which also proves the Chebyshev codes defined using its solution must correct the Lindblad span Eq. (3.126) as required.

Next we show the near-optimality of the Chebyshev code. First we calculate an upper bound of the optimal asymptotic QFI Eq. (3.115), since

$$\begin{aligned} \left\|(a^\dagger a)^s - \mathcal{S}\right\| &= M^s \min_{\forall \chi_i \in \mathbb{R}} \max_{\substack{k\in\mathbb{Z},\\ k\in[0,M]}} \left| \left(\frac{k}{M}\right)^s - \sum_{i=0}^{s-1} \chi_i \left(\frac{k}{M}\right)^i \right| \\ &\leq \left(\frac{M}{2}\right)^s \min_{\forall \chi_i \in \mathbb{R}} \max_{x\in[-1,1]} \left| x^s - \sum_{i=0}^{s-1} \chi_i x^i \right| = 2\left(\frac{M}{4}\right)^s, \end{aligned} \tag{D.14}$$

we have

$$F_{\text{opt}}(T) = 4T^2 \min_{S\in\mathcal{S}} \left\|(a^\dagger a)^s - S\right\|_\infty^2 \leq 16T^2 \left(\frac{M}{4}\right)^{2s} \equiv F_{\text{opt}}^\infty(t). \tag{D.15}$$

According to Lemma D.1,

$$\sum_{k=0}^{s}(-1)^k c_k^2 \left(\sin^2\frac{k\pi}{2s}\right)^i = 0 \implies A\,\boldsymbol{c} = \boldsymbol{e}, \tag{D.16}$$

where $\boldsymbol{c} = (|c_0|^2\ |c_1|^2\ \cdots\ |c_s|^2)^T$, $A_{ik} = (-1)^k \left(\sin^2\frac{k\pi}{2s}\right)^i$ when $0 \leq i \leq s-1$ and $\tilde{A}_{sk} = 1$. Note that

$$\sin^2\frac{k\pi}{2s} - \frac{1}{M} \leq \frac{\left\lfloor M\sin^2\left(\frac{k\pi}{2s}\right)\right\rfloor}{M} \leq \sin^2\left(\frac{k\pi}{2s}\right). \tag{D.17}$$

As $M$ becomes sufficiently large, we have

$$\tilde{\boldsymbol{c}} = \tilde{A}^{-1}\boldsymbol{e} = A + (\tilde{A}-A)^{-1}\boldsymbol{e} = (I + A^{-1}(\tilde{A}-A))^{-1}\boldsymbol{c} = \boldsymbol{c} + O\left(\frac{1}{M}\right). \tag{D.18}$$

Then

$$\begin{aligned}
&\langle\mathfrak{c}_0|(a^\dagger a)^s|\mathfrak{c}_0\rangle - \langle\mathfrak{c}_1|(a^\dagger a)^s|\mathfrak{c}_1\rangle \\
&\quad= \frac{(-M)^s}{2^{2s-2}} + \sum_{k=0}^{s}(-1)^k\left(\tilde{c}_k^2\left\lfloor M\sin^2\frac{k\pi}{2s}\right\rfloor^s - c_k^2\left(M\sin^2\frac{k\pi}{2s}\right)^s\right) \\
&\quad\geq \frac{(-M)^s}{2^{2s-2}} - M^s\left\|\boldsymbol{c}-\tilde{\boldsymbol{c}}\right\|_2^2
\end{aligned} \tag{D.19}$$

where $\|\boldsymbol{c}-\tilde{\boldsymbol{c}}\|_2^2 = \sum_{k=0}^s |\tilde{c}_k - c_k|^2$ is the two-norm. Therefore

$$\frac{F_{\text{opt}}(T) - F(T)}{F_{\text{opt}}(T)} \leq 1 - \frac{F(T)}{F_{\text{opt}}^\infty(t)} = O\left(\frac{1}{M^2}\right), \tag{D.20}$$

where $F(T)$ denotes the QFI obtained using the input state $\frac{1}{\sqrt{2}}(|\mathfrak{c}_0\rangle + |\mathfrak{c}_1\rangle)$, proving its near-optimality. The numerical value of $F(T)/F_{\text{opt}}^\infty(t)$ is plotted in Figure 3.5 as a lower bound of $F(T)/F_{\text{opt}}(T)$.

Consider the $(s-1, \frac{M}{s}-1)$ binomial code (suppose $M$ is a multiple of $s$) [Michael et al.,

2016]

$$|\mathfrak{c}_0^{\text{bin}}/\mathfrak{c}_1^{\text{bin}}\rangle = \sum_{k \text{ even/odd}}^{[0,s]} \frac{1}{\sqrt{2^{s-1}}} \sqrt{\binom{s}{k}} \left|\frac{k}{s}M\right\rangle. \tag{D.21}$$

We have

$$\begin{aligned}
&\langle \mathfrak{c}_0^{\text{bin}}| (a^\dagger a)^\ell |\mathfrak{c}_0^{\text{bin}}\rangle - \langle \mathfrak{c}_1^{\text{bin}}| (a^\dagger a)^\ell |\mathfrak{c}_1^{\text{bin}}\rangle \\
&= \frac{M^\ell}{2^{s-1}s^\ell} \sum_{k=0}^{s} \binom{s}{k} k^\ell (-1)^k \\
&= \frac{M^\ell}{2^{s-1}s^\ell} \left(x\frac{d}{dx}\right)^\ell (1+x)^s \Big|_{x=-1} = \begin{cases} 0, & \ell = 1, \dots, s-1, \\ \frac{(-1)^s s! M^s}{2^{s-1}s^s}, & \ell = s. \end{cases}
\end{aligned} \tag{D.22}$$

Clearly the $(s-1, \frac{M}{s}-1)$ binomial code also corrects the Lindblad span, but the strength of the signal is exponentially smaller with respect to $s$:

$$\frac{F^{\text{bin}}(t)}{F_{\text{opt}}(T)} \approx \left(\frac{2^{s-1}s!}{s^s}\right)^2 = O\left(s\left(\frac{2}{e}\right)^{2s}\right). \tag{D.23}$$

# Appendix E

# Optimizing the recovery channel when HNKS fails

To derive Eqs. (4.51)-(4.53), we first expand $T$ and $E_0E_1^\dagger$ around $\varepsilon = 0$

$$T = e^{i\varepsilon G} = 1 + i\varepsilon G - \frac{\varepsilon^2}{2}G^2 + O(\varepsilon^3), \tag{E.1}$$

$$E_0E_1^\dagger = \sigma - i\varepsilon\tilde{\sigma} - \varepsilon^2(FF^\dagger + EE^\dagger) + O(\varepsilon^3), \tag{E.2}$$

where $\sigma = EE^\dagger$ and $\tilde{\sigma} = i(FE^\dagger - EF^\dagger)$. Then

$$\mathrm{Tr}(TE_0E_1^\dagger) = 1 - 2\varepsilon^2 - \frac{\varepsilon^2}{2}\mathrm{Tr}(G^2\sigma) + i\varepsilon\mathrm{Tr}(G\sigma) + \varepsilon^2\mathrm{Tr}(G\tilde{\sigma}) + O(\varepsilon^3), \tag{E.3}$$

$$\mathrm{Tr}(T(\dot{E}_0E_1^\dagger + E_0\dot{E}_1^\dagger)) = i\varepsilon\mathrm{Tr}(G\dot{\sigma}) + O(\varepsilon^2), \tag{E.4}$$

where we used $\mathrm{Tr}(F^\dagger F) = 1$ and $\mathrm{Tr}(\tilde{\sigma}) = 0$ because $\mathrm{Tr}(E^\dagger F) = 0$. Then

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\theta}) = \max_G \frac{|\mathrm{Tr}(G\dot{\sigma})|^2}{4 - 2\mathrm{Tr}(G\tilde{\sigma}) + \mathrm{Tr}(G^2\sigma) - |\mathrm{Tr}(G\sigma)|^2} + O(\varepsilon) \tag{E.5}$$

$$= \max_{G,x} \frac{|\mathrm{Tr}(G\dot{\sigma})|^2}{4x^2 + 2x\mathrm{Tr}(G\tilde{\sigma}) + \mathrm{Tr}(G^2\sigma) - |\mathrm{Tr}(G\sigma)|^2} + O(\varepsilon) \tag{E.6}$$

$$= \max_G \frac{|\mathrm{Tr}(G\dot{\sigma})|^2}{-\frac{|\mathrm{Tr}(G\tilde{\sigma})|^2}{4} + \left(\mathrm{Tr}(G^2\sigma) - |\mathrm{Tr}(G\sigma)|^2\right)} + O(\varepsilon), \tag{E.7}$$

where in the second step we used the fact that any rescaling of $G$ ($G \leftarrow -G/x$) should not change the optimal QFI. Note that to obtain the solution of the original $G$ in $T$, one need

to rescale the final solution back using $G \leftarrow 4G/\mathrm{Tr}(G\tilde{\sigma})$.

To find the optimal $G$, we first observe that $\mathrm{Tr}(\dot{\sigma}) = \mathrm{Tr}(\tilde{\sigma}) = 0$. Therefore, WLOG, we assume $\mathrm{Tr}(G\sigma) = 0$ because $G \leftarrow G - \mathrm{Tr}(G)\frac{I}{r}$ does not change the target function. Let the derivative of Eq. (E.7) be zero, we have

$$2\dot{\sigma}\Big(\mathrm{Tr}(G^2\sigma) - \frac{|\mathrm{Tr}(G\tilde{\sigma})|^2}{4}\Big) - \mathrm{Tr}(G\dot{\sigma})\Big((\sigma G + G\sigma) - \frac{2\mathrm{Tr}(G\tilde{\sigma})\tilde{\sigma}}{4}\Big) = 0, \tag{E.8}$$

$$\Leftrightarrow \quad \frac{\dot{\sigma}}{\mathrm{Tr}(G\dot{\sigma})}\Big(\mathrm{Tr}(G^2\sigma) - \frac{|\mathrm{Tr}(G\tilde{\sigma})|^2}{4}\Big) + \frac{\mathrm{Tr}(G\tilde{\sigma})\tilde{\sigma}}{4} = \frac{1}{2}(\sigma G + G\sigma), \tag{E.9}$$

$$\Leftrightarrow \quad G = L_\sigma[x\dot{\sigma} + y\tilde{\sigma}], \quad 4y = \mathrm{Tr}(G\tilde{\sigma}) = \mathrm{Tr}(L_\sigma[x\dot{\sigma} + y\tilde{\sigma}]\tilde{\sigma}), \tag{E.10}$$

$$\Leftarrow \quad x = 4 - \mathrm{Tr}(L_\sigma[\tilde{\sigma}]\tilde{\sigma}), \quad y = \mathrm{Tr}(L_\sigma[\dot{\sigma}]\tilde{\sigma}). \tag{E.11}$$

Note that in Eq. (E.10) we used $x\dot{\sigma} + y\tilde{\sigma} = \frac{1}{2}(G\sigma + \sigma G)$ and $\mathrm{Tr}(G^2\sigma) = \mathrm{Tr}(G(x\dot{\sigma} + y\tilde{\sigma}))$. Plug the optimal $G = L_\sigma[x\dot{\sigma} + y\tilde{\sigma}]$ into Eq. (E.7) where $x, y$ satisfies Eq. (E.11), we get

$$\mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\theta}) = \mathrm{Tr}(L_\sigma[\dot{\sigma}]\dot{\sigma}) + \frac{\mathrm{Tr}(L_\sigma[\dot{\sigma}]\tilde{\sigma})^2}{4 - \mathrm{Tr}(L_\sigma[\tilde{\sigma}]\tilde{\sigma})} + O(\varepsilon). \tag{E.12}$$

# Appendix F

# $\mathfrak{F}_{\mathbf{sql}}$ as a function of $(C, \tilde{C})$

To derive Eq. (4.56), we use an orthonormal basis $\{|i\rangle\rangle\}_{i=1}^{d^2}$, where $|i\rangle\rangle = \frac{1}{\sqrt{\lambda_i}}|K_i C\rangle\rangle$ for $1 \le i \le r$. We have

$$\sigma = \begin{pmatrix} (\lambda_i \delta_{ij}) & 0 \\ 0 & 0 \end{pmatrix}, \quad \dot{\sigma} = \begin{pmatrix} \left(\langle\langle K_i C|\dot{K}_j C\rangle\rangle\sqrt{\frac{\lambda_j}{\lambda_i}} + \sqrt{\frac{\lambda_i}{\lambda_j}}\langle\langle \dot{K}_i C|K_j C\rangle\rangle\right) & (\langle\langle \dot{K}_i C|j'\rangle\rangle\sqrt{\lambda_i}) \\ (\langle\langle i'|\dot{K}_j C\rangle\rangle\sqrt{\lambda_j}) & 0 \end{pmatrix}, \tag{F.1}$$

$$\tilde{\sigma} = \begin{pmatrix} \left(i\langle\langle K_i C|K_j D\rangle\rangle\sqrt{\frac{\lambda_j}{\lambda_i}} - i\sqrt{\frac{\lambda_i}{\lambda_j}}\langle\langle K_i D|K_j C\rangle\rangle\right) & (-i\langle\langle K_i D|j'\rangle\rangle\sqrt{\lambda_i}) \\ (i\langle\langle i'|K_j D\rangle\rangle\sqrt{\lambda_j}) & 0 \end{pmatrix}, \tag{F.2}$$

where $1 \le i, j \le r$ and $r+1 \le i', j \le d^2$. Then

$$\begin{aligned} \mathrm{Tr}(L_\sigma[\dot{\sigma}]\dot{\sigma}) &= 2\sum_{i,j:\lambda_i+\lambda_j>0} \frac{|(\dot{\sigma})_{ij}|^2}{\lambda_i+\lambda_j} \\ &= 2\sum_{i,j=1}^{r} \frac{\left|\langle\langle K_i C|\dot{K}_j C\rangle\rangle\sqrt{\frac{\lambda_j}{\lambda_i}} + \sqrt{\frac{\lambda_i}{\lambda_j}}\langle\langle \dot{K}_i C|K_j C\rangle\rangle\right|^2}{\lambda_i+\lambda_j} + 4\sum_{i'=r+1}^{d^2}\sum_{j=1}^{r} \frac{|\langle\langle i'|\dot{K}_j C\rangle\rangle\sqrt{\lambda_j}|^2}{\lambda_j} \\ &= 4\mathrm{Tr}(C^\dagger \dot{\mathbf{K}}^\dagger \dot{\mathbf{K}} C) + 2\sum_{i,j=1}^{r} \frac{\left|\langle\langle K_i C|\dot{K}_j C\rangle\rangle\sqrt{\frac{\lambda_j}{\lambda_i}} + \sqrt{\frac{\lambda_i}{\lambda_j}}\langle\langle \dot{K}_i C|K_j C\rangle\rangle\right|^2}{\lambda_i+\lambda_j} - 2\frac{|\langle\langle K_i C|\dot{K}_j C\rangle\rangle|^2}{\lambda_i} \\ &= 4\mathrm{Tr}(C^\dagger \dot{\mathbf{K}}^\dagger \dot{\mathbf{K}} C) - 2\sum_{i,j=1}^{r} \frac{|\tau'_{ij}|^2}{\lambda_i+\lambda_j} = 4\mathrm{Tr}(C^\dagger \dot{\mathbf{K}}^\dagger \dot{\mathbf{K}} C) - \mathrm{Tr}(L_\tau[\tau']\tau'), \end{aligned} \tag{F.3}$$

$$
\begin{aligned}
\mathrm{Tr}(L_\sigma[\tilde{\sigma}]\tilde{\sigma}) &= 2\sum_{i,j:\lambda_i+\lambda_j>0}\frac{|(\tilde{\sigma})_{ij}|^2}{\lambda_i+\lambda_j} \\
&= 2\sum_{i,j=1}^{r}\frac{\left|i\langle\!\langle K_iC|K_jD\rangle\!\rangle\sqrt{\frac{\lambda_j}{\lambda_i}}-i\sqrt{\frac{\lambda_i}{\lambda_j}}\langle\!\langle K_iD|K_jC\rangle\!\rangle\right|^2}{\lambda_i+\lambda_j}+4\sum_{i'=r+1}^{d^2}\sum_{j=1}^{r}\frac{\left|i\langle\!\langle i'|K_jD\rangle\!\rangle\sqrt{\lambda_j}\right|^2}{\lambda_j} \\
&= 4+2\sum_{i,j=1}^{r}\frac{\left|i\langle\!\langle K_iC|K_jD\rangle\!\rangle\sqrt{\frac{\lambda_j}{\lambda_i}}-i\sqrt{\frac{\lambda_i}{\lambda_j}}\langle\!\langle K_iD|K_jC\rangle\!\rangle\right|^2}{\lambda_i+\lambda_j}-2\frac{|\langle\!\langle K_iC|K_jD\rangle\!\rangle|^2}{\lambda_i} \\
&= 4-2\sum_{ij}\frac{|\tilde{\tau}_{ij}|^2}{\lambda_i+\lambda_j}=4-\mathrm{Tr}(L_\tau[\tilde{\tau}]\tilde{\tau}),
\end{aligned}
\tag{F.4}
$$

and

$$
\begin{aligned}
\mathrm{Tr}(L_\sigma[\dot{\sigma}]\tilde{\sigma}) &= 2\sum_{i,j:\lambda_i+\lambda_j\neq 0}\frac{\dot{\sigma}_{ij}\tilde{\sigma}_{ji}}{\lambda_i+\lambda_j} \\
&= 2\sum_{i,j=1}^{r}\frac{\dot{\sigma}_{ij}\tilde{\sigma}_{ji}}{\lambda_i+\lambda_j}+2\sum_{i'=r+1}^{d^2}\sum_{j=1}^{r}\frac{\dot{\sigma}_{i'j}\tilde{\sigma}_{ji'}}{\lambda_j}+2\sum_{i'=r+1}^{d^2}\sum_{j=1}^{r}\frac{\dot{\sigma}_{ji'}\tilde{\sigma}_{i'j}}{\lambda_j} \\
&= -2\mathrm{Tr}(\tilde{C}H)+2\sum_{i,j=1}^{r}\frac{\dot{\sigma}_{ij}\tilde{\sigma}_{ji}}{\lambda_i+\lambda_j}+2i\sum_{i,j=1}^{r}\frac{\langle\!\langle K_jD|K_iC\rangle\!\rangle\langle\!\langle K_iC|\dot{K}_jC\rangle\!\rangle}{\lambda_i}-\frac{\langle\!\langle \dot{K}_jC|K_iC\rangle\!\rangle\langle\!\langle K_iC|K_jD\rangle\!\rangle}{\lambda_i} \\
&= -2\mathrm{Tr}(\tilde{C}H)+2\sum_{i,j=1}^{r}\frac{\tau'_{ij}\tilde{\tau}_{ji}}{\lambda_i+\lambda_j}=-2\mathrm{Tr}(\tilde{C}H)+\mathrm{Tr}(L_\tau[\tau']\tilde{\tau}).
\end{aligned}
\tag{F.5}
$$

# Appendix G

# QFIs for qubit depolarizing channels

Now we calculate $\mathfrak{F}_1$, $\mathfrak{F}_{\rm SQL}$ and $\mathfrak{F}_{\rm HL}$ for general depolarizing channels $\mathcal{N}_\theta(\rho)=\sum_{i=1}^4 K_i\rho K_i^\dagger$, where

$$\mathbf{K}=\begin{pmatrix}\sqrt{1-p}\\ \sqrt{p_x}\sigma_x\\ \sqrt{p_y}\sigma_y\\ \sqrt{p_z}\sigma_z\end{pmatrix}e^{-\frac{i\theta}{2}\sigma_z},\quad \dot{\mathbf{K}}=\begin{pmatrix}-\frac{i}{2}\sqrt{1-p}\sigma_z\\ -\frac{1}{2}\sqrt{p_x}\sigma_y\\ \frac{1}{2}\sqrt{p_y}\sigma_x\\ -\frac{i}{2}\sqrt{p_z}\end{pmatrix}e^{-\frac{i\theta}{2}\sigma_z}, \tag{G.1}$$

$$\beta=i\mathbf{K}^\dagger(\dot{\mathbf{K}}-ih\mathbf{K})=\frac{1}{2}\sigma_z+\mathbf{K}^\dagger h\mathbf{K}. \tag{G.2}$$

$$\beta=0\quad\Rightarrow\quad\begin{aligned}&(1-p)h_{11}+p_xh_{22}+p_yh_{33}+p_zh_{44}=0,\\ &\sqrt{(1-p)p_x}(h_{12}+h_{21})+i\sqrt{p_yp_z}h_{34}-i\sqrt{p_yp_z}h_{43}=0,\\ &\sqrt{(1-p)p_y}(h_{13}+h_{31})-i\sqrt{p_xp_z}h_{24}+i\sqrt{p_xp_z}h_{42}=0,\\ &\tfrac{1}{2}+\sqrt{(1-p)p_z}(h_{14}+h_{41})+i\sqrt{p_xp_y}h_{23}-i\sqrt{p_xp_y}h_{32}=0.\end{aligned} \tag{G.3}$$

Clearly, HNKS is satisfied if and only if $p_x=p_z=0$ or $p_y=p_z=0$. It is easy to see that when $h_{ij}=0$ for all $i,j$ except $h_{23}$, $h_{32}$, $h_{14}$ and $h_{41}$, $\alpha=\|\alpha\|\,I$, $\|\alpha\|$ takes its minimum

and

$$\|\alpha\| = \frac{1}{4} + \sqrt{(1-p)p_z}(h_{14}+h_{41}) + i\sqrt{p_x p_y}(h_{23}-h_{32}) \\ + (1-p+p_z)\left|h_{14}\right|^2 + (p_x+p_y)\left|h_{23}\right|^2 \quad \text{(G.4)}$$

Then

$$\mathfrak{F}_1(\mathcal{N}_\omega) = 4\min_h \|\alpha\| = 1 - 4\left(\frac{p_x p_y}{p_x+p_y} + \frac{(1-p)p_z}{1-p+p_z}\right). \quad \text{(G.5)}$$

When HNKS is satisfied,

$$\mathfrak{F}_{\text{HL}}(\mathcal{N}_\omega) = 4\min_h \|\beta\|^2 = 1, \quad \text{(G.6)}$$

and when HNKS is violated,

$$\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega) = 4\min_{h:\beta=0} \|\alpha\| = -1 + \frac{1}{4}\left(\frac{p_x p_y}{p_x+p_y} + \frac{(1-p)p_z}{1-p+p_z}\right)^{-1}. \quad \text{(G.7)}$$

# Appendix H

# Solving the optimal QEC code for amplitude damping channels

In this appendix, we use the algorithm in Section 4.4.3 to solve for the optimal QEC protocol analytically for amplitude damping channels with two Kraus operators:

$$\mathbf{K} = \begin{pmatrix} |0\rangle\langle 0| + \sqrt{1-p}\,|1\rangle\langle 1| \\ \\ \sqrt{p}\,|0\rangle\langle 1| \end{pmatrix} e^{-i\frac{\omega}{2}\sigma_z} = \begin{pmatrix} |0\rangle\langle 0|\, e^{-i\frac{\omega}{2}} + \sqrt{1-p}\,|1\rangle\langle 1|\, e^{i\frac{\omega}{2}} \\ \\ \sqrt{p}\,|0\rangle\langle 1|\, e^{i\frac{\omega}{2}} \end{pmatrix}. \tag{H.1}$$

Clearly, $H = i\mathbf{K}^\dagger\dot{\mathbf{K}} = \sigma_z/2$.

## H.1 Finding the optimal $C$

First, we want to find a full rank normalized $C^\diamond$, such that $\min_{h:\beta=0} 4\mathrm{Tr}(C^{\diamond\dagger}\alpha C^\diamond)$ is close to $\mathfrak{F}^{(u)}_{\mathrm{SQL}}(\mathcal{E}_\omega)$ and we will follow the algorithm described in Section 4.4.3.

We first compute $\alpha$ and $\beta$. Note that

$$\dot{\mathbf{K}} = \begin{pmatrix} -\frac{i}{2}\,|0\rangle\langle 0| + \frac{i}{2}\sqrt{1-p}\,|1\rangle\langle 1| \\ \\ \frac{i}{2}\sqrt{p}\,|0\rangle\langle 1| \end{pmatrix} e^{-i\frac{\omega}{2}\sigma_z}. \tag{H.2}$$

We first observe that in order to make $\beta = i\mathbf{K}^\dagger(\dot{\mathbf{K}} - ih\mathbf{K}) = 0$, $h$ has to be diagonal and

then

$$\dot{\mathbf{K}} - ih\mathbf{K} = \begin{pmatrix} (-\frac{i}{2} - ih_{11}) |0\rangle\langle 0| + (\frac{i}{2} - ih_{11})\sqrt{1-p}\, |1\rangle\langle 1| \\ (\frac{i}{2} - ih_{22})\sqrt{p}\, |0\rangle\langle 1| \end{pmatrix} e^{-i\frac{\omega}{2}\sigma_z}. \tag{H.3}$$

We will also assume $\omega = 0$ for simplicity.

$$\beta = \left(\frac{1}{2} + h_{11}\right)|0\rangle\langle 0| + \left(-\frac{1}{2} + h_{11}\right)(1-p)\,|1\rangle\langle 1| + \left(-\frac{1}{2} + h_{22}\right)p\,|1\rangle\langle 1| = 0, \tag{H.4}$$

$$\Rightarrow \quad h_{11} = -\frac{1}{2}, \quad h_{22} = \frac{2-p}{2p}. \tag{H.5}$$

Therefore $\alpha = (\dot{\mathbf{K}} - ih\mathbf{K})^\dagger(\dot{\mathbf{K}} - ih\mathbf{K}) = \left((1-p) + \frac{(1-p)^2}{p}\right)|1\rangle\langle 1| = \frac{1-p}{p}|1\rangle\langle 1|$.

Since there is only one solution of $h$ such that $\beta = 0$, there is no need to solve $\min_{h:\beta=0} \|\alpha\|$ using a SDP and the only solution is: $\alpha^\diamond = \frac{1-p}{p}|1\rangle\langle 1|$ and $\mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}}) = 4(1-p)/p$. We could take $C^\diamond = \sin\delta\,|0\rangle\langle 0| + \cos\delta\,|1\rangle\langle 1|$ where $\delta$ is small. Note that here we use the small constant $\delta$ instead of $\eta'$ in Eq. (4.71) for simplicity. They are related by $\eta'/2 = \sin^2(\delta)$.

## H.2 Finding the optimal $\tilde{C}$

Next we find the optimal $\tilde{C}^\diamond$ which minimizes

$$\frac{\left(-\operatorname{Tr}(\tilde{C}H) + \sum_{i,j=1}^r \frac{\tilde{\tau}_{ij}\tau'_{ji}}{\lambda_i+\lambda_j}\right)^2}{\frac{1}{2}\sum_{i,j=1}^r \frac{|\tilde{\tau}_{ij}|^2}{\lambda_i+\lambda_j}} = \frac{|\langle\!\langle \tilde{C}|f_2(C)\rangle\!\rangle|^2}{\langle\!\langle \tilde{C}|f_3(C)|\tilde{C}\rangle\!\rangle}, \tag{H.6}$$

and the solution is $|\tilde{C}^\diamond\rangle\!\rangle = f_3(C^\diamond)^{-1}|f_2(C^\diamond)\rangle\!\rangle$.

We first compute

$$\tau = \begin{pmatrix} \sin^2\delta + (1-p)\cos^2\delta & 0 \\ 0 & p\cos^2\delta \end{pmatrix} \approx \begin{pmatrix} 1-p & 0 \\ 0 & p \end{pmatrix}, \tag{H.7}$$

$$\tau' = \begin{pmatrix} \sin^2\delta - (1-p)\cos^2\delta & 0 \\ 0 & -p\cos^2\delta \end{pmatrix} \approx \begin{pmatrix} -(1-p) & 0 \\ 0 & -p \end{pmatrix}, \tag{H.8}$$

$$\tilde{\tau} = \begin{pmatrix} \mathrm{Tr}(\tilde{C}\begin{pmatrix} 1 & 0 \\ 0 & 1-p \end{pmatrix}) & \mathrm{Tr}(\tilde{C}\begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}) \\ \mathrm{Tr}(\tilde{C}\begin{pmatrix} 0 & 0 \\ \sqrt{p} & 0 \end{pmatrix}) & \mathrm{Tr}(\tilde{C}\begin{pmatrix} 0 & 0 \\ 0 & p \end{pmatrix}) \end{pmatrix}, \tag{H.9}$$

where by "≈" we ignore the small contribution of $O(\delta)$. Then we have

$$f_2(C^\diamond) \approx -|00\rangle, \tag{H.10}$$

$$f_3(C^\diamond)^{-1} \approx p|01\rangle\langle 01| + p|10\rangle\langle 10| + (|00\rangle + (1-p)|11\rangle)(\langle 00| + (1-p)\langle 11|)\frac{1}{2(1-p)} + (p|11\rangle)(p\langle 11|)\frac{1}{2p}, \tag{H.11}$$

$$f_3(C^\diamond)^{-1} \approx \frac{1}{p}|01\rangle\langle 01| + \frac{1}{p}|10\rangle\langle 10| + \frac{2}{p}((1-p)|00\rangle\langle 00| + |11\rangle\langle 11| - (1-p)|00\rangle\langle 11| - (1-p)|11\rangle\langle 00|). \tag{H.12}$$

Then $|\tilde{C}^\diamond\rangle\!\rangle \approx |00\rangle - |11\rangle$ and we could take $D^\diamond = \cos\delta|0\rangle\langle 0| - \sin\delta|1\rangle\langle 1|$.

## H.3 Attaining the asymptotic QFI

Now we have the optimal code from the previous two steps:

$$|\mathfrak{c}_0\rangle = \sin(\delta+\varepsilon)|0\rangle_S|00\rangle_A + \cos(\delta+\varepsilon)|1\rangle_S|10\rangle_A, \tag{H.13}$$

$$|\mathfrak{c}_1\rangle = \sin(\delta-\varepsilon)|0\rangle_S|01\rangle_A + \cos(\delta-\varepsilon)|1\rangle_S|11\rangle_A. \tag{H.14}$$

where $\delta$ and $\varepsilon = o(\delta)$ are small values. The last step is to find the exact relation between $\delta$ and $\varepsilon$ and $\mathfrak{F}_{\mathrm{SQL}}(\mathcal{N}_\omega^{\mathrm{ad}}) - \mathfrak{F}_{\mathrm{SQL}}(\mathcal{D}_{L,\omega})$.

To do so, we need the near-optimal recovery channel computed using Eq. (4.52):

$$G_{\text{opt}} = \frac{2i}{\sqrt{1-p}} |00\rangle\langle 11| + \frac{-2i}{\sqrt{1-p}} |11\rangle\langle 00| := \begin{pmatrix} 0 & 0 & 0 & \frac{2i}{\sqrt{1-p}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{-2i}{\sqrt{1-p}} & 0 & 0 & 0 \end{pmatrix}, \tag{H.15}$$

and

$$T_{\text{opt}} = e^{i\varepsilon G_{\text{opt}}} = \begin{pmatrix} \cos\left(\frac{2\varepsilon}{\sqrt{1-p}}\right) & 0 & 0 & -\sin\left(\frac{2\varepsilon}{\sqrt{1-p}}\right) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sin\left(\frac{2\varepsilon}{\sqrt{1-p}}\right) & 0 & 0 & \cos\left(\frac{2\varepsilon}{\sqrt{1-p}}\right) \end{pmatrix}. \tag{H.16}$$

Then using Eq. (4.50), and

$$E_{0,1} = \begin{pmatrix} |K_1 A_{0,1}\rangle\rangle & |K_2 A_{0,1}\rangle\rangle \end{pmatrix} = \begin{pmatrix} \sin(\delta \pm \varepsilon) & 0 \\ 0 & \sqrt{p}\cos(\delta \pm \varepsilon) \\ 0 & 0 \\ \sqrt{1-p}\cos(\delta \pm \varepsilon) & 0 \end{pmatrix}, \tag{H.17}$$

we finally get

$$\begin{aligned} \xi = p(\cos(2\delta) + \cos(2\varepsilon))\sin^2\left(\frac{\varepsilon}{\sqrt{1-p}}\right) + \sqrt{1-p}\sin(2\varepsilon)\sin\left(\frac{2\varepsilon}{\sqrt{1-p}}\right) \\ + \cos(2\varepsilon)\cos\left(\frac{2\varepsilon}{\sqrt{1-p}}\right) = 1 - \frac{2p\sin^2\delta}{1-p}\varepsilon^2 + O(\varepsilon^4), \end{aligned} \tag{H.18}$$

$$\dot{\xi} = -i\sqrt{1-p}\sin(2\delta)\sin\left(\frac{2\varepsilon}{\sqrt{1-p}}\right) = -2i\sin(2\delta)\varepsilon + O(\varepsilon^3). \tag{H.19}$$

Clearly,

$$\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\omega}) = \frac{4(1-p)\cos^2\delta}{p} + O(\varepsilon^2), \tag{H.20}$$

as expected. When $\delta$ is small and $\varepsilon = o(\delta)$, we would have $\mathfrak{F}_{\text{SQL}}(\mathcal{D}_{L,\omega}) \approx \mathfrak{F}_{\text{SQL}}(\mathcal{N}_\omega^{\text{ad}})$.

# Bibliography


Aaronson, S. (2019). Shadow tomography of quantum states. *SIAM Journal on Computing*, 49(5):STOC18–368.

Acosta, V. M., Bauch, E., Ledbetter, M. P., Santori, C., Fu, K.-M., Barclay, P. E., Beausoleil, R. G., Linget, H., Roch, J. F., Treussart, F., et al. (2009). Diamonds with a high density of nitrogen-vacancy centers for magnetometry applications. *Physical Review B*, 80(11):115202.

Aharonov, D. and Ben-Or, M. (1999). Fault-tolerant quantum computation with constant error rate. *arXiv preprint quant-ph/9906129*.

Albarelli, F., Friel, J. F., and Datta, A. (2019). Evaluating the holevo cramér-rao bound for multiparameter quantum metrology. *Physical Review Letters*, 123(20):200503.

Albarelli, F., Rossi, M. A. C., Paris, M. G. A., and Genoni, M. G. (2017). Ultimate limits for quantum magnetometry via time-continuous measurements. *New Journal of Physics*, 19(12):123011.

Albarelli, F., Rossi, M. A. C., Tamascelli, D., and Genoni, M. G. (2018). Restoring Heisenberg scaling in noisy quantum metrology by monitoring the environment. *Quantum*, 2:110.

Albert, V. V., Noh, K., Duivenvoorden, K., Young, D. J., Brierley, R. T., Reinhold, P., Vuillot, C., Li, L., Shen, C., Girvin, S. M., Terhal, B. M., and Jiang, L. (2018). Performance and structure of single-mode bosonic codes. *Physical Review A*, 97:032346.

André, A., Sørensen, A., and Lukin, M. (2004). Stability of atomic clocks based on entangled atoms. *Physical Review letters*, 92(23):230801.

Appel, J., Windpassinger, P. J., Oblak, D., Hoff, U. B., Kjærgaard, N., and Polzik, E. S. (2009). Mesoscopic atomic entanglement for precision measurements beyond the standard quantum limit. *Proceedings of the National Academy of Sciences*, 106(27):10960–10965.

Arrad, G., Vinkler, Y., Aharonov, D., and Retzker, A. (2014). Increasing sensing resolution with error correction. *Physical Review Letters*, 112:150801.

Audenaert, K. and De Moor, B. (2002). Optimizing completely positive maps using semidefinite programming. *Physical Review A*, 65(3):030302.

Banaszek, K., Demkowicz-Dobrzański, R., and Walmsley, I. A. (2009). Quantum states made to measure. *Nature Photonics*, 3(12):673–676.

Barndorff-Nielsen, O. E. and Gill, R. D. (2000). Fisher information in quantum statistics. *Journal of Physics A: Mathematical and General*, 33(24):4481–4490.

Barnum, H. and Knill, E. (2002). Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43(5):2097–2106.

Beckey, J. L., Cerezo, M., Sone, A., and Coles, P. J. (2020). Variational quantum algorithm for estimating the quantum fisher information. *arXiv preprint arXiv:2010.10488v1*.

Bennett, C. H., DiVincenzo, D. P., Smolin, J. A., and Wootters, W. K. (1996). Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824.

Bennett, C. H. and Shor, P. W. (2004). Quantum channel capacities. *Science*, 303(5665):1784–1787.

Bény, C. and Oreshkov, O. (2010). General conditions for approximate quantum error correction and near-optimal recovery channels. *Physical Review Letters*, 104:120501.

Bény, C., Zimborás, Z., and Pastawski, F. (2018). Approximate recovery with locality and symmetry constraints. *arXiv preprint arXiv:1806.10324v2*.

Blatt, R. and Wineland, D. (2008). Entangled states of trapped atomic ions. *Nature*, 453(7198):1008–1015.

Boixo, S., Flammia, S. T., Caves, C. M., and Geremia, J. M. (2007). Generalized limits for single-parameter quantum estimation. *Physical Review Letters*, 98(9):090401.

Bollinger, J. J., Itano, W. M., Wineland, D. J., and Heinzen, D. J. (1996). Optimal frequency measurements with maximally correlated states. *Physical Review A*, 54:R4649–R4652.

Bondurant, R. S. and Shapiro, J. H. (1984). Squeezed states in phase-sensing interferometers. *Physical Review D*, 30(12):2548.

Borregaard, J. and Sørensen, A. S. (2013). Near-heisenberg-limited atomic clocks in the presence of decoherence. *Physical Review Letters*, 111(9):090801.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Brandão, F. G. S. L., Crosson, E., Şahinoğlu, M. B., and Bowen, J. (2019). Quantum error correcting codes in eigenstates of translation-invariant spin chains. *Physical Review Letters*, 123:110502.

Braunstein, S. L. and Caves, C. M. (1994). Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72:3439–3443.

Bravyi, S. and König, R. (2013). Classification of topologically protected gates for local stabilizer codes. *Physical Review Letters*, 110:170503.

Breuer, H.-P., Petruccione, F., et al. (2002). *The theory of open quantum systems*. Oxford University Press on Demand.

Buzek, V., Derka, R., and Massar, S. (1999). Optimal quantum clocks. *Physical Review Letters*, 82:2207–2210.

Calderbank, A. R., Rains, E. M., Shor, P. W., and Sloane, N. J. (1997). Quantum error correction and orthogonal geometry. *Physical Review Letters*, 78(3):405.

Calderbank, A. R. and Shor, P. W. (1996). Good quantum error-correcting codes exist. *Physical Review A*, 54(2):1098–1105.

Caves, C. M. (1981). Quantum-mechanical noise in an interferometer. *Physical Review D*, 23:1693–1708.

Chabuda, K., Dziarmaga, J., Osborne, T. J., and Demkowicz-Dobrzański, R. (2020). Tensor-network approach for quantum metrology in many-body quantum systems. *Nature Communications*, 11(1):1–12.

Chaves, R., Brask, J. B., Markiewicz, M., Kołodyński, J., and Acín, A. (2013). Noisy metrology beyond the standard quantum limit. *Physical Review Letters*, 111:120401.

Chen, Y., Chen, H., Liu, J., Miao, Z., and Yuan, H. (2020). Fluctuation-enhanced quantum metrology. *arXiv preprint arXiv:2003.13010v1*.

Chen, Y. and Yuan, H. (2019). Zero-error quantum hypothesis testing in finite time with quantum error correction. *Physical Review A*, 100(2):022336.

Chin, A. W., Huelga, S. F., and Plenio, M. B. (2012). Quantum metrology in non-markovian environments. *Physical Review Letters*, 109:233601.

Chiribella, G., D'ariano, G., and Sacchi, M. (2005). Optimal estimation of group transformations using entanglement. *Physical Review A*, 72(4):042338.

Chiribella, G., D'Ariano, G. M., and Perinotti, P. (2008). Memory effects in quantum channel discrimination. *Physical Review Letters*, 101:180501.

Chiribella, G., D'Ariano, G. M., and Perinotti, P. (2009). Realization schemes for quantum instruments in finite dimensions. *Journal of mathematical physics*, 50(4):042101.

Chuang, I. L., Leung, D. W., and Yamamoto, Y. (1997). Bosonic quantum codes for amplitude damping. *Physical Review A*, 56(2):1114.

Czajkowski, J., Pawłowski, K., and Demkowicz-Dobrzański, R. (2019). Many-body effects in quantum metrology. *New Journal of Physics*, 21(5):053031.

Dankert, C. (2005). Efficient simulation of random quantum states and operators. *arXiv preprint quant-ph/0512217*.

Degen, C. L., Reinhard, F., and Cappellaro, P. (2017). Quantum sensing. *Reviews of Modern Physics*, 89:035002.

Del Moral, P. and Niclas, A. (2018). A taylor expansion of the square root matrix function. *Journal of Mathematical Analysis and Applications*, 465(1):259–266.

Demkowicz-Dobrzański, R., Banaszek, K., and Schnabel, R. (2013). Fundamental quantum interferometry bound for the squeezed-light-enhanced gravitational wave detector geo 600. *Physical Review A*, 88:041802.

Demkowicz-Dobrzański, R., Czajkowski, J., and Sekatski, P. (2017). Adaptive quantum metrology under general markovian noise. *Physical Review X*, 7:041009.

Demkowicz-Dobrzański, R. and Maccone, L. (2014). Using entanglement against noise in quantum metrology. *Physical Review Letters*, 113:250801.

Demkowicz-Dobrzański, R., Dorner, U., Smith, B., Lundeen, J., Wasilewski, W., Banaszek, K., and Walmsley, I. (2009). Quantum phase estimation with lossy interferometers. *Physical Review A*, 80(1):013825.

Demkowicz-Dobrzański, R., Górecki, W., and Guţă, M. (2020). Multi-parameter estimation beyond quantum fisher information. *Journal of Physics A: Mathematical and Theoretical*, 53(36):363001.

Demkowicz-Dobrzański, R., Kołodyński, J., and Guţă, M. (2012). The elusive heisenberg limit in quantum-enhanced metrology. *Nature Communications*, 3:1063.

do Rosário Grossinho, M. and Tersian, S. A. (2001). *An introduction to minimax theorems and their applications to differential equations*, volume 52. Springer Science & Business Media.

Dorner, U. (2012). Quantum frequency estimation with trapped ions and atoms. *New Journal of Physics*, 14(4):043011.

Dorner, U., Demkowicz-Dobrzański, R., Smith, B., Lundeen, J., Wasilewski, W., Banaszek, K., and Walmsley, I. (2009). Optimal quantum phase estimation. *Physical Review letters*, 102(4):040403.

Dowling, J. P. (1998). Correlated input-port, matter-wave interferometer: Quantum-noise limits to the atom-laser gyroscope. *Physical Review A*, 57(6):4736.

Dür, W., Skotiniotis, M., Fröwis, F., and Kraus, B. (2014). Improved quantum metrology using quantum error correction. *Physical Review Letters*, 112:080801.

Dutt, M. G., Childress, L., Jiang, L., Togan, E., Maze, J., Jelezko, F., Zibrov, A., Hemmer, P., and Lukin, M. (2007). Quantum register based on individual electronic and nuclear spin qubits in diamond. *Science*, 316(5829):1312–1316.

Eastin, B. and Knill, E. (2009). Restrictions on transversal encoded quantum gate sets. *Physical Review Letters*, 102:110502.

Escher, B., Davidovich, L., Zagury, N., and de Matos Filho, R. (2012). Quantum metrological limits via a variational approach. *Physical Review letters*, 109(19):190404.

Escher, B., de Matos Filho, R., and Davidovich, L. (2011). General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology. *Nature Physics*, 7(5):406.

Faist, P., Nezami, S., Albert, V. V., Salton, G., Pastawski, F., Hayden, P., and Preskill, J. (2020). Continuous symmetries and approximate quantum error correction. *Physical Review X*, 10(4):041018.

Fang, K., Fawzi, O., Renner, R., and Sutter, D. (2020). Chain rule for the quantum relative entropy. *Physical Review Letters*, 124:100501.

Fletcher, A. S., Shor, P. W., and Win, M. Z. (2007). Optimum quantum error recovery using semidefinite programming. *Physical Review A*, 75:012338.

Fu, R. R., Weiss, B. P., Lima, E. A., Harrison, R. J., Bai, X.-N., Desch, S. J., Ebel, D. S., Suavet, C., Wang, H., Glenn, D., et al. (2014). Solar nebula magnetic fields recorded in the semarkona meteorite. *Science*, 346(6213):1089–1092.

Fujiwara, A. (2006). Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. *Journal of Physics A: Mathematical and General*, 39(40):12489–12504.

Fujiwara, A. and Imai, H. (2008). A fibre bundle over manifolds of quantum channels and its application to quantum statistics. *Journal of Physics A: Mathematical and Theoretical*, 41(25):255304.

Gefen, T., Herrera-Martí, D. A., and Retzker, A. (2016). Parameter estimation with efficient photodetectors. *Physical Review A*, 93(3):032133.

Genoni, M. G., Olivares, S., and Paris, M. G. A. (2011). Optical phase estimation in the presence of phase diffusion. *Physical Review Letters*, 106(15).

Gessner, M., Pezzè, L., and Smerzi, A. (2018). Sensitivity bounds for multiparameter quantum metrology. *Physical Review Letters*, 121(13):130503.

Gilchrist, A., Langford, N. K., and Nielsen, M. A. (2005). Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71:062310.

Gill, R. D. and Massar, S. (2000). State estimation for large ensembles. *Physical Review A*, 61:042312.

Giovannetti, V., Lloyd, S., and Maccone, L. (2001). Quantum-enhanced positioning and clock synchronization. *Nature*, 412(6845):417–419.

Giovannetti, V., Lloyd, S., and Maccone, L. (2004). Quantum-enhanced measurements: Beating the standard quantum limit. *Science*, 306(5700):1330–1336.

Giovannetti, V., Lloyd, S., and Maccone, L. (2006). Quantum metrology. *Physical Review Letters*, 96:010401.

Giovannetti, V., Lloyd, S., and Maccone, L. (2011). Advances in quantum metrology. *Nature Photonics*, 5(4):222.

Górecki, W., Zhou, S., Jiang, L., and Demkowicz-Dobrzański, R. (2020). Optimal probes and error-correction schemes in multi-parameter quantum metrology. *Quantum*, 4:288.

Gorini, V., Kossakowski, A., and Sudarshan, E. C. G. (1976). Completely positive dynamical semigroups of n-level systems. *Journal of Mathematical Physics*, 17(5):821–825.

Gottesman, D. (1996). Class of quantum error-correcting codes saturating the quantum hamming bound. *Physical Review A*, 54(3):1862.

Gottesman, D. (2009). An introduction to quantum error correction and fault-tolerant quantum computation. *arXiv preprint arXiv:0904.2557*.

Grant, M. and Boyd, S. Cvx: Matlab software for disciplined convex programming http://cvxr.com/cvx/.

Haah, J., Harrow, A. W., Ji, Z., Wu, X., and Yu, N. (2017). Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641.

Hanson, R., Dobrovitski, V., Feiguin, A., Gywat, O., and Awschalom, D. (2008). Coherent dynamics of a single spin interacting with an adjustable spin bath. *Science*, 320(5874):352–355.

Harlow, D. and Ooguri, H. (2018). Symmetries in quantum field theory and quantum gravity. *arXiv preprint arXiv:1810.05338v2*.

Harlow, D. and Ooguri, H. (2019). Constraints on symmetries from holography. *Physical Review Letters*, 122:191601.

Hayashi, M. (2002). Two quantum analogues of fisher information from a large deviation viewpoint of quantum estimation. *Journal of Physics A: Mathematical and General*, 35(36):7689.

Hayashi, M. (2005). *Asymptotic theory of quantum statistical inference: selected papers*. World Scientific.

Hayashi, M. (2009). Discrimination of two channels by adaptive methods and its application to quantum system. *IEEE Transactions on Information Theory*, 55(8):3807–3820.

Hayashi, M. (2016). *Quantum information theory*. Springer.

Hayashi, M. and Matsumoto, K. (2008). Asymptotic performance of optimal state estimation in qubit system. *Journal of Mathematical Physics*, 49(10):102101.

Hayden, P., Horodecki, M., Winter, A., and Yard, J. (2008). A decoupling approach to the quantum capacity. *Open Systems & Information Dynamics*, 15(01):7–19.

Hayden, P., Nezami, S., Popescu, S., and Salton, G. (2017). Error correction of quantum reference frame information. *arXiv preprint arXiv:1709.04471v1*.

Heeres, R. W., Reinhold, P., Ofek, N., Frunzio, L., Jiang, L., Devoret, M. H., and Schoelkopf, R. J. (2017). Implementing a universal gate set on a logical qubit encoded in an oscillator. *Nature Communications*, 8(1):94.

Heisenberg, W. (1949). *The physical principles of the quantum theory*. Courier Corporation.

Helstrom, C. (1968). The minimum variance of estimates in quantum signal detection. *IEEE Transactions on Information Theory*, 14(2):234–242.

Helstrom, C. W. (1967). Minimum mean-squared error of estimates in quantum statistics. *Physics letters A*, 25(2):101–102.

Helstrom, C. W. (1974). Estimation of a displacement parameter of a quantum system. *International Journal of Theoretical Physics*, 11(6):357–378.

Helstrom, C. W. (1976). *Quantum detection and estimation theory*. Academic press New York.

Herrera-Martí, D. A., Gefen, T., Aharonov, D., Katz, N., and Retzker, A. (2015). Quantum error-correction-enhanced magnetometer overcoming the limit imposed by relaxation. *Physical Review Letters*, 115(20):200501.

Hirose, M. and Cappellaro, P. (2016). Coherent feedback control of a single qubit in diamond. *Nature*, 532(7597):77–80.

Holevo, A. (1979). Covariant measurements and uncertainty relations. *Reports on Mathematical Physics*, 16(3):385–400.

Holevo, A. (1998). The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273.

Holevo, A. S. (1982). *Probabilistic and statistical aspects of quantum theory*. North Holland.

Holland, M. and Burnett, K. (1993). Interferometric detection of optical phase shifts at the heisenberg limit. *Physical Review Letters*, 71(9):1355.

Horodecki, R., Horodecki, P., Horodecki, M., and Horodecki, K. (2009). Quantum entanglement. *Reviews of Modern Physics*, 81(2):865.

Hotta, M., Karasawa, T., and Ozawa, M. (2005). Ancilla-assisted enhancement of channel estimation for low-noise parameters. *Physical Review A*, 72(5):052334.

Hotta, M., Karasawa, T., and Ozawa, M. (2006). N-body-extended channel estimation for low-noise parameters. *Journal of Physics A: Mathematical and General*, 39(46):14465.

Hu, L., Ma, Y., Cai, W., Mu, X., Xu, Y., Wang, W., Wu, Y., Wang, H., Song, Y., Zou, C.-L., et al. (2019). Quantum error correction and universal gate set operation on a binomial bosonic logical qubit. *Nature Physics*, 15(5):503–508.

Huang, H.-Y., Kueng, R., and Preskill, J. (2020). Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057.

Hübner, M. (1992). Explicit computation of the bures distance for density matrices. *Physics Letters A*, 163(4):239–242.

Huelga, S. F., Macchiavello, C., Pellizzari, T., Ekert, A. K., Plenio, M. B., and Cirac, J. I. (1997). Improvement of frequency standards with quantum entanglement. *Physical Review Letters*, 79:3865–3868.

Huffman, W. C. and Pless, V. (2010). *Fundamentals of error-correcting codes*. Cambridge university press.

Humphreys, P. C., Barbieri, M., Datta, A., and Walmsley, I. A. (2013). Quantum enhanced multiple phase estimation. *Physical Review Letters*, 111(7):070403.

Imai, H. and Fujiwara, A. (2007). Geometry of optimal estimation scheme for su (d) channels. *Journal of Physics A: Mathematical and Theoretical*, 40(16):4391.

Ji, Z., Wang, G., Duan, R., Feng, Y., and Ying, M. (2008). Parameter estimation of quantum channels. *IEEE Transactions on Information Theory*, 54(11):5172–5185.

Jochym-O'Connor, T., Kubica, A., and Yoder, T. J. (2018). Disjointness of stabilizer codes and limitations on fault-tolerant logical gates. *Physical Review X*, 8:021047.

Kahn, J. and Guţă, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Communications in Mathematical Physics*, 289(2):597–652.

Kapourniotis, T. and Datta, A. (2019). Fault-tolerant quantum metrology. *Physical Review A*, 100:022335.

Katariya, V. and Wilde, M. M. (2020a). Geometric distinguishability measures limit quantum channel estimation and discrimination. *arXiv preprint arXiv:2004.10708v1*.

Katariya, V. and Wilde, M. M. (2020b). Rld fisher information bound for multiparameter estimation of quantum channels. *arXiv preprint arXiv:2008.11178v1*.

Kay, S. M. (1993). *Fundamentals of statistical signal processing*. Prentice Hall PTR.

Kessler, E. M., Lovchinsky, I., Sushkov, A. O., and Lukin, M. D. (2014). Quantum error correction for metrology. *Physical Review Letters*, 112:150802.

Kitagawa, M. and Ueda, M. (1993). Squeezed spin states. *Physical Review A*, 47:5138–5143.

Knill, E. and Laflamme, R. (1996). Concatenated quantum codes. *arXiv preprint quant-ph/9608012*.

Knill, E. and Laflamme, R. (1997). Theory of quantum error-correcting codes. *Physical Review A*, 55:900–911.

Knill, E., Laflamme, R., and Viola, L. (2000). Theory of quantum error correction for general noise. *Physical Review Letters*, 84(11):2525.

Knysh, S. I., Chen, E. H., and Durkin, G. A. (2014). True limits to precision via unique quantum probe. *arXiv preprint arXiv:1402.0495v1*.

Kobayashi, H., Mark, B. L., and Turin, W. (2011). *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance*. Cambridge University Press.

Koczor, B., Endo, S., Jones, T., Matsuzaki, Y., and Benjamin, S. C. (2020). Variational-state quantum metrology. *New Journal of Physics*, 22(8):083038.

Kołodyński, J. and Demkowicz-Dobrzański, R. (2013). Efficient tools for quantum metrology with uncorrelated noise. *New Journal of Physics*, 15(7):073043.

Komiya, H. (1988). Elementary proof for sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7.

Koschorreck, M., Napolitano, M., Dubost, B., and Mitchell, M. (2010). Sub-projection-noise sensitivity in broadband atomic magnetometry. *Physical Review Letters*, 104(9):093602.

Kosut, R. L. and Lidar, D. A. (2009). Quantum error correction via convex optimization. *Quantum Information Processing*, 8(5):443–459.

Kubica, A. and Demkowicz-Dobrzański, R. (2020). Using quantum metrological bounds in quantum error correction: A simple proof of the approximate eastin-knill theorem. *arXiv preprint arXiv:2004.11893v1*.

Kura, N. and Ueda, M. (2018). Finite-error metrological bounds on multiparameter hamiltonian estimation. *Physical Review A*, 97(1):012101.

Laurenza, R., Lupo, C., Spedalieri, G., Braunstein, S. L., and Pirandola, S. (2018). Channel simulation in quantum metrology. *Quantum Measurements and Quantum Metrology*, 5(1):1–12.

Layden, D. and Cappellaro, P. (2018). Spatial noise filtering through error correction for quantum sensing. *npj Quantum Information*, 4(1):30.

Layden, D., Chen, M., and Cappellaro, P. (2020). Efficient quantum error correction of dephasing induced by a common fluctuator. *Physical Review Letters*, 124(2).

Layden, D., Zhou, S., Cappellaro, P., and Jiang, L. (2019). Ancilla-free quantum error correction codes for quantum metrology. *Physical Review Letters*, 122:040502.

Le Sage, D., Arai, K., Glenn, D. R., DeVience, S. J., Pham, L. M., Rahn-Lee, L., Lukin, M. D., Yacoby, A., Komeili, A., and Walsworth, R. L. (2013). Optical magnetic imaging of living cells. *Nature*, 496(7446):486–489.

Lee, H., Kok, P., and Dowling, J. P. (2002). A quantum rosetta stone for interferometry. *Journal of Modern Optics*, 49(14-15):2325–2338.

Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation*. Springer.

Leibfried, D., Barrett, M. D., Schaetz, T., Britton, J., Chiaverini, J., Itano, W. M., Jost, J. D., Langer, C., and Wineland, D. J. (2004). Toward heisenberg-limited spectroscopy with multiparticle entangled states. *Science*, 304(5676):1476–1478.

Leroux, I. D., Schleier-Smith, M. H., and Vuletić, V. (2010). Implementation of cavity squeezing of a collective atomic spin. *Physical Review Letters*, 104(7):073602.

Leung, D. W., Nielsen, M. A., Chuang, I. L., and Yamamoto, Y. (1997). Approximate quantum error correction can lead to better codes. *Physical Review A*, 56(4):2567.

Li, K. et al. (2014). Second-order asymptotics for quantum hypothesis testing. *Annals of Statistics*, 42(1):171–189.

Lidar, D. A. and Brun, T. A. (2013). *Quantum error correction*. Cambridge university press.

Lidar, D. A., Chuang, I. L., and Whaley, K. B. (1998). Decoherence-free subspaces for quantum computation. *Physical Review Letters*, 81(12):2594.

LIGO Collaboration (2011). A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nature Physics*, 7(12):962.

LIGO Collaboration (2013). Enhanced sensitivity of the ligo gravitational wave detector by using squeezed states of light. *Nature Photonics*, 7(8):613–619.

Lindblad, G. (1976). On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130.

Liu, J. and Yuan, H. (2017a). Control-enhanced multiparameter quantum estimation. *Physical Review A*, 96(4):042114.

Liu, J. and Yuan, H. (2017b). Quantum parameter estimation with optimal control. *Physical Review A*, 96(1):012117.

Louchet-Chauvet, A., Appel, J., Renema, J. J., Oblak, D., Kjaergaard, N., and Polzik, E. S. (2010). Entanglement-assisted atomic clock beyond the projection noise limit. *New Journal of Physics*, 12(6):065032.

Lu, X.-M., Yu, S., and Oh, C. (2015). Robust quantum metrological schemes based on protection of quantum fisher information. *Nature Communications*, 6:7282.

Maccone, L. and Giovannetti, V. (2011). Beauty and the noisy beast. *Nature Physics*, 7(5):376–377.

MacWilliams, F. J. and Sloane, N. J. A. (1977). *The theory of error correcting codes*, volume 16. Elsevier.

Mason, J. C. and Handscomb, D. C. (2002). *Chebyshev polynomials*. Chapman and Hall/CRC.

Matsumoto, K. (2002). A new approach to the cramér-rao-type bound of the pure-state model. *Journal of Physics A: Mathematical and General*, 35(13):3111.

Matsumoto, K. (2010). On metric of quantum channel spaces. *arXiv preprint arXiv:1006.0300*.

Matsuzaki, Y. and Benjamin, S. (2017). Magnetic-field sensing with quantum error detection under the effect of energy relaxation. *Physical Review A*, 95(3):032303.

Matsuzaki, Y., Benjamin, S. C., and Fitzsimons, J. (2011). Magnetic field sensing beyond the standard quantum limit under the effect of decoherence. *Physical Review A*, 84:012103.

Meyer, J. J., Borregaard, J., and Eisert, J. (2020). A variational toolbox for quantum multi-parameter estimation. *arXiv preprint arXiv:2006.06303v1*.

Michael, M. H., Silveri, M., Brierley, R., Albert, V. V., Salmilehto, J., Jiang, L., and Girvin, S. M. (2016). New class of quantum error-correcting codes for a bosonic mode. *Physical Review X*, 6(3):031006.

Mirsky, L. (1960). Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, 11(1):50–59.

Mitchell, M. W., Lundeen, J. S., and Steinberg, A. M. (2004). Super-resolving phase measurements with a multiphoton entangled state. *Nature*, 429(6988):161–164.

Nagaoka, H. (1988). An asymptotic efficient estimator for a one-dimensional parametric model of quantum statistical operators. *Proceedings of IEEE International Symposium on Information Theory*, page 198.

Nagaoka, H. (1989a). A new approach to cramér-rao bounds for quantum state estimation. *IEICE Technical Report*, IT 89-42:9–14.

Nagaoka, H. (1989b). On the parameter estimation problem for quantum statistical models. *Proceedings of 12th Symposium on Information Theory and Its Applications*, page 577–582.

Nagata, T., Okamoto, R., O'brien, J. L., Sasaki, K., and Takeuchi, S. (2007). Beating the standard quantum limit with four-entangled photons. *Science*, 316(5825):726–729.

Neumann, P., Mizuochi, N., Rempp, F., Hemmer, P., Watanabe, H., Yamasaki, S., Jacques, V., Gaebel, T., Jelezko, F., and Wrachtrup, J. (2008). Multipartite entanglement among single spins in diamond. *Science*, 320(5881):1326–1329.

Ng, H. K. and Mandayam, P. (2010). Simple approach to approximate quantum error correction based on the transpose channel. *Physical Review A*, 81:062342.

Nielsen, M. A. and Chuang, I. L. (2010). *Quantum Computation and Quantum Information*. Cambridge University Press.

O'Donnell, R. and Wright, J. (2016). Efficient quantum tomography. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 899–912.

Ofek, N., Petrenko, A., Heeres, R., Reinhold, P., Leghtas, Z., Vlastakis, B., Liu, Y., Frunzio, L., Girvin, S., Jiang, L., et al. (2016). Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature*, 536(7617):441–445.

Ouyang, Y., Shettell, N., and Markham, D. (2019). Robust quantum metrology with explicit symmetric states. *arXiv preprint arXiv:1908.02378v3*.

Ozeri, R. (2013). Heisenberg limited metrology using quantum error-correction codes. *arXiv preprint arXiv:1310.3432*.

Paris, M. G. (2009). Quantum estimation for quantum technology. *International Journal of Quantum Information*, 7(supp01):125–137.

Pastawski, F. and Yoshida, B. (2015). Fault-tolerant logical gates in quantum error-correcting codes. *Physical Review A*, 91:012305.

Petz, D. and Ghinea, C. (2010). Introduction to quantum fisher information. *arXiv preprint arXiv:1008.2417*.

Pezzé, L. and Smerzi, A. (2009). Entanglement, nonlinear dynamics, and the heisenberg limit. *Physical Review Letters*, 102:100401.

Pezzè, L., Smerzi, A., Oberthaler, M. K., Schmied, R., and Treutlein, P. (2018). Quantum metrology with nonclassical states of atomic ensembles. *Reviews of Modern Physics*, 90:035005.

Pirandola, S., Bardhan, B. R., Gehring, T., Weedbrook, C., and Lloyd, S. (2018). Advances in photonic quantum sensing. *Nature Photonics*, 12(12):724.

Pirandola, S., Laurenza, R., Lupo, C., and Pereira, J. L. (2019). Fundamental limits to quantum channel discrimination. *npj Quantum Information*, 5(1):1–8.

Pirandola, S., Laurenza, R., Ottaviani, C., and Banchi, L. (2017). Fundamental limits of repeaterless quantum communications. *Nature communications*, 8(1):1–15.

Pirandola, S. and Lupo, C. (2017). Ultimate precision of adaptive noise estimation. *Physical Review Letters*, 118:100502.

Plenio, M. B. and Huelga, S. F. (2016). Sensing in the presence of an observed environment. *Physical Review A*, 93:032123.

Preskill, J. (2000). Quantum clock synchronization and quantum error correction. *arXiv preprint quant-ph/0010098*.

Ragy, S., Jarzyna, M., and Demkowicz-Dobrzański, R. (2016). Compatibility in multiparameter quantum metrology. *Physical Review A*, 94(5):052108.

Reimpell, M. and Werner, R. F. (2005). Iterative optimization of quantum error correcting codes. *Physical Review Letters*, 94(8):080501.

Reiter, F., Sørensen, A. S., Zoller, P., and Muschik, C. (2017). Dissipative quantum error correction and application to quantum sensing with trapped ions. *Nature Communications*, 8(1):1822.

Resch, K. J., Pregnell, K. L., Prevedel, R., Gilchrist, A., Pryde, G. J., O'Brien, J. L., and White, A. G. (2007). Time-reversal and super-resolving phase measurements. *Physical Review Letters*, 98(22):223601.

Rojkov, I., Layden, D., Cappellaro, P., Home, J., and Reiter, F. (2021). Bias in error-corrected quantum sensing. *arXiv preprint arXiv:2101.05817v1*.

Roos, C. F., Chwalla, M., Kim, K., Riebe, M., and Blatt, R. (2006). 'designer atoms' for quantum metrology. *Nature*, 443(7109):316–319.

Rosenband, T., Hume, D., Schmidt, P., Chou, C.-W., Brusch, A., Lorini, L., Oskay, W., Drullinger, R. E., Fortier, T. M., Stalnaker, J. E., et al. (2008). Frequency ratio of al+ and hg+ single-ion optical clocks; metrology at the 17th decimal place. *Science*, 319(5871):1808–1812.

Rosenblum, S., Gao, Y. Y., Reinhold, P., Wang, C., Axline, C. J., Frunzio, L., Girvin, S. M., Jiang, L., Mirrahimi, M., Devoret, M. H., and Schoelkopf, R. J. (2018). A CNOT gate between multiphoton qubits encoded in two cavities. *Nature Communications*, 9(1).

Safranek, D. (2017). Discontinuities of the quantum fisher information and the bures metric. *Physical Review A*, 95(5):052320.

Sanders, B. C. and Milburn, G. J. (1995). Optimal quantum measurements for phase estimation. *Physical Review Letters*, 75:2944–2947.

Schmidt, P. O., Rosenband, T., Langer, C., Itano, W. M., Bergquist, J. C., and Wineland, D. J. (2005). Spectroscopy using quantum logic. *Science*, 309(5735):749–752.

Schumacher, B. (1996). Sending entanglement through noisy quantum channels. *Physical Review A*, 54:2614–2628.

Schumacher, B. and Westmoreland, M. D. (1997). Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138.

Schumacher, B. and Westmoreland, M. D. (2002). Approximate quantum error correction. *Quantum Information Processing*, 1(1):5–12.

Sekatski, P., Skotiniotis, M., Kołodyński, J., and Dür, W. (2017). Quantum metrology with full and fast quantum control. *Quantum*, 1:27.

Sewell, R., Koschorreck, M., Napolitano, M., Dubost, B., Behbood, N., and Mitchell, M. (2012). Magnetic sensitivity beyond the projection noise limit by spin squeezing. *Physical Review Letters*, 109(25):253605.

Shen, C., Noh, K., Albert, V. V., Krastanov, S., Devoret, M. H., Schoelkopf, R. J., Girvin, S., and Jiang, L. (2017). Quantum channel construction with circuit quantum electrodynamics. *Physical Review B*, 95(13):134501.

Shor, P. W. (1995). Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52(4):R2493.

Shor, P. W. (1999). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41(2):303–332.

Smirne, A., Kołodyński, J., Huelga, S. F., and Demkowicz-Dobrzański, R. (2016). Ultimate precision limits for noisy frequency estimation. *Physical Review Letters*, 116:120801.

Steane, A. (1996a). Multiple-particle interference and quantum error correction. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 452(1954):2551–2577.

Steane, A. M. (1996b). Error correcting codes in quantum theory. *Physical Review Letters*, 77(5):793–797.

Steinert, S., Ziem, F., Hall, L., Zappe, A., Schweikert, M., Götz, N., Aird, A., Balasubramanian, G., Hollenberg, L., and Wrachtrup, J. (2013). Magnetic spin imaging under ambient conditions with sub-cellular resolution. *Nature Communications*, 4(1):1–6.

Sun, L., Petrenko, A., Leghtas, Z., Vlastakis, B., Kirchmair, G., Sliwa, K., Narla, A., Hatridge, M., Shankar, S., Blumoff, J., et al. (2014). Tracking photon jumps with repeated quantum non-demolition parity measurements. *Nature*, 511(7510):444–448.

Takeoka, M. and Wilde, M. M. (2016). Optimal estimation and discrimination of excess noise in thermal and amplifier channels. *arXiv preprint arXiv:1611.09165v1*.

Tan, K. C., Omkar, S., and Jeong, H. (2019). Quantum-error-correction-assisted quantum metrology without entanglement. *Physical Review A*, 100(2):022312.

Taylor, J., Cappellaro, P., Childress, L., Jiang, L., Budker, D., Hemmer, P., Yacoby, A., Walsworth, R., and Lukin, M. (2008). High-sensitivity diamond magnetometer with nanoscale resolution. *Nature Physics*, 4(10):810–816.

Tomamichel, M. and Hayashi, M. (2013). A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59(11):7693–7710.

Tsang, M., Nair, R., and Lu, X.-M. (2016). Quantum theory of superresolution for two incoherent optical point sources. *Physical Review X*, 6(3):031033.

Tyson, J. (2010). Two-sided bounds on minimum-error quantum measurement, on the reversibility of quantum dynamics, and on maximum overlap using directional iterates. *Journal of Mathematical Physics*, 51(9):092204.

Ulam-Orgikh, D. and Kitagawa, M. (2001). Spin squeezing and decoherence limit in ramsey spectroscopy. *Physical Review A*, 64:052106.

Unden, T., Balasubramanian, P., Louzon, D., Vinkler, Y., Plenio, M. B., Markham, M., Twitchen, D., Stacey, A., Lovchinsky, I., Sushkov, A. O., Lukin, M. D., Retzker, A., Naydenov, B., McGuinness, L. P., and Jelezko, F. (2016). Quantum metrology enhanced by repetitive quantum error correction. *Physical Review Letters*, 116:230502.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge university press.

Vidal, G., Latorre, J., Pascual, P., and Tarrach, R. (1999). Optimal minimal measurements of mixed states. *Physical Review A*, 60(1):126.

Waldherr, G., Wang, Y., Zaiser, S., Jamali, M., Schulte-Herbrüggen, T., Abe, H., Ohshima, T., Isoya, J., Du, J. F., Neumann, P., and Wrachtrup, J. (2014). Quantum error correction in a solid-state hybrid spin register. *Nature*, 506(7487):204–207.

Walls, D. F. and Milburn, G. J. (2007). *Quantum optics*. Springer Science & Business Media.

Walther, P., Pan, J.-W., Aspelmeyer, M., Ursin, R., Gasparoni, S., and Zeilinger, A. (2004). De broglie wavelength of a non-local four-photon state. *Nature*, 429(6988):158–161.

Wang, D.-S., Zhu, G., Okay, C., and Laflamme, R. (2020). Quasi-exact quantum computation. *Physical Review Research*, 2(3):033116.

Wang, X. and Wilde, M. M. (2019). Resource theory of asymmetric distinguishability for quantum channels. *Physical Review Research*, 1(3):033169.

Watrous, J. (2018). *The theory of quantum information*. Cambridge University Press.

Wilde, M. M. (2013). *Quantum information theory*. Cambridge University Press.

Wilde, M. M., Berta, M., Hirche, C., and Kaur, E. (2020). Amortized channel divergence for asymptotic quantum channel discrimination. *Letters in Mathematical Physics*, pages 1–60.

Wilde, M. M., Tomamichel, M., and Berta, M. (2017). Converse bounds for private communication over quantum channels. *IEEE Transactions on Information Theory*, 63(3):1792–1817.

Wineland, D. J., Bollinger, J. J., Itano, W. M., and Heinzen, D. (1994). Squeezed atomic states and projection noise in spectroscopy. *Physical Review A*, 50(1):67.

Wineland, D. J., Bollinger, J. J., Itano, W. M., Moore, F. L., and Heinzen, D. J. (1992). Spin squeezing and reduced quantum noise in spectroscopy. *Physical Review A*, 46:R6797–R6800.

Wineland, D. J., Monroe, C., Itano, W. M., Leibfried, D., King, B. E., and Meekhof, D. M. (1998). Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *Journal of research of the National Institute of Standards and Technology*, 103(3):259.

Woods, M. P. and Alhambra, Á. M. (2020). Continuous groups of transversal gates for quantum error correcting codes from finite clock reference frames. *Quantum*, 4:245.

Xu, H., Li, J., Liu, L., Wang, Y., Yuan, H., and Wang, X. (2019). Generalizable control for quantum parameter estimation through reinforcement learning. *npj Quantum Information*, 5(1):1–8.

Yamagata, K., Fujiwara, A., and Gill, R. D. (2013). Quantum local asymptotic normality based on a new quantum likelihood ratio. *Annals of Statistics*, 41(4):2197–2217.

Yang, Y. (2019). Memory effects in quantum metrology. *Physical Review Letters*, 123:110501.

Yang, Y., Chiribella, G., and Hayashi, M. (2019). Attaining the ultimate precision limit in quantum state estimation. *Communications in Mathematical Physics*, 368(1):223–293.

Yuan, H. (2016). Sequential feedback scheme outperforms the parallel scheme for hamiltonian parameter estimation. *Physical Review Letters*, 117:160801.

Yuan, H. and Fung, C.-H. F. (2017). Fidelity and fisher information on quantum channels. *New Journal of Physics*, 19(11):113039.

Yuen, H. and Lax, M. (1973). Multiple-parameter quantum estimation and measurement of nonselfadjoint observables. *IEEE Transactions on Information Theory*, 19(6):740–750.

Yurke, B. (1986). Input states for enhancement of fermion interferometer sensitivity. *Physical Review Letters*, 56(15):1515–1517.

Yurke, B., McCall, S. L., and Klauder, J. R. (1986). Su(2) and su(1,1) interferometers. *Physical Review A*, 33:4033–4054.

Zhou, S. and Jiang, L. (2019). An exact correspondence between the quantum fisher information and the bures metric. *arXiv preprint arXiv:1910.08473v1*.

Zhou, S. and Jiang, L. (2020a). Asymptotic theory of quantum channel estimation. *arXiv preprint arXiv:2003.10559v2*.

Zhou, S. and Jiang, L. (2020b). Optimal approximate quantum error correction for quantum metrology. *Physical Review Research*, 2:013235.

Zhou, S., Liu, Z.-W., and Jiang, L. (2020). New perspectives on covariant quantum error correction. *arXiv preprint arXiv:2005.11918v2*.

Zhou, S., Zhang, M., Preskill, J., and Jiang, L. (2018). Achieving the heisenberg limit in quantum metrology using quantum error correction. *Nature Communications*, 9(1):78.

Zhuang, Q., Preskill, J., and Jiang, L. (2020). Distributed quantum sensing enhanced by continuous-variable error correction. *New Journal of Physics*, 22(2):022001.