Class of Sciences

PhD thesis in Nanosciences

XXXIII Cycle, Academic year 2020/2021

Quantum statistical inference and communication

Candidate
Marco Fanizza

Supervisor
Vittorio Giovannetti

Abstract

This thesis studies the limits on the performances of inference tasks with quantum data and quantum operations. Our results can be divided in two main parts.

In the first part, we study how to infer relative properties of sets of quantum states, given a certain amount of copies of the states. We investigate the performance of optimal inference strategies according to several figures of merit which quantifies the precision of the inference. Since we are not interested in obtaining a complete reconstruction of the states, optimal strategies do not require to perform quantum tomography. In particular, we address the following problems:

- We evaluate the asymptotic error probabilities of optimal learning machines for quantum state discrimination. Here, a machine receives a number of copies of a pair of unknown states, which can be seen as training data, together with a test system which is initialized in one of the states of the pair with equal probability. The goal is to implement a measurement to discriminate in which state the test system is, minimizing the error probability. We analyze the optimal strategies for a number of different settings, differing on the prior incomplete information on the states available to the agent.

- We evaluate the limits on the precision of the estimation of the overlap between two unknown pure states, given N and M copies of each state. We find an asymptotic expansion of a Fisher information associated with the estimation problem, which gives a lower bound on the mean square error of any estimator. We compute the minimum average mean square error for random pure states, and we evaluate the effect of depolarizing noise on qubit states. We compare the performance of the optimal estimation strategy with the performances of other intuitive strategies, such as the swap test and measurements based on estimating the states.

- We evaluate how many samples from a collection of N d-dimensional states are necessary to understand with high probability if the collection is made of identical states or they differ more than a threshold ϵ according to a motivated closeness measure. The access to copies of the states in the collection is given as follows: each time the agent ask for a copy of the states, the agent receives one of the states

with some fixed probability, together with a different label for each state in the collection. We prove that the problem can be solved with $O(\sqrt{N}d/\epsilon^2)$ copies, and that this scaling is optimal up to a constant independent on d, N, ϵ.

In the second part, we study optimal classical and quantum communication rates for several physically motivated noise models.

- The quantum and private capacities of most realistic channels cannot be evaluated from their regularized expressions. We design several degradable extensions for notable channels, obtaining upper bounds on the quantum and private capacities of the original channels. We obtain sufficient conditions for the degradability of flagged extensions of channels which are convex combination of other channels. These sufficient conditions are easy to verify and simplify the construction of degradable extensions.

- We consider the problem of transmitting classical information with continuous variable systems and an energy constraint, when it is impossible to maintain a shared reference frame and in presence of losses. At variance with phase-insensitive noise models, we show that, in some regimes, squeezing improves the communication rates with respect to coherent state sources and with respect to sources producing up to two-photon Fock states. We give upper and lower bounds on the optimal coherent state rate and show that using part of the energy to repeatedly restore a phase reference is strictly suboptimal for high energies.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Vittorio Giovannetti, for his guidance, his encouragement and for sharing his vision on many aspects of research. I also thank Andrea Mari, who helped me with my first steps in the PhD. I thank the GiQ group at UAB, Barcelona for their warm hospitality, and especially Matteo Rosati, Michalis Skotiniotis and John Calsamiglia, who kindly made my visit possible. We shared countless discussions on our projects, and I thank them for many precious inputs. Matteo, in particular, has decisively helped and influenced me since we started working together. I would also like to thank the Quantum Gravity group at CPT, Marseille, where I started learning how to do research, especially Pietro Donà, Giorgio Sarno, Gabriele V. Stagno, Carlo Rovelli and Simone Speziale.

Thanks to the mobility opportunities of SNS, I had the possibility to attend several conference and schools and receive valuable feedback on my work. I acknowledge helpful discussions with Matthias Christandl, Masahito Hayashi, Ludovico Lami, Felix Leditzky, Ashley Montanaro, Gael Sentís, Xin Wang, Andreas Winter. I would also like to thank the external referees for this thesis, Felix Leditzky and Paolo Perinotti, for their time and precious comments and suggestions.

Part of the results of this thesis come from a collaboration with my colleague and friend Farzad Kianvash: I thank him for all our classical and quantum adventures. I also thank Raffaele Salvia for his enthusiasm in our joint work. At SNS, I also had the pleasure to work with Stefano Chessa, Nicola Dalla Pozza, and Fabio Zoratti, and to have many helpful discussions with Stefano Chessa and Salvatore Tirone.

I thank all the people in the Condensed Matter and Quantum Information Group at SNS, and especially my PhD colleagues during these years: Ashkan, Bibek, Donato, Fabio, Farzad, Federico, Gianmichele, Stefano Chessa, Stefano Cusumano, Marcello, Max, Paolo, Pietro, Raffaele, Salvatore, Vasco, Yu. I learned a lot from all of them and I will keep many great memories of the time spent together.

I am extremely grateful to my family. My parents, Franco and Lina, and my sister Giulia, are always there for me. Finally, a special thanks to Laura: her love, support and kindness have been essential.

Contents

1 Introduction **1**
- 1.1 Learning with quantum data . . . 3
- 1.2 Classical and quantum communication over quantum channels . . . 5
- 1.3 Outline . . . 7
- 1.4 Papers . . . 8

2 Statistical inference in the quantum domain **10**
- 2.1 Quantum states . . . 10
- 2.2 Quantum measurements . . . 12
- 2.3 Quantum channels . . . 13
- 2.4 Quantum statistical inference . . . 14
  - 2.4.1 Quantum discrimination . . . 15
  - 2.4.2 Quantum estimation and tomography . . . 18
  - 2.4.3 Finite size effects and property testing . . . 21

3 Quantum Shannon theory **23**
- 3.1 Von Neumann entropy . . . 24
- 3.2 Capacities of quantum channels . . . 25
  - 3.2.1 Classical communication . . . 25
  - 3.2.2 Private communication . . . 28
  - 3.2.3 Quantum communication . . . 29
- 3.3 Bounds on private and quantum capacities . . . 32

4 Symmetries and quantum information processing **37**
- 4.1 Group theory and representation theory . . . 37
  - 4.1.1 Groups and representations . . . 38
  - 4.1.2 Schur's lemma and orthogonality relations . . . 40
  - 4.1.3 Representations of the symmetric group . . . 41
  - 4.1.4 Representations of SU(d) . . . 43
  - 4.1.5 A special case: SU(2) . . . 46

4.2 Symmetries and optimal measurements 47
4.2.1 Covariant and invariant measurements 47
4.2.2 Schur-Weyl duality 51
4.3 Pauli channels 58
4.3.1 Qubit Pauli group 58
4.3.2 Qudit Pauli group 59
4.3.3 Pauli channels 60
4.4 Gaussian channels 60
4.4.1 Displacements 61
4.4.2 Gaussian states 62
4.4.3 Symplectic transformations and the structure of Gaussian states . 63
4.4.4 Gaussian channels 65

5 Learning machines for quantum state discrimination 70
5.1 Introduction 70
5.2 The model 72
5.3 Symmetries of average states 76
5.3.1 Scenario i): mixed states with fixed purity 79
5.3.2 Scenario ii): mixed states with hard sphere prior 85
5.3.3 Scenario iii): pure states with fixed overlap 87
5.3.4 Compatibility between optimal machines 91
5.4 Implementation of the optimal POVM 92
5.5 Remarks 93

6 Optimal overlap estimation 96
6.1 Introduction 96
6.2 Optimal measurements for overlap estimation 98
6.3 Alternative strategies 102
6.4 Measurement invasiveness 105
6.5 Noise-robustness 106
6.6 Gate complexity and noisy implementations 107
6.7 Optimal global mean squared error 108
6.8 Remarks 112

7 Identity testing of collection of quantum states 115
7.1 Introduction 115
7.1.1 Results 116
7.1.2 Related work 118
7.2 Distance measures for collection of distributions 119
7.3 Upper bound on the sample complexity 122

7.3.1 Building the estimator for $\mathcal{M}^2_{HS}$ . . . 122
7.4 Lower bound on the sample complexity . . . 125
7.5 Implementation of the optimal measurement . . . 130
7.6 Remarks . . . 131

8 Designing degradable extensions 132
8.1 Introduction . . . 132
8.2 Sufficient conditions for degradability of flagged extensions . . . 133
8.3 Degradable extensions of Pauli channels . . . 137
8.3.1 Depolarizing channel . . . 140
8.3.2 BB84 channel . . . 141
8.4 Degradable extensions of the generalized amplitude damping channel . . . 144
8.5 Degradable extensions of single mode gauge-covariant Gaussian channels . 147
8.5.1 Flagged extension of the additive gaussian noise . . . 150
8.5.2 Extension of the thermal attenuator . . . 152
8.5.3 Upper bounds for the thermal amplifier . . . 156
8.5.4 Combining data-processing and direct bounds . . . 156
8.6 Remarks . . . 158

9 Classical communication in absence of a shared phase reference 159
9.1 Introduction . . . 159
9.2 Phase-noise model . . . 163
9.2.1 Constrained rates . . . 164
9.3 Phase-noise channel: capacity, covariant and Gaussian rates . . . 164
9.3.1 Classical capacity of the phase-noise channel and Fock encodings . 165
9.3.2 Covariant encodings . . . 166
9.3.3 Gaussian encodings . . . 167
9.4 Bounds on Gaussian communication rates . . . 168
9.4.1 Maximum coherent-state rate and its upper bounds . . . 168
9.4.2 Lower bounds on Gaussian rates via discrete-pulse encodings . . . 171
9.4.3 Limiting behaviour of the maximum coherent-state rate . . . 172
9.4.4 Comparison of all strategies and squeezing advantage . . . 175
9.5 Comparison with Fock encodings in presence of loss . . . 178
9.6 Communication cost of establishing a phase reference . . . 179
9.7 Remarks . . . 179

10 Conclusions 183

A Appendix: statistics of quantum invariant measurements 186
A.1 Spectrum of the average operator of states at fixed overlap . . . 186

A.2 Averaged multiqubit states 188
A.3 Asymptotic expansion of weighted sums 189
A.4 Asymptotics of the Fisher information 193
A.5 Overlap estimation with depolarizing noise 199
A.6 Equivalence of sampling model and Poissonized model 202
A.7 Proof of Proposition 7.3.2 203
A.7.1 Bound on V_1 204
A.7.2 Bound on V_2 207
A.7.3 Bound on $V_1 + V_2$ 208

B Appendix: miscellanea on Gaussian states and channels 210
B.1 Degradability of flagged additive gaussian noise 210
B.2 Coherent information of flagged additive gaussian noise 212
B.3 Coherent information of extended thermal attenuator 212
B.4 Decomposition into irreducible representations of $\mathrm{U}(m)$ 213
B.5 Pure-state ensembles are always optimal among Gaussian encodings . . . 215
B.6 Communicate with phase reference 215
B.7 Squeezed-coherent encodings 217
B.8 Photon number distribution of single mode Gaussian states 218

Chapter 1

Introduction

The development of science and technology depends fundamentally on the possibility to make inference on the basis of experimental data. The wit of the experimenter in choosing the right measurement procedure has always been crucial to make a good observation. In quantum mechanics, however, the choice of the measurement is elevated to a more fundamental position; at variance with the classical idealization of a measurement, where measurements differ in the way they perform a coarse graining of some absolute truth, in principle accessible, different quantum measurements can be fundamentally incompatible; they cannot be seen as different imprecise versions of an idealized measurement. Different measurements that solve two different inference problems, can be completely useless for an inference problem they are not designed for.

This thesis addresses the study of the fundamental limits on several quantum statistical inference tasks of particular interest. The problems we consider are examples of hypothesis testing and communication, generalized to settings where quantum data and operations are allowed. This possibility generates a rich variety of scenarios. In particular, we explore two kinds of problems:

- Optimal testing and estimation of unitarily invariant properties of sets of quantum states.
- Classical and quantum capacities of discrete and continuous variables quantum channels of physical relevance.

The interest in these topics is motivated by both fundamental and technological reasons. On a fundamental level, the study of optimal strategies and performances for quantum inference tasks is an interesting physical problem, where the quantum mechan-

ical constraints on the ability of the observer to make inferences play a manifest role. On the technological side, studying fundamental limits clarifies the ideal goal for any relevant practical strategy. In the light of the steady progress of quantum technology towards quantum sensors [DRC17], quantum internet [WEH18], universal quantum computers [Pre18], a detailed study of the fundamental limits of quantum statistical inference is of pressing importance. In particular, the choice of the problems addressed in this thesis is inspired by practical applications in computation and communication.

On the computation side, the fine-grained control of quantum systems is steadily progressing towards fully programmable and correctable quantum computers, capable of doing tasks for which classical computation is believed to be insufficient. The most natural application is the one advocated originally by Feynman [Fey82], that is to simulate quantum systems. However, Grover's algorithm [Gro96], which gives a quadratic speedup for unstructured search, and especially Shor's algorithm [Sho97] for polynomial time factorization of integers, suggest that quantum computers are definitely more powerful than classical ones. More than two decades of work have been dedicated to explore quantum algorithms, and many more applications are likely unforeseeable. Large-scale quantum computers are believed to be realizable thanks to quantum error correction [Sho95], particularly in its fault-tolerant incarnations [Sho96; Got10], which allow to protect quantum states by encoding them in a logical space which is a subspace of the physical space, such that errors on the logical space can be corrected. Ultimately, due to the intrinsic probabilistic nature of quantum measurement, any quantum algorithm solves an inference problem. Moreover, in a future where fully functional quantum computers are a reality, it is likely that large amount of high dimensional data will be available, similarly to how classical computers motivated the development of statistical learning theory and machine learning [Vap98; LBH15]. In the quantum case, data will be inherently quantum, and fully quantum information processing will be required to probe this data in the most efficient way.

On the communication side, quantum theory imposes fundamental constraints on the optimal communication rates of classical and quantum information in presence of noise, as all the communication channels are fundamentally quantum. Indeed, there are regimes where quantum effects cannot be neglected: this happens already in the familiar setting of classical communication mediated by electromagnetic waves with finite average power [CD94], and it is perhaps more drastic for quantum communication, that studies how to protect quantum states from noise: in the latter case, even if the noise is not so strong that any input is replaced by the same output, quantum communication may be not possible at all, at variance with classical communication [BDS97]. Moreover, quantum effects are not only a hindrance but also a resource: sharing quantum entanglement allows to transfer any quantum state between different physical systems

with a universal protocol using local operations and classical communication (*quantum teleportation* [Ben+93]), communicate at higher rates than permitted by classical communication (*superdense coding* [BW92]), while the probabilistic nature of quantum measurement allows to devise schemes where classical information can be communicated with the guarantee that it cannot be overheard by any third party (a classic example is the *BB84 protocol* [BB14]). It may be helpful to stress that the model of a communication channel is appropriate for any noise acting on a quantum system, being environmental noise in a transmission line between distant parties or local noise acting in time on a quantum memory or processor. For this reason, quantum communication rates for noise models appropriate for a quantum memory give the ultimate limits for the rate between reliable logical and physical qubits that can be guaranteed through quantum error correction.

In the following sections I present the original contributions of this thesis.

1.1 Learning with quantum data

As already emphasized, a fundamental departure from classical hypothesis testing and estimation is the non-uniqueness of the measurement. Optimizing over quantum strategies is thus much more difficult than in the classical case. However, when the prior information is generic, optimal strategies can be obtained using symmetry principles. This principle has a long history of success in quantum information theory [Hay17a], and it is still very fruitful. In particular, we consider settings where the goal is to devise a unique measurement that works well for solving different inference problems, either with average or worst case figures of merit. For such universal devices, representation theory helps to reduce the degrees of freedom of the optimization problem, sometimes in such a way that the optimization can be fully characterized. The quantities of interest may not be available in a closed form, but a detailed analysis can give at least the leading orders of asymptotic expansions in the extensive parameters of the problem.

Chapter 5: Learning machines for quantum state discrimination [FMG19]

The most basic inference task is state discrimination: given a copy of a state with the promise that it is either ρ or σ with equal probability, guess the correct state with a test achieving the lowest probability of error on average. The determination of the optimal test and an analytical formula for the probability of error are among the first results obtained in quantum information theory, known under the name of Holevo-Helstrom theorem [Hel69; Hol73]. We studied the setting where the agent is not provided with the classical description of ρ and σ, but with an equal number n of labeled copies of ρ and σ

which can be used as a training set. In this way, the problem becomes an instance of a supervised quantum learning task. Several previous work have studied different settings for this problem under the name of programmable state discrimination [BH05]. We address the optimization of the minimum error probability when the states ρ and σ are assumed to be randomly extracted from unitarily invariant distributions over the set of states, and the goal is to minimize the average probability of error for random ρ and σ. In the limit of infinite number of copies of ρ and σ, the optimal probability of error reduces to the one given by the Holevo-Helstrom theorem. In addition to the determination of the optimal measurement, we determine the leading orders of the asymptotic expansion in n of the optimal average error probability, improving on previous results [HHH05a; Sen+10; AH11], in various settings: we find such expansion for pure states in any dimension at fixed overlap, for qubits of fixed purities but otherwise random, and for mixed qubit states distributed according to a hard sphere prior. We also produced a circuit to implement the optimal measurement for $n = 1$ and pure qubit states on a physical machine, evaluating the maximum amount of noise tolerable such that the measurement can be carried out sufficiently close to the ideal performances.

Chapter 6: Estimation of overlap between quantum states [Fan+20a]

We consider optimal estimation of the overlap for two unknown pure states in dimension d, given N and M copies of each. We considered an average case scenario, where copies of states with equal overlap are provided after the same Haar-random distributed unitary acts on them. The optimal measurement can be determined using representation theory, therefore one can reduce the analysis to estimators of the overlap given the classical probability distribution of the optimal measurement [BIMT06; GI06; LSB06]. A minimum variance unbiased estimator of the Hilbert-Schmidt distance of two (possibly mixed) states had also been found in [BOW19], together with an analytical expression of the variance. We look at this problem from an information-theoretic viewpoint, evaluating the asymptotic Quantum Fisher information of the family of average input states at fixed overlap, giving a lower bound on the mean square error of any estimator according to the Cramér-Rao bound. We also complete the investigation of the Bayesian estimation initiated by [BIMT06; GI06; LSB06], giving an analytical solution for any d for the average mean square error for pairs of independent random pure states, for any d. We also considered the effect of depolarizing noise on qubit states, giving an asymptotic expression of the optimal average mean square error. These optimality results have been compared with the performances of intuitive local strategies. Most importantly, the results illustrate how a simple measurement for this problem known as the swap test [Buh+01] is substantially suboptimal with respect to the optimal measurement when the overlap is small.

Chapter 7: Testing identity of a collection of quantum states in the sampling model [FSG21]

Given access to copies of a collection of N quantum states in a Hilbert space of dimension d, we study a procedure to decide whether they are all equal or their minimum average distance from a state is larger than a small constant. The copies are labeled and each copy can be one of the states with some prescribed probability. The problem was solved for two states with fixed number of copies by [BOW19], and we develop their construction to deal with this more general case, following the solution of the analogous problem for classical distributions [DK16]. In this case we are not interested in an accurate determination of the average case probability of error, while we care about understanding the required number of copies such that the worst case probability of error is still higher than a fixed threshold. From a realistic quality control perspective, the setting we consider is particularly meaningful: one could imagine that some preparation procedure is ended by some measurement, but different outcomes of the measurement are expected to correspond to the same desired state. Since the outcome of the measurement at the preparation stage is random, the procedure prepares in principle different states for each measurement outcome. We devise a universal procedure to test the hypothesis that the states produced are equal or sufficiently far, with the guarantee that producing $O(\sqrt{N}d)$ copies of random states suffice to guarantee a correct answer with high probability. This dependence on the dimension and the number of states is actually optimal.

1.2 Classical and quantum communication over quantum channels

Quantum Shannon theory [Wil17; Hol19] provides a characterization of the maximum achievable transmission rates (capacities) for classical or quantum data through a quantum channel, as maximizations of entropic functionals. Available characterizations of most capacities cannot be computed algorithmically, since they involve a limit of an infinite sequence of optimization problems, one for each number of uses of the channel. Superadditivity of quantum entropic functionals makes such regularization necessary and can hinder the evaluation of capacities even for simple fundamental channels; see, e.g., [SS96; SY08; Has09; Li+09; SSY11; ZZS17; Zhu+19]. While it is hard to get past regularized expressions in the general case, it is important to get our best understanding of the capacities of physically motivated channels. The depolarizing channel is a paradigmatic example of this tension: despite its simple definition, the evaluation of its quantum capacity has resisted more than 20 years of attempts. We obtained bounds on quantum and private capacities of several fundamental channels, including the depolarizing channel, by tailoring general results to the peculiar structure of the channels of interest. Even

when the capacity is known, it is often difficult to find a code working at good rates which is also practically feasible. We address the problem in the context of continuous variable communication without a phase reference, where we study the performances of codes constructed with Gaussian states, which can be produced by standard optical elements [Ser17].

Chapter 8: Degradable flagged extensions [FKG20; KFG20; FKG21]

For degradable channels [DS05], the coherent information and the private information are additive and equal, therefore the quantum and private capacity are given by a single-letter formula (meaning that a single optimization involving one use of the channel is required to compute the capacity). In [FKG20], we introduced a new upper bound for the quantum capacity of the depolarizing channel using a degradable flagged extension, where an environment is assumed to "friendly" provide partial helpful information to the receiver. To be specific, for a given noisy quantum channel $\mathcal{N} = \sum_j p_j \mathcal{N}_j$ with probability distribution $\{p_j\}$, $\mathcal{N}_j$ channels, and a collection of states σ_j, a flagged extension is $\widehat{\mathcal{N}} = \sum_j p_j \mathcal{N}_j \otimes \sigma_j$. In contrast to previous flagged extensions [SSW08; SS08; Ouy14], our extension of depolarizing channel uses non orthogonal flags which gives a tighter upper bound. The idea was further exploited by bounding the quantum capacity of the depolarizing channel, BB84 channel and amplitude damping channel using approximate degradability [Sut+17; LLS18a] applied to flagged extensions [Wan21]. In [KFG20], we improve these results constructing new exactly degradable flagged extensions [KFG20], which use more than two flags: we exhibit a candidate degrading map for which a sufficient condition for degradability is expressed in terms of simple algebraic conditions involving the Kraus operators of the original channels and the flags. We obtain non-trivial degradable extensions for any mixture of a unitary operator and another channel, with the probability associated to the unitary operator being larger than $1/2$, and also for convex combination of unitary operators. For Pauli channels, these conditions characterize a rich family of degradable flagged extensions, for which an explicit formula for the quantum capacity can be obtained. We also apply these methods to the generalized amplitude damping channel, which is a realistic model of noise acting on superconducting qubits. In [FKG21] we design other degradable extensions for single-mode phase-insensitive Gaussian channels, modeling loss and amplification of electromagnetic signals at fixed frequency. These channels can be classified as thermal attenuators, thermal amplifiers and additive noise channels. We find new degradable Gaussian extensions of the thermal attenuator and the thermal amplifier by improving previous construction based on weak-degradability, while a proper flagged extension is obtained for the Gaussian additive noise, by adapting the construction of [KFG20] to the infinite dimensional setting. With these techniques we can obtain bounds on the quantum and private capacity which are state-of-the-art at the time of the writing.

Chapter 9: Gaussian codes in absence of a shared reference [Fan+20b]

A milestone result in quantum Shannon theory is the computation of the classical capacity for phase-insensitive Gaussian channels [Gio+14]. For these channels the maximum classical information transmission rate is attained by sending coherent states, which can be considered as the most classical quantum states of light. However, this optimal rate can be attained only by assuming the possibility of maintaining a shared reference frame [BRS07] between the sender and the receiver. We considered a memory channel for which complete phase-decoherence takes place after M subsequent uses of the transmission line, which effectively models the loss of a common phase reference within a finite time period. This channel is non-Gaussian, its classical capacity is the same as the identity channel and it is achievable using an encoding with Fock states. Since Fock states are hard to produce, it is important to study the performance of restricted encodings. We showed that the strategy of using part of the energy for preparing a phase-reference state in one mode and using the other modes to communicate with coherent states is in general suboptimal, even at large energies, with respect to random coding with coherent states. For coherent states encodings the channel is a classical-quantum generalization of the Poisson channel, and we can upper bound its capacity using recent results on the classical capacity of the Poisson channel [CR19]. This upper bound is sufficient to show, in the case $M = 1$ (equivalent to photon counting at the receiver's end) that sub-Poissonian squeezed-coherent states surpass the best coherent states rates. This is an example of an advantage of non-classical Gaussian light in a physically-motivated communication context. We show numerical evidence that this result is robust to noise represented by a zero temperature attenuator. Moreover, with sufficient amount of noise and in the right energy regime, binary or ternary encodings with squeezed-coherent states perform better than binary or ternary encodings with low photon number states or coherent states, for $M = 1$ and $M = 2$.

1.3 Outline

The thesis is structured as follows. In Chapter 2, after recalling the fundamental objects in quantum information theory and their properties, we review fundamental quantum statistical inference tasks. In Chapter 3, we review basics of quantum Shannon theory and definitions and characterizations of capacities of quantum channels, with particular attention to the properties of degradable channels. In Chapter 4, after recalling basics of representation theory, we review group theoretic approaches to hypothesis testing and estimation, and we present Pauli channels, Gaussian channels and the group theoretic structures behind them. Chapter 5, 6, 7, 8, 9 present the original results just discussed, and Chapter 9 contains final remarks and perspectives.

1.4 Papers

This thesis is based on the following articles:

- Marco Fanizza, Andrea Mari, and Vittorio Giovannetti. "Optimal Universal Learning Machines for Quantum State Discrimination". In: *IEEE Transactions on Information Theory* 65.9 (2019), pp. 5931–5944. DOI: 10.1109/TIT.2019.2916646. arXiv: 1805.03477
- Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. "Quantum Flags and New Bounds on the Quantum Capacity of the Depolarizing Channel". In: *Physical Review Letters* 125.2 (2020), p. 020503. DOI: 10.1103/PhysRevLett.125.020503. arXiv: 1911.01977
- M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti. "Beyond the Swap Test: Optimal Estimation of Quantum State Overlap". In: *Physical Review Letters* 124.6 (2020), p. 060503. DOI: 10.1103/PhysRevLett.124.060503. arXiv: 1906.10639
- Farzad Kianvash, Marco Fanizza, and Vittorio Giovannetti. *Bounding the quantum capacity with flagged extensions*. 2020. arXiv: 2008.02461
- Marco Fanizza, Matteo Rosati, Michalis Skotiniotis, John Calsamiglia, and Vittorio Giovannetti. *Classical capacity of quantum Gaussian codes without a phase reference: when squeezing helps*. 2020. arXiv: 2006.06522
- Marco Fanizza, Raffaele Salvia, and Vittorio Giovannetti. *Testing identity of collections of quantum states: sample complexity analysis*. 2021. arXiv: 2103.14511
- Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. *Estimating Quantum and Private capacities of Gaussian channels via degradable extensions*. 2021. arXiv: 2103.09569

During my PhD I also coauthored the following papers, which are not included in this thesis:

- Pietro Donà, Marco Fanizza, Giorgio Sarno, and Simone Speziale. "SU(2) graph invariants, Regge actions and polytopes". In: *Classical and Quantum Gravity* 35.4 (2018), p. 045011. DOI: 10.1088/1361-6382/aaa53a. arXiv: 1708.01727
- Farzad Kianvash, Marco Fanizza, and Vittorio Giovannetti. "Optimal quantum subtracting machine". In: *Physical Review A* 99.5 (2019), p. 052319. DOI: 10.1103/PhysRevA.99.052319. arXiv: 1811.07187
- Stefano Chessa, Marco Fanizza, and Vittorio Giovannetti. "Quantum-capacity bounds in spin-network communication channels". In: *Physical Review A* 100.3

(2019), p. 032311. DOI: 10.1103/PhysRevA.100.032311. arXiv: 1905.11920

- Pietro Donà, Marco Fanizza, Giorgio Sarno, and Simone Speziale. "Numerical study of the Lorentzian Engle-Pereira-Rovelli-Livine spin foam amplitude". In: *Physical Review D* 100.10 (2019), p. 106003. DOI: 10.1103/PhysRevD.100.106003. arXiv: 1903.12624

- Pietro Dona, Marco Fanizza, Pierre Martin-Dussaud, and Simone Speziale. *Asymptotics of* $\mathrm{SL}(2,\mathbb{C})$ *coherent invariant tensors*. 2020. arXiv: 2011.13909

Chapter 2

Statistical inference in the quantum domain

2.1 Quantum states

Any system which can be the object of an experiment can be described in quantum theory as a unit-trace positive semi-definite operator on a Hilbert space $\mathcal{H}$. Such objects, called *quantum states*, allow to predict the probability distribution of the outcomes of measurements on the system [Hol11b]. We will consider Hilbert spaces of finite dimension $\dim \mathcal{H} = d$ as well as Hilbert spaces of infinite dimension isomorphic to the space of square integrable functions on $\mathbb{R}^n$, $L^2(\mathbb{R}^n)$. The former case models physical systems with d distinguishable states, such as the states of a spin system, and it is used to describe discrete variable quantum computation and communication. The latter case models n distinguishable quantum particles, and we will use it to describe a system of n modes of the radiation field. The space of states of a Hilbert space $\mathcal{H}$ will be denoted as $\Sigma(\mathcal{H})$ and it is a convex set. The extremal points $\Sigma(\mathcal{H})$ are *pure states* and have a special role in quantum information theory. The projectors on vectors of $\mathcal{H}$ are pure states. Given an orthonormal basis of $\mathcal{H}$ and its associated set of pure state projectors, the set of states that are convex combination of basis projectors is in one-to-one correspondence with the set of probability distribution on the set $[d] := \{1, ..., d\}$ for d-dimensional $\mathcal{H}$ or $\mathbb{N}^n$ if $\mathcal{H} = L^2(\mathbb{R}^n)$. This observation shows that quantum information theory is a non-commutative generalization of classical probability theory. The study of quantum information tasks show how quantum objects differ from classical objects, and several interesting phenomena arise: quantumness makes inference tasks harder, quantum effects

improve performances with respect to classical resources, and new information processing tasks which have no classical analogue are possible.

Using Dirac's notation, we will often denote Hilbert space vectors $v \in \mathcal{H}$ as *kets* $|v\rangle$, their conjugate transpose as *bras* $\langle v|$, and inner products of two vectors w, v as $\langle w|v\rangle$. Following this notation, the image of operator A acting on a vector v is also denoted as $|Av\rangle$, and its conjugate transpose as $\langle Av|$. We also use the notation $\langle v| A |w\rangle := \langle v|Av\rangle$. For an operator $A : H_A \to H_B$ the adjoint $A^\dagger : H_B \to H_A$ is the unique operator (if it exists) satisfying $\langle A^\dagger w|v\rangle_A = \langle w|Av\rangle_B$ for all $|w\rangle \in H_A$ and $|v\rangle \in H_B$, where $\langle\cdot|\cdot\rangle_{A/B}$ are the inner products of $\mathcal{H}_A$ and $\mathcal{H}_A$ respectively. In finite dimension $A^\dagger$ is the conjugate transpose of A. For linear operators between two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ we sometimes uses the notation $V_{A\to B}$ to clarify domain and codomain, and we may write V_A if $\mathcal{H}_A = \mathcal{H}_B$.

An operator $V_{A\to B}$ is an isometry if it preserves the inner product: $\langle v|w\rangle_A = \langle V_{A\to B}v|V_{A\to B}w\rangle_B$; $V_{A\to B}$ is an isometry if and only if $V^\dagger_{B\to A}V_{A\to B} = I_A$, where I_A is the identity operator of A. Moreover, if $V_{A\to B}$ is an isometry $V_{A\to B}V^\dagger_{B\to A}$ is the projector on the range of $V_{A\to B}$. If $\mathcal{H}_A = \mathcal{H}_B$ and $V_A^\dagger$ is also an isometry, V_A is called a unitary operator and $V_A^\dagger V_A = V_A V_A^\dagger = I_A$. We will make use of the Schatten operator norms [Hay17c]: $||A||_p := \mathrm{Tr}\left[\sqrt{A^\dagger A}^p\right]^{1/p}$.

Two separate quantum systems A and B, with associated Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, can be joined as a unique system AB with associated Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. We will sometimes use the notation ρ_{AB} for a *bipartite state* of a system AB. The marginal state ρ_A is defined as $\rho_A := \mathrm{tr}_B[\rho_{AB}]$, where the partial trace is computed as $\mathrm{tr}_B[\rho_{AB}] := \sum_{i\in I}(I_A \otimes \langle i|)\rho_{AB}(I_A \otimes |i\rangle)$, where $\{|i\rangle\}_{i\in I}$ is a basis of $\mathcal{H}_B$, the result being independent on the choice of the basis. To streamline the notation, we will often denote sets with indexed element dropping the subset indicating the index set, e.g. writing $\{|i\rangle\}$ instead of $\{|i\rangle\}_{i\in I}$. ρ_A is sufficient to predict the probability of the outcome of any measurement of the state ρ_{AB} acting only system A. *Product states* are of the form $\rho_{AB} = \rho_A \otimes \rho_B$. *Separable states* are such that they can be written as convex combinations of product states. *Entangled states* are non-separable states, and pure entangled states can always be written according to the *Schmidt decomposition* as $|\psi\rangle_{AB} = \sum_i \sqrt{\lambda_i}\,|i\rangle_A \otimes |i\rangle_B$, with λ_i positive real numbers which correspond to the non-zero eigenvalues of ρ_A and ρ_B, and $\{|i\rangle_A\}$ and $\{|i\rangle_B\}$ are sets of orthonormal vectors of respectively $\mathcal{H}_A$ and $\mathcal{H}_B$. Any mixed state ρ_A can be written in terms of a *purification*, i.e. a pure bipartite state $|\phi\rangle_{AB}$ such that the dimension of $\mathcal{H}_B$ is equal or larger than the dimension of $\mathcal{H}_A$, and $\mathrm{tr}_B[|\phi\rangle\langle\phi|_{AB}] = \rho_A$. The *purity* of a state ρ_A is measured as $\mathrm{tr}\left[\rho_A^2\right] \leq 1$, with the equality holding if and only if ρ_A is pure. A maximally entangled state is a state of the form $|\Gamma\rangle_{AB} = \sum_{i=1}^{\min(d_A,d_B)} \frac{1}{\sqrt{\min(d_A,d_B)}}\,|i\rangle_A \otimes |i\rangle_B$. In inference problems, we usually

assume to have at our disposal many copies of the same source. In this case the Hilbert space we consider is $\mathcal{H}^{\otimes n} := \underbrace{\mathcal{H} \otimes ... \otimes \mathcal{H}}_{n\,times}$ and we denote n copies of ρ as $\rho^{\otimes n}$.

2.2 Quantum measurements

A measurement on $\Sigma(\mathcal{H})$ is described as a *positive operator valued measures* (POVM), i.e. [Hol11b] a function M from the set $\mathcal{A}(\Omega)$ of measurable subsets of a measurable space Ω to positive operators, such that:

- $M(\emptyset) = 0, \qquad M(\Omega) = I;$
- $M(B) \geq 0, \qquad \forall B \in \mathcal{A}(\Omega);$
- For any at most countable disjoint decomposition $\{B_i\}$ of $B \in \mathcal{A}(\Omega)$, $M(B) = \sum_i M(B_i)$, where the series is weakly convergent in the operator sense (a sequence of bounded operators $\{A_i\}$ on $\mathcal{H}$ converges weakly to A if for any $|v\rangle, |w\rangle \in \mathcal{H}$, $\lim_{i\to\infty} \langle v| A_i |w\rangle = \langle v| A |w\rangle$).

We denote the set of POVM on Ω as $\mathcal{M}(\Omega)$. The probability for the event B to occur when the measurement associated to M is performed on a state ρ is

$$p(B) := \mathrm{tr}[M(B)\rho]. \tag{2.1}$$

For discrete probability spaces $[m]$, one can assign a positive semi-definite operator E_i to each $i \in [m]$, and define M as $M(B) = \sum_{i\in B} E_i$. The condition for M to be a POVM is simply $\sum_{i=1}^m E_i = I$. The probability of the event i is $p(i) = \mathrm{tr}[E_i\rho]$.

An *observable* is a random variable $X : \Omega \to \mathbb{R}$ that can be sampled from a measurement. In the finite dimensional case, Hermitian operators $H = H^\dagger = \sum_i \lambda_i P_i$ with eigenvalues $\{\lambda_i\}$ define real observables, which can be sampled by the POVM constructed from the projectors P_i on the eigenspaces of H with eigenvector λ_i: $E_i = P_i$. The expectation value of H on ρ can be calculated as

$$\mathbb{E}_\rho[H] := \mathrm{tr}[H\rho] = \sum_i \mathrm{tr}[P_i\rho]\lambda_i \tag{2.2}$$

More generally, the expectation value of a POVM defined on a subset of $\mathbb{R}^n$, such that $\{E_\lambda\}$, $\lambda \in \mathbb{R}^n$ are positive semi-definite operators and $p(B) = \int_B \mathrm{d}^n\lambda\, \mathrm{tr}[E_\lambda\rho]$, is $\mathbb{E}_\rho[\{E_\lambda\}] := \int \mathrm{d}^n\lambda\, \mathrm{tr}[E_\lambda\rho]\lambda$. In the infinite dimensional case compact Hermitian operators can still be represented in the form $H = \sum_i \lambda_i P_i$, while for self-adjoint non-bounded operators on $L^2(\mathbb{R})$ one has to resort to the continuous spectral representation $H = \int \lambda P(d\lambda)$, where $P(d\lambda)$ is a projector valued measure. For example, the

position operator Q will be written as $Q = \int_{\mathbb{R}} xE(dx)$, where $E(dx)$ is defined by $\mathrm{tr}\left[\int_B E(dx)\,|\psi\rangle\langle\psi|\right] = \int_B dx|\psi(x)|^2$ for all B measurable subsets of $\mathbb{R}$, where $\psi(x)$ is the wave function of $|\psi\rangle$. This also defines a probability density for the position measurement, $p(x) = |\psi(x)|^2$.

The central moments of the random variable associated to the observable H, if they exists, can be computed as $\mu_\rho^{(n)}[H] := \mathrm{tr}[(H - \mathrm{tr}[H\rho])^n\rho]$, and we denote the variance as $\mathrm{Var}_\rho[H] := \mu_\rho^{(2)}[H]$. Analogous definitions can be given for observable arising from general POVM.

2.3 Quantum channels

A generic linear map between operators (super-operator) on $\mathcal{H}_A$ and $\mathcal{H}_B$ will be denoted as $\mathcal{N}_{A\to B}$, with subscript omitted when not needed. The identity map on operators on $\mathcal{H}_A$ will be denoted as $\mathcal{I}_A$. The tensor product operation of two maps $\mathcal{N} \otimes \mathcal{N}'$ is defined by linearity from the action on a basis of product operators, on which it acts as $\mathcal{N} \otimes \mathcal{N}'[X \otimes Y] = \mathcal{N}[X] \otimes \mathcal{N}'[Y]$. The biggest class of transformations of states that is considered physically meaningful is the set of completely positive trace preserving (CPTP) linear maps, i.e. maps $\mathcal{N}_{A\to B}$ such that for any auxiliary Hilbert space $\mathcal{H}_C$:

- $\mathcal{N}_{A\to B} \otimes \mathcal{I}_C[X] \geq 0$ if $X \geq 0$ (completely positive);
- $\mathrm{tr}[\mathcal{N}[X]] = \mathrm{tr}[X]$ (trace preserving).

This class of transformations models noise on quantum hardware and communication lines, therefore it is of fundamental importance in quantum information theory and also referred to as *quantum channels*. In fact, states and measurements can be considered as a special case of quantum channels: states can be written as channels from the trivial Hilbert space of dimension 1 to states, while the probability distribution associated with the outcomes of a POVM $M = \{E_i\}$ acting on a state ρ can be seen as the output of the channel $\mathcal{N}_M[\rho] := \sum_i \mathrm{tr}[E_i\rho]\,|i\rangle\langle i|$.

In addition to the definition we just gave, quantum channels can be also characterized through the *Kraus representation* from a collection of *Kraus operators* $\{K_i\}$ such that:

- $\mathcal{N}[X] = \sum_i K_i X K_i^\dagger$;
- $\sum_i K_i^\dagger K_i = I$.

Any channel between finite-dimensional Hilbert spaces admits a finite Kraus representation. For infinite dimensional channels, a set of Kraus operators of infinite cardinality may be needed. Different sets of Kraus operators $\{K_i'\}$, $\{K_i\}$ for the same channels

are related by an isometry, i.e. $K'_i = \sum_j V_{ij} K_j$, $V^\dagger V = I$ if $|\{K'_i\}| \geq |\{K_i\}|$. Another important representation of quantum channel is the *Stinespring representation*, which shows how any channel $\mathcal{N}_{A\to B}$ can be realized as a unitary interaction $U_{AE\to BE'}$ with an environment prepared in a suitable state $|\tau\rangle\langle\tau|$:

$$\mathcal{N}_{A\to B}[X] = \mathrm{tr}_{E'}[U_{AE\to BE'} X_A \otimes |\tau\rangle\langle\tau|_E U^\dagger_{AE\to BE'}]. \tag{2.3}$$

Equivalently, $U_{AE\to BE'}|\tau\rangle$ can be seen as an arbitrary isometry from A to BE', $V_{A\to BE'}$. Different Stinespring representations of the same channel, given by isometries $V_{A\to BE'}$ and $V'_{A\to BE''}$ with $\dim \mathcal{H}_{E''} \geq \dim \mathcal{H}_{E'}$, are such that $V'_{A\to BE''} = W_{E'\to E''} V_{A\to BE'}$ for some isometry $W_{E'\to E''}$.

Finally, the *Choi-Jamiołkowski representation* establishes a correspondence between channels and states. The *Choi state* $\mathcal{E}_{AR}$ of a linear superoperator $\mathcal{N}_{A\to B}$ is defined as

$$\mathcal{E}_{BR}(\mathcal{N}) := \mathcal{N}_{A\to B} \otimes \mathcal{I}_{R\to R}[|\Gamma\rangle\langle\Gamma|_{AR}], \tag{2.4}$$

where $|\Gamma\rangle_{AR}$ is a maximally entangled state of AR, with $\mathcal{H}_R \cong \mathcal{H}_A$. A linear superoperator $\mathcal{N}_{A\to B}$ is completely positive if and only if its Choi state $\mathcal{E}_{BR}(\mathcal{N}) \geq 0$. For infinite-dimensional systems, an analogue of the Choi-Jamiołkowski representation can be defined [Hol11a].

2.4 Quantum statistical inference

The basic inference problem is to distinguish between two or more hypotheses on the basis of observations. In the language of quantum states, a *property* is a function from quantum states to $\{0, 1\}$, for example a statement that can be verified to be true or false by looking at the mathematical description of the state. A general inference problem would be, given a collection of properties $\{\mathcal{P}_i\}$ and a state ρ, to determine $\mathcal{P}_i(\rho)$ with some quantum strategy. By the Choi-Jamiołkowski representation this construction is immediately lifted to channels. Hypothesis testing solves the following type of problems: given

- a collection of mutually exclusive properties, i.e. if $\mathcal{P}_i(\rho) = 1$ for some i, then $\mathcal{P}_{i'}(\rho) = 0$ for $i \neq i'$, for any state ρ;
- an unknown state ρ for which there exists an i such that $\mathcal{P}_i(\rho) = 1$,

determine which i satisfies $\mathcal{P}_i(\rho) = 1$.

In this section, we review fundamental results on several quantum hypothesis testing paradigms.

2.4.1 Quantum discrimination

A fundamental hypothesis testing problem is state discrimination [BK15; BC09; Ber10]. Here ρ can be one of a collection of states $\{\sigma_i\}$, and the goal is to understand which state it is, performing a measurement on one copy of ρ. To evaluate the performance of the inference strategy one can consider various figures of merit.

In *minimum error discrimination*, one assumes that the state ρ is with probability p_i equal to ρ_i, $i = 1, .., m$ and the goal is to maximize the average probability of success over the possible POVMs $M = \{E_i\}$:

$$p_{\text{succ}}^{\text{opt}} = \sup_{\{E_i\} \in \mathcal{M}[m]} \sum_{i=1}^{m} p_i \operatorname{tr}[E_i \rho_i]. \tag{2.5}$$

This has been one of the first problems to be considered in quantum information theory, and necessary and sufficient conditions for the optimality were immediately found [Hol73; YKL75]. It can be cast as a semi-definite program [EMV03] and solved numerically, although there is not an analytical formula for the optimal probability in the general case. For a collection of two states, one can instead solve the problem exactly:

Theorem 2.4.1 (Holevo-Helstrom theorem [Hel69; Hol73]). *For any binary POVM $\{E_1, E_2\}$, the probability of success for quantum state discrimination of two states ρ and σ with prior probabilities p_1 and p_2 satisfies*

$$p_{\text{succ}} = p_1 \operatorname{tr}[E_1 \rho] + p_2 \operatorname{tr}[E_2 \rho] \leq p_{\text{succ}}^{\text{opt}} = \frac{1}{2} + \frac{||p_1 \rho - p_2 \sigma||_1}{2}, \tag{2.6}$$

The inequality is saturated by a projective measurement, with E_1 being the projector on the eigenspaces with positive eigenvalues of $p_1\rho - p_2\sigma$.

The *trace distance* between ρ and σ is defined as

Definition 2.4.1 (Trace distance).

$$D_{\text{Tr}}(\rho, \sigma) := \frac{1}{2} ||\rho - \sigma||_1. \tag{2.7}$$

It appears in Eq. (2.6) in the case $p_1 = p_2 = \frac{1}{2}$, showing its operational meaning. It will be useful to mention some properties of the trace distance:

- it is a proper distance, since it comes from a norm;
- it is invariant under isometries, $D_{\text{Tr}}(U\rho U^\dagger, U\sigma U^\dagger) = D_{\text{Tr}}(\rho, \sigma)$ for $U^\dagger U = I$;
- it satisfies monotonicity $D_{\text{Tr}}(\mathcal{N}[\rho], \mathcal{N}[\sigma]) \leq D_{\text{Tr}}(\rho, \sigma)$.

In the multi-hypothesis case there are analytic results for symmetric sets of states [Ban+97; EF01; Bar01; CH03; EMV04]. A good guess for a POVM is the *pretty good measurement* [HW94; EF01], defined as

Definition 2.4.2 (Pretty good measurement). For any set of states $\{\sigma_i\}$, and a probability distribution $\{p_i\}$ one can define a POVM called the pretty good measurement, with elements:

$$E_i = \left(\sum_{i'=1}^{m} p_{i'}\sigma_{i'}\right)^{-1/2} p_i\sigma_i \left(\sum_{i'=1}^{m} p_{i'}\sigma_{i'}\right)^{-1/2}. \tag{2.8}$$

The pretty good measurement is optimal for sets of symmetric states [EMV04] and gives a powerful bound on the optimal success probability [BK02]: $p_{\text{succ}}^{\text{PGM}} \geq (p_{\text{succ}}^{\text{opt}})^2$. Other upper and lower bounds can be obtained on the optimal probability of success [Qiu08; Mon08], or on probability of success of the pretty good measurement [Mon07; Mon19].

A different setting is *unambiguous discrimination*, where the POVM has the elements $\{E_i\}_{i\in I} \cup \{E_?\}$, where $E_?$ indicates an inconclusive result. It is requested is that the probability of misidentifying ρ is zero, that is $\text{Tr}[\sigma_i E_{i'}] = 0$ for all $i \neq i'$. The problem is to find the maximum average probability of success. For two pure states this problem was solved in [Iva87; Die88; Per88; JS95]. For n pure states a non-trivial solution exists if the states are linearly independent, and an explicit expression for the optimal probability of success can be obtained for uniform success probabilities [Che98], and for equidistant states [CB98; Roa+11]. The necessary and sufficient conditions for unambiguous discrimination of mixed states were found in [RST03; FDY04]. Upper [Zha+01; Fen+02; NUK18] and lower [NUK18; Lü21] bounds can be obtained. As for the minimum error case, the problem can be cast as a semi-definite program [Eld03a].

More general discrimination problems can be designed if one allows both inconclusive outcomes and wrong answers, such as maximum confidence discrimination [Cro+06], and discrimination with bounded probability of inconclusive answers [FJ03; Eld03b].

The discrimination problem can also be extended to channels. For two quantum channels, there is freedom in choosing an input state such that the output states corresponding to the alternatives have the maximum distinguishability. In particular, for two channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$, one looks at maximizing the distinguishability of the states $\mathcal{N}_{A\to B} \otimes I_C[\rho_{AC}]$ and $\mathcal{M}_{A\to B} \otimes I_C[\rho_{AC}]$, where C is an arbitrary auxiliary system, whose dimension is in principle a free parameter in the optimization problem. According to the Holevo-Helstrom theorem, the maximum trace norm of the weighted difference of the outputs gives the optimal success probability. For equal priors this gives to so-called called diamond-norm distance [Kit97]:

Definition 2.4.3 (Diamond-norm distance).

$$D_{\diamond}(\mathcal{N},\mathcal{M}) := \sup_{\mathcal{H}_C,\sigma_{AC}} D_{\mathrm{Tr}}(\mathcal{N}_{A\to B}\otimes I_C[\rho_{AC}],\mathcal{M}_{A\to B}\otimes I_C[\rho_{AC}]). \tag{2.9}$$

This quantity can be evaluated with a semidefinite program [Wat09], and it can be proven that the dimension of system C can be chosen to be equal to the dimension of A (see [Wat18] for a proof of this fact and many other properties).

We also mention an asymmetric setting for binary state discrimination. In this case ρ is promised to be σ_1 or σ_2. The two type of errors are treated asymmetrically

- Type I error: $\rho=\sigma_1$, outcome 2, probability of error $q_1 := \mathrm{tr}[E_2\sigma_1]$;
- Type II error: $\rho=\sigma_2$, outcome 1, probability of error $q_2 := \mathrm{tr}[E_1\sigma_2]$.

The goal of asymmetric state discrimination is to minimize q_2 when $q_1<\epsilon$, giving a different importance to the errors of the two types. The minimum q_2 at fixed ϵ is also related to the *smooth min entropy* $D^{\epsilon}_{\min}(\rho||\sigma)$ [Dat09; BD10; BD11; WR12]:

Definition 2.4.4 (Smooth min entropy).

$$D^{\epsilon}_{\min}(\rho||\sigma) := -\log_2 \inf_{0\le E\le I}\{\mathrm{tr}[E\rho] : \mathrm{tr}[E\sigma]>1-\epsilon\}. \tag{2.10}$$

The smooth min entropy finds an operational meaning in the resource theory of asymmetric distinguishability [Mat10; WW19a].

If an arbitrarily large number n of copies of the state are available, the probabilities of error will go to zero, as long as the trace distance between each pair of states is bounded below by a constant. This is a fundamental consequence of the fact that as long as the probability of error of pairwise discrimination is bounded from 1/2, repeating the measurement many times will give the right answer with a probability of error decreasing exponentially with respect to n. One can then ask how fast this decay can be for a general measurement. If the decay of the error probability $p_{\mathrm{err}}(n)$, which is considered as a figure of merit, is exponential as a function of n, the error exponent is $p_{err}(n)=\lim_{n\to\infty}-\frac{\log p_{\mathrm{err}}(n)}{n}$. In particular, the asymptotics of $D^{\epsilon}_{\min}(\rho^{\otimes n}||\sigma^{\otimes n})$ gives an operational interpretation to the quantum relative entropy:

Definition 2.4.5 (Quantum relative entropy). If $\mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma)$, we define

$$D(\rho||\sigma) := \mathrm{tr}[\rho(\log\rho-\log\sigma)], \tag{2.11}$$

otherwise $D(\rho||\sigma):=+\infty$.

For commuting states, with eigenvalues given by probability distributions $p = \{p_i\}$ and $q = \{q_i\}$, this definition reduces to the classical relative entropy between two distributions, $D(p||q) = \sum_i p_i \log \frac{p_i}{q_i}$ if $\text{supp}(p) \subseteq \text{supp}(q)$ and $D(p||q) = +\infty$ otherwise. The following result holds, generalizing the classical Stein's lemma [CT05].

Definition 2.4.6 (Quantum Stein lemma [HP91; ON05]).

$$\lim_{n\to\infty} \frac{D^{\epsilon}_{\min}(\rho^{\otimes n}||\sigma^{\otimes n})}{n} = D(\rho||\sigma) \tag{2.12}$$

From the definition, it follows that the relative entropy is additive: $D(\rho_1 \otimes \rho_2||(\sigma_1 \otimes \sigma_2) = D(\rho_1||\rho_2) + D(\sigma_1||\sigma_2)$. The quantum relative entropy is not a distance, because it is not symmetric and it does not satisfy the triangle inequality. However, it satisfies monotonicity $D(\mathcal{N}[\rho]||\mathcal{N}[\sigma]) \leq D(\rho||\sigma)$, and it follows that it is non-negative if ρ and σ are states. The monotonicity of the relative entropy is a deep, non-trivial result which is equivalent to *strong subadditivity*, which we will recall in the context of quantum Shannon theory as Theorem 3.1.1.

If the probability of type I error is constrained to vanish with some error exponent one instead finds the quantum Hoeffding bound [OH04; Nag06; Hay07]. The error exponent of the minimum error probability of discrimination with multiple copies of the state can also be determined, and the result is known as quantum Chernoff bound. With two states this has been done in [NS09; Aud+08], later generalized to multiple states [Li16].

As far as channels are involved, when more copies of the channels can be used, characterizing the optimal strategy is much more complicated [CDP08], since the channels can be accessed in parallel or in sequence. Several results have been obtained in both symmetric and asymmetric settings [Har+10; PW17; Pir+19; WW19b; ZP20a; ZP20b; Wil+20; SHW20].

2.4.2 Quantum estimation and tomography

Another important problem is the following: for a family of continuously parametrized states ρ_θ, $\theta \in \Theta$, suppose to have access to the state ρ_{θ^*}. How well can one infer the value of θ^*? We denote as $\{E_{\tilde{\theta}}\}$ a POVM that estimates the value of θ as $\tilde{\theta}$ when the outcome is $\tilde{\theta}$. We will refer to an estimator $\tilde{\theta}$ implying that it is obtained from a suitable POVM. The probability density of getting outcome $\tilde{\theta}$ when the true parameter is θ is $p(\tilde{\theta}|\theta) = \text{tr}\big[E_{\tilde{\theta}}\rho_\theta\big]$. The expectation value of the estimator $\tilde{\theta}$ is the expectation value of the corresponding POVM, which for ease of notation we now denote $\mathbb{E}_\theta[\tilde{\theta}] := \mathbb{E}_\theta[\{E_{\tilde{\theta}}\}]$, since the statements we will make are valid for any estimator. This problem can be still framed as hypothesis testing, with mutually exclusive properties that are continuously parametrized. However, it is clear that it would be too demanding to ask

for exact determination of the parameter θ^*, even in the classical case. A figure of merit appropriate for this case is the *mean square error* (MSE), defined as

Definition 2.4.7 (Mean square error).

$$\mathrm{MSE}(\theta) := \int_{\tilde{\theta}\in\Theta} \mathrm{d}\tilde{\theta}\, p(\tilde{\theta}|\theta)(\tilde{\theta}-\theta)^2 \tag{2.13}$$

If a prior knowledge on θ is available, expressed in the form of a probability distribution $p(\theta)$, we can also define the *average mean square error* (AvMSE):

Definition 2.4.8 (Average mean square error).

$$\mathrm{AvMSE} := \int_{\theta\in\Theta} p(\theta) \int_{\tilde{\theta}\in\Theta} \mathrm{d}\tilde{\theta}\, p(\tilde{\theta}|\theta)(\tilde{\theta}-\theta)^2. \tag{2.14}$$

Given a family of states $\{\rho_\theta\}$, the choice of an estimator such that the average mean square error is minimized is the subject of Bayesian estimation. This figure of merit can be suitably generalized in the case where there are multiple parameters to estimate, i.e. θ is a vector. In particular, if the family of states is $\{\rho(\theta)^{\otimes n}\}$ when the parametrization is such that any states of $\mathcal{H}$ can be written as $\rho(\theta)$ for some θ, the problem of finding θ^* is known as *quantum tomography*. In the single parameter case, a bound on $\mathrm{MSE}(\theta)$ is given by the quantum version of the Cramér-Rao bound [Hel69; Hol11b], which we state without making regularity requirements precise. The classical Cramér-Rao bound states the following, for any estimator with probability distribution $p(\tilde{\theta}|\theta)$, with expectation value $\mathbb{E}_\theta\tilde{\theta}$.

Theorem 2.4.2 (Cramér-Rao bound [CT05]). *The mean square error of any estimator $\tilde{\theta}$ satisfies*

$$\mathrm{MSE}(\theta) \geq \left[\frac{d\mathbb{E}_\theta[\tilde{\theta}]}{d\theta}\right]^2 \frac{1}{F(\theta)} + (\mathbb{E}_\theta[\tilde{\theta}]-\theta)^2, \tag{2.15}$$

where

$$F(\theta) = \int_{\tilde{\theta}\in\Theta} \frac{1}{p(\tilde{\theta}|\theta)} \left(\frac{\mathrm{d}p(\tilde{\theta}|\theta)}{d\theta}\right)^2. \tag{2.16}$$

$F(\theta)$ is called *Fisher information*. The Fisher information of a POVM applied to a family of states will generally depend on the details of the POVM, but can be bounded from above by an information quantity which depends only on the family of states $\{\rho_\theta\}$, which is called *quantum Fisher information*. It is defined starting from the *symmetric logarithmic derivative*, that is the operator L_θ satisfying $\frac{\mathrm{d}\rho}{d\theta} = \frac{L_\theta\rho_\theta+\rho_\theta L_\theta}{2}$.

Theorem 2.4.3 (Quantum Cramér-Rao bound [Hol11b; Par09]). *The Fisher information $F(\theta)$ of any estimator $\tilde{\theta}$ is bounded as*

$$F(\theta) \leq H(\theta) := \operatorname{tr}\left[\rho_\theta L_\theta^2\right], \tag{2.17}$$

A POVM attaining the bound is given by the projectors on the eigenvectors of L_θ. Therefore, the MSE of any estimator satisfies

$$\mathrm{MSE}(\theta) \geq \left[\frac{d\mathbb{E}_\theta[\tilde{\theta}]}{d\theta}\right]^2 \frac{1}{H(\theta)} + (\mathbb{E}_\theta[\tilde{\theta}] - \theta)^2. \tag{2.18}$$

$H(\theta)$ is called *quantum Fisher information.*

The multi-parameter case is much richer than the single-parameter one [Hol11b; SBD16], and it is beyond the scope of the present thesis. We mention that a class of bounds on the covariance matrix of the estimator, known as Holevo-Cramér-Rao bounds can be computed as a semidefinite program [AFD19], and are asymptotically attainable for the tomography problem, under mild conditions [KG08; YCH19]. We also remark that the problem of finding fundamental bounds for the estimation of a subset of the complete set of parameters is studied, and known as estimation with nuisance parameters [SYH20].

In the single-parameter case, the solution for the Bayesian estimation problem is given in [Per71]:

Theorem 2.4.4 (Optimal Bayesian estimator). *If an estimator $\tilde{\theta}_B$ minimizing the average mean square error is constructed from an observable S, then S satisfies*

$$\frac{S\Gamma + \Gamma S}{2} = \eta, \tag{2.19}$$

where

$$\begin{aligned} \Gamma &:= \int p(\theta)\, \rho_\theta \, \mathrm{d}\theta, \\ \eta &:= \int \theta\, p(\theta)\, \rho_\theta \, \mathrm{d}\theta. \end{aligned} \tag{2.20}$$

In the case of Γ positive definite S is uniquely determined as:

$$S = \int_0^\infty e^{-\alpha\Gamma}\, \eta\, e^{-\alpha\Gamma} \, \mathrm{d}\alpha. \tag{2.21}$$

2.4.3 Finite size effects and property testing

The information-theoretic approach to estimation based on generalization of the Fisher information is appropriate to investigate the asymptotic limit of large number of copies, where the number of copies is larger than any other extensive parameter of the problem. However, it fails to give concrete answers when other parameters can be considered to be large, such as the dimension or the rank. The *sample complexity* of quantum tomography is the number of copies that are necessary and sufficient in order to guarantee that any unknown state can be determined to accuracy ϵ with high probability. Since it is hard to characterize the sample complexity exactly, the goal is to determine its general dependence on the dimension of the Hilbert space. This problem remained unsolved until recently [Haa+17; OW16; OW17], when a solution has been found, establishing the sample complexity of quantum tomography as $O(dr/\epsilon^2)$, where d is the dimension of the supporting Hilbert space, and r is the rank of the state to be determined. This performance is achieved with nonlocal measurements on many copies of the states. Beyond the sample complexity with optimal measurements, several works have studied performances of local measurements with efficient reconstruction algorithms (e.g. [Gro+10; Fla+12; KRT17; Gut+18; AKG19]).

The study of the sample complexity of quantum tomography reflects the general spirit of *property testing*, a concept developed in computer science [Gol17a], and applied to hypothesis testing of distributions [Can20] and quantum states and channels [MW16]. At variance with optimal asymptotic error rates studied in statistical classical and quantum hypothesis testing [Hay17c], the sample complexity captures finite size effects in inference problems. A general binary hypothesis testing problem can be framed as a property testing problem in the following way: given a property $\mathcal{P}$ associated to the null hypothesis, the property associated to the alternative hypothesis is to be ϵ-far from the states satisfying $\mathcal{P}$, for example in trace distance. A typical property testing problem will have the following structure. Find a two-outcome test (binary POVM) acting on $\mathcal{H}^{\otimes n}$ with outcomes `"accept"` and `"reject"` such that, for any $\rho \in \mathcal{H}$ satisfying either case A or B, where

- **Case** A: $\mathcal{P}(\rho) = \text{true}$;
- **Case** B: $\min_{\sigma:\mathcal{P}(\sigma)=\text{true}} D_{\text{Tr}}(\rho,\sigma) > \epsilon$,

when the test is applied to m copies of ρ, the probability of getting `"accept"` is larger than $2/3$ in case A, and smaller than $1/3$ in case B, i.e.

$$\begin{cases} P(\texttt{test} \mapsto \texttt{"accept"} \,|\text{Case A}) > 2/3\,, \\ \\ P(\texttt{test} \mapsto \texttt{"accept"} \,|\text{Case B}) < 1/3\,. \end{cases} \tag{2.22}$$

Note that the values $2/3$ and $1/3$ are entirely conventional, and can be replaced by any constant in $(1/2, 1)$ and $(0, 1/2)$ respectively. The reason is that if for m copies such test exists, one can repeat it a constant number of times and take the majority vote. The probability of error for the test obtained with this classical post-processing goes down exponentially fast in the number of repetitions, independently of the particular test. Therefore, the dependence on the extensive parameters of the problem is already determined by the test with constant probability of error.

Examples of property testing problems are: certifying productness [HM10], certifying stabilizerness [GNW21], certifying identity with a known or unknown state [Mon07; BOW19].

Another important problem which is studied in the finite copies setting is shadow tomography, i.e. estimate expectation values of a set of observables [Aar07; Aar18; AR19; HKP20].

For a review of many other results in the context of certification of quantum states and channels in the non-asymptotic regime, we refer to [KR21].

Chapter 3

Quantum Shannon theory

In a landmark paper [Sha48], Shannon understood how the law of large numbers allows to compress messages and to protect them from noise, founding a mathematical theory of communication. In his model, a source of messages is represented by a probability distribution $\mathbf{p} = (p_1, ..., p_n)$ on d symbols, and $\lceil \log_2 d \rceil$ bits are required to faithfully represent each symbol. A sequence of n symbols will thus require $n\lceil \log_2 d \rceil$ bits. The first of Shannon's results states that as n goes to infinity sequences are contained in a *typical* set of cardinality $2^{nS(\mathbf{p})}$ with probability arbitrarily close to one, where $S(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log_2 p_i$ is the *Shannon entropy*. Therefore, a compression and decompression algorithm using $nS(\mathbf{p})$ bits is able to reliably reconstruct the original message. The second of Shannon's result shows how to characterize the maximum number M of sequences of n symbols of an alphabet $[d]$, such that if some noise (channel) acts independently on each symbol, the receiver can reliably recover the original sequence, in the limit of large n. In fact, the result characterizes the maximum ratio $\frac{\log_2 M}{n}$, and it is called the capacity of the channel. The capacity is never zero, unless the channel is completely noisy, which came as a surprise at the time [CT05].

Quantum Shannon theory generalizes this approach to quantum states and quantum channels. At variance with the classical case, there are several possible communication tasks. In the following we give some definitions and characterizations. Proofs of the results cited can be found in [Wil17; Hol19], beyond the original papers we refer to. We will need these facts in Chapters 8, 9.

3.1 Von Neumann entropy

The generalization of the Shannon entropy for quantum states is the von Neumann entropy, defined as:

Definition 3.1.1 (von Neumann entropy).

$$S(\rho_A) := -\operatorname{Tr}[\rho_A \log_2 \rho_A] \tag{3.1}$$

The von Neumann entropy of a state coincides with the Shannon entropy of the probability distribution associated to the eigenvalues of the state. Therefore, there is no clash in using the same symbol for the two quantities. This notation is also suitable to denote marginal entropies. For a state $\rho_{AB} \in \Sigma(\mathcal{H}_A \otimes \mathcal{H}_B)$, the marginal entropy $S(\rho_A)$ is the entropy of the marginal state ρ_A. We will also use the notation $S(A)_\rho := S(\rho_A)$, when we will need emphasis is on the system A rather than the state ρ.

The von Neumann entropy also characterize the maximum rate of compression of a quantum states: for n large, $\rho^{\otimes n}$ is arbitrarily close to a state supported on a Hilbert space of dimension $2^{nS(\rho)}$. This fact was proven by [Sch95].

We list some properties of von Neumann entropy.

- Positivity: $S(\rho) \geq 0$ for any state ρ, with equality if and only if ρ is pure.
- Maximum value: $S(\rho) \leq \log d$ if $\rho \in \Sigma(\mathbb{C}^d)$.
- Invariance under isometries: $S(\rho_B) = S(V_{A\to B}\rho_A V_{A\to B}^\dagger)$ for any isometry $V_{A\to B}$.
- Concavity: $S(\sum_{i=1}^n p_i\rho_i) \geq \sum_{i=1}^n p_i S(\rho_i)$ for any probability distribution $\mathbf{p} = (p_1, ..., p_n)$ and any collection of states $\{\rho_i\}_{i=1,...,n}$.
- Additivity for tensor products: $S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$.
- Chain rule for classical quantum states: For states of the form $\rho_{AB} = \sum_{i=1}^n p_i \left|i\right\rangle\!\left\langle i\right|_A \otimes (\rho_i)_B$ we have $S(\rho_{AB}) = S(\mathbf{p}) + \sum_{i=1}^n p_i S(\rho_i)$
- Equality of marginal entropies for pure states: $S(\rho_A) = S(\rho_B)$ for ρ_{AB} pure.
- Conditional entropy $S(A|B)_\rho := S(AB)_\rho - S(B)_\rho$ can be negative.

The last property is peculiar for quantum states and it is not a generalization of a classical property, since it is due to the existence of entanglement. From the von Neumann entropy other entropic quantities can be defined, which have operational meaning by themselves. We define:

Definition 3.1.2 (Mutual information).

$$I(A;B)_\rho := S(A)_\rho + S(B)_\rho - S(AB)_\rho. \tag{3.2}$$

A fundamental properties of the mutual information is the following:

Theorem 3.1.1 (Strong subadditivity).

$$I(A;B)_\rho \leq I(A;BC)_\rho. \tag{3.3}$$

While strong subadditivity is elementary in the classical case, the quantum case requires a much more complicated proof [LR73a; LR73b; Lin75]. It is a cornerstone of quantum Shannon theory, as many results follow from this property. For example, strong subadditivity is equivalent to the monotonicity of the relative entropy:

Theorem 3.1.2 (Monotonicity of the relative entropy). *For any channel* $\mathcal{N}$

$$D(\rho||\sigma) \geq D(\mathcal{N}[\rho]||\mathcal{N}[\sigma]). \tag{3.4}$$

Strong subadditivity also guarantees a data processing inequality for mutual information: for any two channels $\mathcal{N}_{A\to A'}, \mathcal{N}'_{B\to B'}$ and any state ρ_{AB}, defining $\sigma_{A'B'} = \mathcal{N}_{A\to A'} \otimes \mathcal{N}'_{B\to B'}[\rho_{AB}]$, we have $I(A;B)_\rho \geq I(A';B')_\sigma$.

Another important entropic quantity is the coherent information, which is equal to the negative of the conditional entropy:

Definition 3.1.3 (Coherent information).

$$I(A\rangle B)_\rho := S(B)_\rho - S(AB)_\rho \tag{3.5}$$

Coherent information also satisfies a data processing inequality: for any channel $\mathcal{N}_{A\to A'}$, and any state ρ_{AB}, defining $\sigma_{AB'} = \mathcal{I}_A \otimes N_{B\to B'}[\rho_{AB}]$, we have $I(A\rangle B)_\rho \geq I(A\rangle B')_\sigma$.

3.2 Capacities of quantum channels

3.2.1 Classical communication

For a classical quantum state of the form

$$\rho_{AB} = \sum_{i=1}^{k} p_i \, |i\rangle\langle i|_A \otimes (\rho_i)_B, \tag{3.6}$$

we can define:

Definition 3.2.1 (Holevo quantity).

$$\chi(\{p_i, \rho_i\}) = S\left(\sum_{i=1}^{k} p_i \rho_i\right) - \sum_{i=1}^{k} p_i S(\rho_i) \tag{3.7}$$

One can verify that $\chi(\{p_i, \rho_i\}) = I(A;B)_\rho$ for a classical quantum state. The Holevo quantity was first introduced as un upper bound for the classical mutual information $I(A;X)$ for the joint probability distribution obtained applying a POVM to the register B of a classical quantum state ρ_{AB} as in Eq. 3.6 [Hol73]. In fact, it has an even more fundamental role in the characterization of the classical capacity of a quantum channel.

We now illustrate the scheme of a classical communication task. Alice wants to send Bob one of $|M|$ messages, or *codewords*. She can do it by sending states $(\rho_m)_{A^n} \in \Sigma\left(\mathcal{H}^{\otimes n}\right)$, where $m \in M$ label the codeword. Bob receives a state $(\mathcal{N}_{A\to B})^{\otimes n}[(\rho_m)_{A^n}]$, and performs a POVM $\{M_{\hat{m}}\}$ to decode the message as $\hat{m}$. The choice of states $(\rho_m)_{A^n}$ and POVM $\{E_{\hat{m}}\}$ defines a code $\mathcal{C}$. The worst-case error probability of the code $\mathcal{C}$ is then

$$p_e(\mathcal{C}) = \max_{m\in M} \mathrm{tr}\big[(I - E_m)\mathcal{N}^{\otimes n}_{A\to B}[(\rho_m)_{A^n}]\big]. \tag{3.8}$$

The rate of a code is

$$R_c(\mathcal{C}) = \frac{\log_2 M}{n} \tag{3.9}$$

and we denote a code as $\mathcal{C}(n, r, \epsilon)$ if it uses n uses of the channel with rate $r = R_c(\mathcal{C})$ and probability of error $\epsilon = p_e(\mathcal{C})$. A rate R_c is called achievable with $\mathcal{N}$ if for any $\delta > 0$, $\epsilon > 0$ there exists a code $C(n, R_c - \delta, \epsilon)$ for sufficiently large n. We can then give the following definition.

Definition 3.2.2 (Definition of classical capacity). The classical capacity of a quantum channel $\mathcal{N}_{A\to B}$ is defined as

$$C = \sup\{R_c | R_c \,\text{achievable with}\, \mathcal{N}\}. \tag{3.10}$$

A characterization of the classical capacity in terms of the Holevo quantity was found in [Hol98; SW97]. First, any value assumed by the Holevo quantity $\chi(\{p_i, \mathcal{N}[\rho_i]\})$ is an achievable rate, therefore

Theorem 3.2.1 (Achievable classical rate: Holevo information of a channel). *The Holevo information of $\mathcal{N}$ defined as*

$$\chi(\mathcal{N}) := \sup_{\{p_i,\rho_i\}} \chi(\{p_i, \mathcal{N}[\rho_i]\}), \tag{3.11}$$

is an achievable rate. It follows that $\chi(\mathcal{N}^{\otimes})/n$ is also achievable.

Second, the family of achievable rates characterized in term of the Holevo quantity saturate the classical capacity:

Theorem 3.2.2 (Classical capacity). *The classical capacity of $\mathcal{N}$ is*

$$C(\mathcal{N}) = \lim_{n\to\infty} \frac{\chi(\mathcal{N}^{\otimes n})}{n}. \tag{3.12}$$

Unfortunately, this formula does not give a computable characterization of the capacity, since it requires to solve, in principle, an infinite sequence of optimization problems. It is computable when the Holevo information is additive, meaning $\chi(\mathcal{N}^{\otimes n}) = n\chi(\mathcal{N})$, which is not always guaranteed. It holds for entanglement breaking channels, reviewed in the next section [Sho02a], for unital qubit channels [Kin02], for depolarizing channels [Kin03], Hadamard channels [Kin06; Kin+05], phase-insensitive Gaussian channels [Gio+04; Gio+14]. The latter are a realistic noise model for electromagnetic waves in vacuum or optical fiber. In Chapter 9 we will consider a variation of this noise model, in which phase decoherence occurs. We defer the presentation of Gaussian channels to Chapter 4. By continuity results on channel capacities with respect to the diamond norm distance [LS09; Win16; Shi17], channels close to channels with additive Holevo information have approximately additive Holevo information [Led+18]. However, there exists channels for which $\chi(\mathcal{N} \otimes \mathcal{M}) > \chi(\mathcal{N}) + \chi(\mathcal{M})$ [HW08; Has09], a fact proved non-constructively.

Note that, for channels between infinite-dimensional Hilbert spaces, it is necessary to impose a further constraint on the encoding protocol to have finite results: typically, as motivated by the case of electromagnetic signals, one considers only protocols which have an average input state with bounded energy, or in general a positive definite observable. In optical communication, this constraint is motivated by a practical limit in the source power. In this case the formula for the Holevo information of the channel is modified restricting the supremum to ensemble satisfying the constraint, and the classical capacity of the channel is given by the regularized constrained Holevo information

The *superadditivity* of the Holevo information tells us that the characterization of the classical capacity is not entirely satisfactory. However, the situation is dramatically different when the sender and the receiver share unlimited entanglement, and the encoding of classical messages is done with a collection of CPTP maps acting on the sender's part of shared entangled states, with outputs in the inputs of n channel uses. In this case the optimal rate is completely characterized as the mutual information of the channel [AC97; Ben+99; Ben+02]. .

Theorem 3.2.3 (Entanglement assisted classical capacity.). *The classical entanglement assisted classical capacity of $\mathcal{N}$ is*

$$C_e(\mathcal{N}) = I(\mathcal{N}) := \sup_{\rho_{AA'}} I(B;A')_{\mathcal{N}_{A\to B}\otimes\mathcal{I}_{A'}[\rho_{AA'}]}. \tag{3.13}$$

This characterization holds since the mutual information gives achievable rates and it is additive. The state ρ_{AB} can be chosen to be pure. This characterization can be seen as a noisy version of superdense coding [BW92], as the achievability can be shown by random codes of unitary operations applied on the sender's share of the typical subspace of $\rho_{AB}^{\otimes n}$.

An equivalent expression for the mutual information of the channel $I(\mathcal{N})$ is $I(\mathcal{N}) = \sup_{\rho_A} I(\mathcal{N}, \rho_A)$, where $I(\mathcal{N}, \rho_A) := S(\rho_A) + S(\mathcal{N}_{A\to B}[\rho_A]) - S(\mathcal{N}^c_{A\to E'}[\rho_A])$. $I(\mathcal{N}, \rho_A)$ is concave in ρ_A [Wil17], which helps in solving the maximization.

3.2.2 Private communication

The BB84 protocol [Ben+14] shows how Alice and Bob can establish a secret key using the fact that different ensembles of pure states can have the same average state. This principle allows to devise a general strategy to communicate classical information through quantum channel, in such a way that it can be protected from any eavesdropper Eve that has access to the environment. From the point of view of Alice and Bob, a private classical communication protocol is executed in the same way as the classical protocol already explained. In addition, it is asked that any potential eavesdropper cannot decode Alice messages with probability of success larger than another parameter ϵ'. If Alice sends a message m, any eavesdropper will receive a post-processing of $(\omega_m)_{E'^n} := \left(\mathcal{N}^c_{A\to E'}\right)^{\otimes n}[(\rho_m)_{A^n}]$. The most restrictive condition is that there exists a state $\omega_{E'^n}$ such that

$$D_{\mathrm{Tr}}((\omega_m)_{E'^n}, \omega_{E'^n}) \leq \epsilon' \ \forall m \in M \tag{3.14}$$

A private code can be denoted as $\mathcal{C}_p(n, R_p, \epsilon, \epsilon')$ following the notation established for classical codes, and ϵ' is the such that the privacy condition Eq. 3.14 holds.

A rate for private communication R_p is called achievable with $\mathcal{N}$ if for any $\delta > 0$, $\epsilon > 0$, $\epsilon' > 0$ there exists a code $\mathcal{C}_p(n, R - \delta, \epsilon, \epsilon')$ for sufficiently large n. We can then give the following definition.

Definition 3.2.3 (Definition of classical capacity). The private classical capacity of a quantum channel $\mathcal{N}$ is defined as

$$P = \sup\{R_p | R_p \,\text{achievable with}\, \mathcal{N}\}. \tag{3.15}$$

The entropic functional which characterize the private capacity is the private information:

Theorem 3.2.4 (Achievable private rate: Private information of a channel). *The private information of a channel $\mathcal{N}$ is defined as*

$$I_p(\mathcal{N}) = \max_{\{p_i,\rho_i\}} \chi(\{p_i,\mathcal{N}[\rho_i]\}) - \chi(\{p_i,\mathcal{N}^c[\rho_i]\})\} \tag{3.16}$$

is an achievable rate for private communication. It follows that $I_p(\mathcal{N}^{\otimes})/n$ is also achievable.

The private information gives an achievable rate, obtained from random codewords generated from an ensemble that saturate the maximization. The idea behind the achievability proof is to have multiple codewords for the same message, and arrange the codewords in subsets corresponding to the same message. For each subset, the average state of the subset at the environment output should be a constant state. While all the codewords can be distinguishable for Bob with high probability, Eve sees on average the same state for each codeword, since a random choice of codewords makes the average states of each codewords close to the average state of the ensemble, at the output of the environment. The minimum necessary redundancy allows to communicate at the private information rate.

We have the following regularized expression for the private capacity [Dev05; CWY04]:

Theorem 3.2.5 (Private capacity). *The private capacity of a channel $\mathcal{N}$ is*

$$P(\Lambda) = \lim_{n\to\infty} \frac{I_p(\mathcal{N}^{\otimes n})}{n} \tag{3.17}$$

The private information is not additive, with explicit examples for Pauli channels (e.g. [SRS08]). Moreover, any alternative formula for the private capacity cannot be additive: a channel with zero private capacity together with a channel of classical capacity C can have a private capacity larger than C [Li+09]. A case in which the private information should be evaluated for an infinite number of uses is found in [ES15].

3.2.3 Quantum communication

Quantum communication is a genuinely quantum task, which has no classical analogue. In this case, Alice possesses a share of a quantum state, possibly entangled with Charlie. She wants to send Bob her share of the quantum state, such that the joint state held by Bob and Charlie is close to the original one, independently of the particular state. At her side, Alice acts on her share of $\rho_{A'C} \in \Sigma(\mathcal{H}_{A'} \otimes \mathcal{H}_C)$ with a channel $\mathcal{E}_{A'\to A^n}$. Bob receives $\mathcal{N}^{\otimes n}_{A^n\to B^n} \circ \mathcal{E}_{A'\to A^n}(\rho_{A'C})$ and performs a decoding applying a channel $\mathcal{D}_{B^n\to B'}$. The final state is then $\rho'_{B'C} = \mathcal{D}_{B^n\to B'} \circ \mathcal{N}^{\otimes n}_{A^n\to B^N} \circ \mathcal{E}_{A'\to A^n}(\rho_{A'C})$. The maps $\mathcal{E}_{A'\to A^n}$ and $\mathcal{D}_{B^n\to B'}$ define a quantum code $\mathcal{C}_q$. The rate $R_q(\mathcal{C}_q)$ is

$$R_q(\mathcal{C}_q) = \frac{\log_2 \dim \mathcal{H}_{A'}}{n}. \tag{3.18}$$

The constraint on the error probability is replaced in the quantum case by the following condition on the decoding error

$$D_{\mathrm{Tr}}(\mathcal{D}_{B^n \to B'} \circ \mathcal{N}^{\otimes n}_{A^n \to B^N} \circ \mathcal{E}_{A' \to A^n}(\rho_{A'C}),\ \ \rho_{A'C}) \leq \epsilon,\ \forall \rho_{A'C} \in \Sigma(\mathcal{H}_{A'} \otimes \mathcal{H}_C) \tag{3.19}$$

A quantum code can be denoted as $\mathcal{C}_q(n, R_q, \epsilon)$ if it uses n uses of the channel at rate R_q, with decoding error ϵ such that Eq. 3.19 holds.

Definition 3.2.4 (Definition of quantum capacity). The quantum capacity of a quantum channel $\mathcal{N}$ is defined as

$$Q = \sup\{R_q | R_q \,\text{achievable with}\, \mathcal{N}\}. \tag{3.20}$$

In this case, achievable quantum communication rates are given by the coherent information of the channel. We use the notation

$$I_c(\rho, \mathcal{N})_{\tau_{AB}} := I(A\rangle B)_\tau, \tag{3.21}$$

where $\tau_{AB} = N_{A' \to B} \otimes \mathcal{I}_A[\psi_{AA'}]$ and $\psi_{AA'}$ is a purification of ρ_A.

Theorem 3.2.6 (Achievable quantum rate: coherent information of a channel). *The coherent information of a channel $\mathcal{N}$ defined as:*

$$I_c(\mathcal{N}) = \sup_{\rho_A} I_c(\rho, \mathcal{N})_{\tau_{AB}}, \tag{3.22}$$

is an achievable rate for quantum communication. It follows that $I_c(\mathcal{N}^{\otimes})/n$ is also achievable.

Note that $I_c(\rho, \mathcal{N}) = I(A\rangle B)_\tau = S(\mathcal{N}[\rho]) - S(\mathcal{N}^{\mathrm{c}}[\rho])$. Both expressions of $I_c(\rho, \mathcal{N})$ are useful.

A way to construct a quantum code at the coherent information rate $I_c(\rho, \mathcal{N})$ is to map a basis of A' to a basis of a subspace of the typical subspace of the state $\rho_A^{\otimes n}$. This subspace has to have the property that its image under the channels decouples from the environment, making Bob able to decode correctly [Hay+08]. A detailed strategy involves making a coherent version of a private code constructed from and ensemble of eigenvectors of ρ, which achieves reliable private communication at the coherent information rate [Dev05; Wil17].

The quantum capacity is characterized as [Llo97; Sho02b; Dev05]:

Theorem 3.2.7 (Quantum capacity).

$$Q(\mathcal{N}) = \lim_{n\to\infty} \frac{I_c(\mathcal{N}^{\otimes n})}{n} \tag{3.23}$$

At variance with the non-additivity of the Holevo quantity, the non-additivity of the coherent information was immediately noticed. Indeed, the coherent information of a channel can be zero even if the channel is not a constant channel. By showing that for the coherent information of more uses of the channel is non zero while the one-use coherent information is zero, superadditivity was discovered [SS96; DSS98]. Several subsequent works have explored this phenomenon for a variety of channels [SS07; FW08; SSY11; Cub+15; LLS18a; BL19; SG21; Sid20b; Sid20a; NPJ20; Yu+20], with evidence that an unbounded number of uses of the channel may be necessary to obtain a non zero coherent-information [Cub+15], for a channel with non-zero quantum capacity. As for the private capacity, the quantum capacity cannot have an additive formula in terms of an information quantity which is additive for a tensor product of generic channels, because of the striking *superactivation* phenomenon: two channels with zero quantum capacity can have non-zero quantum capacity if used together [SY08; SSY11].

It can be shown that the private information of a channel is always larger than the coherent information, implying also $P(\mathcal{N}) \geq Q(\mathcal{N})$.

We also mention the *entanglement assisted quantum capacity* $Q_E(\mathcal{N})$, which give the optimal rate for quantum teleportation in presence of noise. Thanks to a duality between superdense coding and quantum teleportation, this capacity is equal to half the entanglement assisted classical capacity [DHW04].

Moreover, while classical communication from the sender to the receiver does not increase the quantum capacity [Ben+96; BKN00], classical feedback does [Ben+96]. The two-way quantum capacity $Q^{\leftrightarrow}(\mathcal{N})$ [Ben+96], defined as the quantum capacity assisted by local operations and classical communication, is a very relevant figure of merit for transmission of quantum information, since it is usually conceivable that Alice and Bob have a good and cheap classical channel to communicate. However, we also stress that on quantum memories this adaptive protocol cannot be applied, since the noise acts at the same time on all the physical systems. In this case, $Q(\mathcal{N})$ sets the ultimate limits for information preservation. In the same way of $Q^{\leftrightarrow}(\mathcal{N})$, one can define the two-way private capacity $P^{\leftrightarrow}(\mathcal{N})$. $Q^{\leftrightarrow}(\mathcal{N})$ and $P^{\leftrightarrow}(\mathcal{N})$ lack a characterization in terms of an entropic functional, but upper and lower bounds are available, which we will mention in the next section since they are obviously bounds for the capacities $Q(\mathcal{N})$ and $P(\mathcal{N})$ as well.

3.3 Bounds on private and quantum capacities

In Chapter 8 we will present a method to obtain upper bounds on the quantum and private capacity of channels which has a wide applicability, and at the time of the writing gives the best bounds available for several channels of practical and fundamental relevance. In this section, we review the current knowledge on bounds on the quantum and private capacity.

Any value of the coherent information for some input state is a lower bound. As we already mentioned, interesting results can be found on lower bounds on the zero quantum capacity threshold, with informed choices of states which are good for the high noise regime. For Pauli channels (see Chapter 4 for a presentation) there are several works showing superadditivity from zero capacity thresholds [SS96; DSS98; SS07; FW08], the most recent and comprehensive being [BL19]. We do not know of any superadditivity evidence for gaussian channels, even in the energy-constrained setting. However, for the single-mode thermal attenuator at constrained energy, superadditivity for input states restricted to Gaussian states has been shown in [NPJ20]. Among qubit channels, the dephrasure channel, which is a concatenation of a dephasing and an erasure channel, exhibits clear superadditivity already at the level of the two-letter coherent information [LLS18a]. Other examples of manifest superadditivity can be found in [SG21; Sid20b; Sid20a].

In the following, we will concentrate on upper bounds, which are also the interest of Chapter 8 of this thesis. First of all, any capacity is monotonic with respect to composition, i.e. $\tilde{C}(\Phi_1 \circ \mathcal{N} \circ \Phi_2) \leq \tilde{C}(\mathcal{N})$ for any triple of channels $\Phi_1, \mathcal{N}, \Phi_2$ and any capacity $\tilde{C}$. Therefore, if the capacity of $\mathcal{N}$ can be computed, this results in a computable upper bound on the capacity of $\Phi_1 \circ \mathcal{N} \circ \Phi_2$. We call $\mathcal{N}$ an extension of $\Phi_1 \circ \mathcal{N} \circ \Phi_2$.

Fortunately, there are classes of channels for which the quantum and private capacity can be computed, since their coherent information is additive.

Degradable and antidegradable channels are defined as follows.

Definition 3.3.1 (Degradable channel). A channel $\mathcal{N}_{A\to B}$ with a complementary channel $\mathcal{N}^c_{A\to E'}$ is degradable if there exists a channel $\mathcal{W}_{B\to E'}$ such that $\mathcal{N}^c_{A\to E'} = \mathcal{W}_{B\to E'} \circ \mathcal{N}_{A\to B}$.

Definition 3.3.2 (Antidegradable channel). A channel $\mathcal{N}_{A\to B}$ with a complementary channel $\mathcal{N}^c_{A\to E'}$ is antidegradable if $\mathcal{N}^c_{A\to E'}$ is degradable.

A subset of antidegradable channels are *entanglement breaking* channels:

Definition 3.3.3 (Entanglement breaking channels). A channel $\mathcal{N}_{A\to B}$ is entanglement breaking if $\mathcal{N}_{A\to B} \otimes \mathcal{I}_C[\rho_{AC}]$ is separable for any ρ_{AC}.

For degradable channels the quantum and private capacity can be exactly computed, and they actually coincide.

Theorem 3.3.1 (Quantum capacity of degradable channels). *For a degradable channel $\mathcal{N}$, $I_p(\mathcal{N}^{\otimes n}) = I_c(\mathcal{N}^{\otimes n}) = I_p(\mathcal{N}) = I_c(\mathcal{N})$, therefore*

$$Q(\mathcal{N}) = I_c(\mathcal{N}) = P(\mathcal{N}). \tag{3.24}$$

Additivity of coherent information for degradable channel was shown in [DS05], while [Smi08] showed the additivity of the private information, and that the private information coincides with coherent information. It is known that the quantum and private capacities do not coincide in general, since there exists channel with positive private information and zero quantum capacity [HHH98].

For antidegradable channels the quantum and private capacity is zero, since $I_c(\mathcal{N}) = 0$ by a no-cloning argument [BDS97; Gio+03; GF05], (see also [Hol08; CRS08]).

Theorem 3.3.2 (Quantum and private capacity of antidegradable channels). *For an antidegradable channel $\mathcal{N}$, $I_c(\mathcal{N}^{\otimes n}) = I_c(\mathcal{N}) = 0$, therefore*

$$Q(\mathcal{N}) = 0. \tag{3.25}$$

Since an antidegradable channel can always be extended to a channel which is both degradable and antidegradable, $P(\mathcal{N}) = 0$.

A degradable extension for antidegradable channels, with zero coherent information, has been found in [SS08]. Examples of channels that are either degradable or antidegradable are dephasing channels [DS05], amplitude damping channels [GF05], quantum limited attenuators and amplifiers [WPGG07], and all qubit channels with Kraus rank two [WPG07].

An important fact is the concavity of the coherent information of degradable channels [YHD08], which makes the optimization easier:

Theorem 3.3.3 (Concavity of coherent information for degradable channels). *If a channel $\mathcal{N}$ is degradable, for any ensemble $\{p_i, \rho_i\}$ we have*

$$I_c(\mathcal{N}, \sum_{i=1}^{n} p_i \rho_i) \geq \sum_{i=1}^{n} p_i I_c(\mathcal{N}, \rho_i). \tag{3.26}$$

Since the quantum and private capacities are easily computed for degradable channels, an established way to find upper bounds on the quantum and private capacity is to finding degradable extensions and computing their coherent information.

For example, the thermal amplifier and attenuator and the additive noise channel are not degradable, but they can be realized as a composition of a quantum limited attenuator and a quantum limited amplifier, which are instead either degradable or antidegradable. In this way, an area of zero capacity which strictly includes entanglement-breaking channels [Hol08] has been found in [CGH06], and upper bounds have been found in [WQ16; Sha+18; RMG18; NAJ19]. Similar techniques have been applied to the generalized amplitude damping channel [KSW20].

The notion of weak-degradability is introduced in [CG06] and used in [CGH06] to classify single-mode Gaussian channels.

Definition 3.3.4 (Weak degradability). A channel with *physical representation*

$$\mathcal{N}_{A\to B}[X] = \mathrm{tr}_{E'}[U_{AE\to BE'} X_A \otimes \rho_E U^\dagger_{AE\to BE'}], \tag{3.27}$$

where ρ_E is a generic mixed state, is called weakly degradable if there exists a channel $\mathcal{W}_{B\to E'}$

$$\mathrm{tr}_B[U_{AE\to BE'} X_A \otimes \rho_E U^\dagger_{AE\to BE'}] = \mathcal{W}_{B\to E'} \circ \mathcal{N}_{A\to B}[X]. \tag{3.28}$$

While degradable channels are weakly degradable, the converse is not true since ρ_E is mixed. However, for a purification $|\tau\rangle\langle\tau|_{EB'}$ of ρ_E, the channel $\tilde{\mathcal{N}}_{A\to BB'} = \mathrm{tr}_{E'}[(U_{AE\to BE'} \otimes I_{B'}) X_A \otimes |\tau\rangle\langle\tau|_{EB'} (U_{AE\to BE'} \otimes I_{B'})^\dagger]$ is a degradable extension of $\mathcal{N}$. Therefore $Q(\mathcal{N}) \leq Q(\tilde{\mathcal{N}})$. This fact was also used in [RMG18] to obtain a bound on the quantum capacity of the thermal attenuator.

Another important type of degradable extensions are flagged extensions, which we will study in detail in Chapter 8. We restate the definition of flagged extension we mentioned in the introduction.

Definition 3.3.5 (Flagged extension of a convex combination of channels). For a channel $\mathcal{N} = \sum_j p_j \mathcal{N}_j$ with probability distribution $\{p_j\}$, $\{\mathcal{N}_j\}$ channels, and a collection of states $\{\sigma_j\}$, a flagged extension is

$$\widehat{\mathcal{N}} = \sum_j p_j \mathcal{N}_j \otimes \sigma_j. \tag{3.29}$$

The case with orthogonal flags $\{\sigma_j\}$ was introduced in [SS08], noting that flagged extensions with orthogonal flags of convex combination of degradable channels are degradable, obtaining upper bounds on the quantum capacity of Pauli channels. This result came after it was observed that the quantum capacity is convex for convex combination of degradable channels [WPG07; SSW08], and after similar bounds could be obtained exploiting the notion of communication with symmetric side channel assistance [SSW08]. This idea was used in [Ouy14] and developed in [LDS18], where optimization of upper

bounds from flagged convex combinations of degradable and antidegradable channels were considered.

The concept of approximate degradability was introduced in [Sut+17], showing that non-additivity effect can be bounded if the channel, composed with a candidate degrading map, is close to the complementary channel in diamond norm distance. The minimum diamond norm distance obtained in this way is the degradability parameter, and it can can be computed by a semidefinite program. Another notion of approximate degradability is given by the minimum distance from degradable channels. In both cases one can obtain an upper bound on the quantum capacity in terms of the single-letter coherent information and the degradability parameter of the channel. [LLS18b] analytically estimates the degradability parameter for low noise channels in terms of the diamond norm distance to the identity channel, using the complementary channel as candidate degrading map.

Approximate degradability was also used in [Sha+18], to bound quantum and private capacities of thermal attenuator and amplifier, with an energy constraint. (see also [WQ16] for energy constrained quantum and private capacity of infinite dimensional systems).

Interestingly, while these results based on degradability give the best bounds available for important finite dimensional channels, bounds valid for the two-way quantum capacity (quantum communication assisted by unlimited forward-backward classical communication) [Pir+17; WTB17] of thermal attenuator and amplifier and additive Gaussian noise, proved to be state-of-the-art in low noise regimes. To our knowledge, this is the only case where upper bounds for two-way capacities are competitive with upper bounds given by (approximate)-degradability.

For completeness, we mention several bounds which are more general and in some cases have also the advantage of providing strong converses for the optimal rate, i.e. the error of a code with rate higher than the upper bound goes to one exponentially fast in the number of channel uses. The Rains information, inspired by the Rains bounds in entanglement theory [Rai01], is a strong converse for the unassisted quantum capacity [TWW17], and more easily computable upper bounds on the Rains information are available [WFD19; ZP20a], which also give strong converses for the two-way quantum capacity [BW18; ZP20a]. Squashed entanglement is an upper bound for the two-way private capacity [TGW14]. Entanglement cost [Ber+13] is a strong converse for two-way private capacity [CMH17]. Relative entropy of entanglement [Pir+17] is a strong converse for unassisted private capacity and two-way private capacity for teleportation covariant channels [Pir+17; WTB17]. Max-relative entropy of entanglement of a channel is a strong converse for two-way private capacity [CMH17]. Other upper bounds to the

quantum capacity are [MHRW16; GJL18].

Chapter 4

Symmetries and quantum information processing

A problem that seems complicated at first glance may be greatly simplified if symmetries are taken into account. Many relevant problems in quantum information theory can be attacked using group representation theory [Hay17b], and this thesis makes extended use of this tool. In Sec. 4.1 we recall elements of representation theory of finite and compact groups. In Sec. 4.2 we present important applications of these results in quantum statistical inference, inspired by the selection of [Hay17a]. These applications will be used in Chapters 5, 6, 7. In Sec. 4.3 and 4.4, we introduce Pauli channels and Gaussian states and channels putting an emphasis of their group theoretic structure, which will be crucial in Chapters 8, 9.

4.1 Group theory and representation theory

The intention of this section is to offer a compact presentation of these classic results that can be used as a quick reference for the reader of this thesis. The level of detail of the presentation is a little more than what actually is needed in the following chapters, with the goal to present the group theoretic objects we use in a clearer way. Sometimes simple arguments are used to justify properties needed for our computations, which are actually corollaries of deeper theorems (e.g. orthogonality of matrix elements of irreducible representation come from the Peter-Weyl theorem [Kna86]). The presentation is inspired by [Hay17b], which introduces group representation theory in the context of quantum physics. Most proofs of the theorems can be found in [Hay17b], while for the more technical results references are [Kna86; Sag01; GW09; Hal15].

4.1.1 Groups and representations

A *group* G is a set together with a product $G \times G \to G$, denoted as $(g_1, g_2) \to g_1 g_2$, with the following properties:

- associativity: $g_1(g_2 g_3) = (g_1 g_2) g_3$
- unique neutral element e such that $ge = g$ for each $g \in G$,
- unique inverse g^{-1} for each $g \in G$ such that $gg^{-1} = e$ (and therefore $g^{-1}g = e$).

The order of the a group is the cardinality of the set $|G|$. A *subgroup* H is a subset of G closed under product and taking inverses. The conjugacy class of the element a is $\{a\} := \{gag^{-1} | g \in G\}$, and such classes establish an equivalence relation on G. Given two groups H and K with cartesian product $H \times K$ one can define the direct product group $H \times K$, with $(h, k)(h', k') = (hh', kk')$.

Groups have a rich structure, and it is fruitful to consider maps between groups that preserve this structure. A map f between two groups G_1, G_2 is a homomorphism if $f(g_1)f(g_2) = f(g_1 g_2)$, and f is furthermore an isomorphism if it is bijective. We write $G_1 \cong G_2$ if an isomorphism exists, and if $G_1 = G_2$ the isomorphism is also called automorphism.

In addition, groups can act on a set in a way that mirrors the group structure. This is physically meaningful, as we can see a group as the set of transformations of a system. The action of a group on a set Θ is a function $T : G \times \Theta \to \Theta$ denoted as $(g, \theta) \to g\theta$ satisfying $(g_1 g_2)\theta = g_1(g_2\theta)$ for all $g_1, g_2 \in G$ and $\theta \in \Theta$, and $e\theta = \theta$ for all $\theta \in \Theta$. A group acts transitively on a non-empty set Θ if for every $\theta_1, \theta_2 \in \Theta$ there is a $g \in G$ such that $g\theta_1 = \theta_2$, and in this case Θ is called a homogenous space. Given an element θ_0, the stabilizer of θ_0 is the subgroup H s.t. $\{h \in G | T(h, \theta_0) = \theta_0\}$. For an element $g \in G$, $[g] := gH$ is a residue class, the set of residue classes is the quotient space G/H, and it encodes the set of non-trivial transformations of θ_0. The action of a group on itself given by the product $f : G \times G \to G$, $f(a, g) := aga^{-1}$ is an automorphism.

We will also mention some important facts that can be obtained by considering the elements of the group as a basis of a vector space, with an additional structure given by the group product. A real (complex) algebra is a linear real (complex) space $\mathfrak{a}$ together with a product operation $\mathfrak{a} \times \mathfrak{a}$ denoted as $(v, w) \to v \cdot w$, which is bilinear, i.e. $\sum_i a_i v_i \cdot \sum_j b_j w_j = \sum_{i,j} a_i b_j v_i \cdot w_j$. An homomorphism of algebras $\mathfrak{a}$ and $\mathfrak{b}$ is a linear function $f : \mathfrak{a} \to \mathfrak{b}$ such that $f(v) \cdot f(w) = f(v \cdot w)$. For a finite group, a group algebra $\mathbb{C}[G]$ can be obtained on a complex vector space with a basis indexed by the elements of the group and the product is given by the product of G.

Some examples of well known groups are:

- permutation group S_n;
- cyclic groups: $\mathrm{U}_k = \{e^{i2\pi j/k} | j \in \mathbb{Z}\}$, $k \in \mathbb{N}$;
- $\mathrm{U}(1) = \{z \in \mathbb{C} || z| = 1\}$;
- general linear group $\mathrm{GL}(\mathcal{H})$ of invertible linear maps on $\mathcal{H}$, subgroup $\mathrm{U}(\mathcal{H})$ of unitary operators;
- subgroups of $\mathrm{GL}(\mathbb{C}^n)$: $\mathrm{SL}(\mathbb{C}^n)$ (matrices with determinant one), $\mathrm{U}(\mathbb{C}^n)$ (unitary matrices), $\mathrm{SU}(\mathbb{C}^n)$ (unitary matrices with determinant one).

Representations give groups of transformations realizable on a Hilbert space $\mathcal{H}$ based on a group G. We now state some basic facts about representations. A representation is a homomorphism $\mathbf{f}$ between G and $\mathrm{GL}(\mathcal{H})$, a unitary representation has $\mathbf{f}(G) \subseteq \mathrm{U}(\mathcal{H})$. Two representations $\mathbf{f}_1$ and $\mathbf{f}_1$ are isomorphic if there exists an invertible linear operator A such that $A\mathbf{f}_1 A^{-1} = \mathbf{f}_2$, and unitarily isomorphic if A unitary. For any two representations $\mathbf{f}_1$ and $\mathbf{f}_2$, we denote by $\mathbf{f}_1 \oplus \mathbf{f}_2$ the direct sum representation, which acts as $\mathbf{f}_1 \oplus \mathbf{f}_2(g) = \mathbf{f}_1(g) \oplus \mathbf{f}_2(g)$. Conversely, a representation is decomposable if it is a direct sum of two representations. A fundamental question is to understand how to decompose a representation as direct sum of representations.

Definition 4.1.1 (Irreducible representation). A representation $\mathbf{f}$ is called reducible if there exists a non-trivial invariant subspace $\mathcal{K} \subseteq \mathcal{H}$, $\mathcal{K} \neq \{0\}$ or $\mathcal{H}$, such that $\mathbf{f}(g)u \in \mathcal{K}$ for each $u \in \mathcal{K}$ and for each $g \in G$. If a non-trivial invariant subspace does not exists, the representation is irreducible.

A representation that can be written as a direct sum of irreducible representation is completely reducible. Unitary representations have the property that if they have an invariant subspace, its orthogonal subspace is also invariant. It follows that finite-dimensional unitary representations are completely reducible. Moreover, representations of finite groups are either irreducible or decomposable, since they are unitary with respect to the following product: $(v, w)_G := \frac{1}{|G|} \sum_{g \in G} \langle v | \mathbf{f}(g)^\dagger \mathbf{f}(g) | w \rangle$. Therefore, finite dimensional representations are completely reducible. For the purposes of this thesis, the knowledge of the reduction of unitary representations acting on Hilbert spaces of quantum systems makes us able to solve quantum information processing tasks. We denote as Λ_G the set of indexes describing the irreducible unitary representations of G, up to unitary isomorphism. For $\lambda \in \Lambda_G$, we denote the corresponding representation space as $\mathcal{U}_\lambda(G)$ of dimension d_λ, and the corresponding representation by $\mathbf{f}_\lambda$. The reduction into irreducible representation of $\mathbf{f}$, where each equivalent irreducible representation appears with multiplicities n_λ, has the following decomposition

$$\mathcal{H} \cong \bigoplus_{\lambda \in \Lambda_G} \mathcal{U}_\lambda(G) \otimes \mathbb{C}^{n_\lambda} \tag{4.1}$$

and we write the decomposition of $\mathbf{f}$ as $\mathbf{f} = \oplus_{\lambda\in\Lambda_G} n_\lambda \mathbf{f}_\lambda$. $\{|j\rangle\}_{j=1}^{n_\lambda}$ is a complete orthonormal set of vectors (CONS) for $\{\mathbb{C}^{n_\lambda}\}$, while a CONS of $\mathcal{U}_\lambda(G)$ is given by $\{|\lambda; j\rangle\}_{j=1}^{d_\lambda}$. These vectors $|\lambda; j\rangle \otimes |j'\rangle = |\lambda; j; j'\rangle$ form a CONS of $\mathcal{H}$. We denote matrix elements of representation matrices as $R_{\lambda,i,j}(g) = \langle \lambda, i| \mathbf{f}_\lambda(g) |\lambda, j\rangle$.

4.1.2 Schur's lemma and orthogonality relations

The following simple but deep result is the cornerstone of many applications:

Theorem 4.1.1 (Schur's lemma). *Let $\mathbf{f}_1$ and $\mathbf{f}_2$ be two irreducible representations on complex vector spaces V and W. If A is a linear map from V to W such that $A\mathbf{f}_1(g) = \mathbf{f}_2(g)A$, $\forall g \in G$, then $A = 0$ or A is an isomorphism. If $\mathbf{f}_1 = \mathbf{f}_2$, A is a multiple of the identity.*

As a corollary, an operator $A \in GL(\mathcal{H})$ satisfying $A\mathbf{f}(g) = \mathbf{f}(g)A$ for a representation with decomposition $\mathbf{f} = \oplus_{\lambda\in\Lambda_G} n_\lambda \mathbf{f}_\lambda$, will have the form

$$A = \bigoplus_{\lambda\in\Lambda_G} I_{\mathcal{U}_\lambda(G)} \otimes A_\lambda \tag{4.2}$$

according to the decomposition of Eq. (4.1), and A_λ acting on the multiplicity space of the irreducible representation λ.

Another important consequence of Schur's lemma are the following orthogonality relations for finite groups:

$$\sum_{g\in G} \frac{1}{|G|} \overline{R_{\lambda,i,j}}(g) R_{\lambda',i',j'}(g) = \frac{\delta_{\lambda\lambda'}\delta_{ii'}\delta_{jj'}}{d_\lambda} \tag{4.3}$$

A similar statement holds if there exists a left and right invariant measure $\mu(dg)$ on the group, i.e. $\int_G \mu(dg) f(g) = \int_{g\in G} \mu(dg) f(ag) = \int_G \mu(dg) f(ga)$ for any $a \in G$, $f : G \to \mathbb{C}$ and $\int_G \mu(dg) = 1$. For finite groups, we used the measure $\mu(B) = \frac{|B|}{|G|}$, which satisfies this property. For compact groups, a left and right invariant measure exists and it is the unique, and it is called *Haar measure* [Kna86; Hal15]. In fact, given a left and right invariant measure μ one can define the space $L^2(G) := \{f : G \to \mathbb{C} | \int_G \mu(dg) \overline{f(g)} f(g) < \infty\}$ and matrix elements of irreducible unitary representations are dense in $L^2(G)$, a result known as Peter-Weyl theorem [Kna86].

The function $\chi(g) := \mathrm{Tr}[\mathbf{f}(g)]$ is called the *character* of the representation $\mathbf{f}$ and it is invariant under conjugation of g, i.e. $\chi(g) = \mathrm{Tr}\big[\mathbf{f}(aga^{-1})\big]$ for all $a, g \in G$, by ciclicity of the trace and since $\mathbf{f}$ is an homomorphism. The space of functions invariant under

conjugation is a vector space called the space of class functions. The character of a representation can be calculated as $\chi(g) = \sum_{\lambda\in\Lambda_G} n_\lambda \chi_\lambda(g)$, where $\chi_\lambda(g)$ is the character of the irreducible representation λ and n_λ its multiplicity. From the orthogonality relations for matrix elements, Eq. (4.3), it follows that characters are also an orthonormal set

$$\sum_{g\in G} \frac{1}{|G|} \chi_{\lambda'}(g)\overline{\chi_\lambda(g)} = \delta_{\lambda,\lambda'} \tag{4.4}$$

Therefore the multiplicities of a representation with character χ can be calculated as $m_\lambda = \sum_{g\in G} \frac{1}{|G|}\chi(g)\overline{\chi_\lambda(g)}$. With a little more work (see e.g. [Hay17b]) one can show that characters are actually a basis of the linear space of class function. Since the latter has dimension equal to the number of conjugacy classes of G for finite G, the irreducible representations of a finite group are in bijection with the conjugacy classes of G.

We complete the list of important consequences of Schur's lemma by stating the following facts

- Unitary irreducible representations of finite groups are finite dimensional.
- Unitary irreducible representations of compact groups are finite dimensional.

The first fact can be checked by picking $v \in \mathcal{U}_\lambda$ and considering the operator $\frac{1}{|G|}\sum_{g\in G} \mathbf{f}(g)^\dagger_\lambda |v\rangle\langle v| \mathbf{f}(g)_\lambda$. This operator has trace 1, and it is a multiple of the identity operator on $\mathcal{U}_\lambda$, by Schur's lemma. Therefore the dimension of $\mathcal{U}_\lambda$ is finite. The same fact for compact groups can be shown by using the properties of the Haar measure.

We will often be interested in tensor product representations, defined as $\mathbf{f}_1 \otimes \mathbf{f}_1(g) := \mathbf{f}_1(g) \otimes \mathbf{f}_2(g)$ for two representations $\mathbf{f}_1$ and $\mathbf{f}_2$. We will denote the decomposition of the tensor product space of the product of two irreducible representation as

$$\mathcal{U}_\lambda(G) \otimes \mathcal{U}_{\lambda'}(G) \cong \bigoplus_{\lambda''\in\Lambda_G} \mathcal{U}_{\lambda''}(G) \otimes \mathbb{C}^{C^{\lambda''}_{\lambda,\lambda'}(G)}, \tag{4.5}$$

where $C^{\lambda''}_{\lambda,\lambda'}$ are the multiplicities of the irreducible representation with label λ'' in the tensor product representation of labels λ, λ'. The isomorphism can be expressed as a change of basis, which is called the Clebsch-Gordan transform.

4.1.3 Representations of the symmetric group

The group of permutations of n objects is also known as the symmetric group on the set $[n]$:

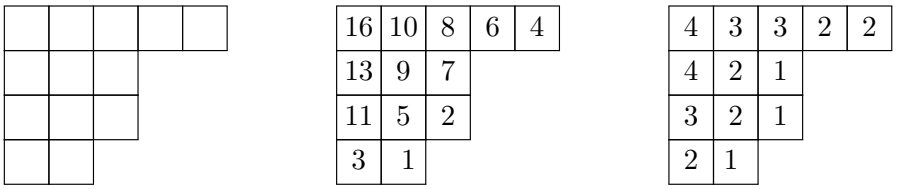
$$S_n := \{\sigma : [n] \to [n] : \sigma \text{ is bijective}\} \tag{4.6}$$

We can denote an element $\sigma \in S_n$ as a vector with components $\sigma_i = \sigma(i)$. The cardinality of this group is $|S_n| = n!$, and the conjugacy classes are given by the cycles structure of the permutations. A cycle of length $k \geq 2$ is a permutation such that there exists an element $i \in [n]$ such that $\sigma(i), \sigma^2(i), ..., \sigma^k(i)$ are all different and they coincide with all the elements on $[n]$ for which $\sigma(i) \neq i$. Two cycles are disjoint if they act non-trivially on disjoint subsets of $[n]$, and it follows that disjoint cycles commute. A cycle of length 2 is called a transposition, and exchanges exactly two elements. Transpositions generate S_n, and the sign of a permutation $\text{sgn}(\sigma)$ is defined as -1 elevated to the number of transpositions whose product is σ. Any permutation can be written in a unique way as a product of cycles up to their order. The list of integers given by the length of the cycles of σ in decreasing order, followed by a number of 1 for as many elements such that $\sigma(i) = i$, gives a partition of n. The cycle structure does not change under conjugation: for a permutation $\sigma = \sigma^{(1)}\sigma^{(2)}...\sigma^{(l)}$ where $\sigma^{(j)}$ are disjoint cycles, we have that for any $\tau \in S_n$, $\tau\sigma\tau^{-1} = \tau\sigma^{(1)}\tau^{-1}\tau\sigma^{(2)}\tau^{-1}...\tau\sigma^{(l)}\tau^{-1}$ and $\tau\sigma^{(j)}\tau^{-1}$ are disjoint cycles since τ is bijective. These observation tells us that the conjugacy classes of S_n are indexed by partitions of n, therefore:

Proposition 4.1.1 (Irreducible representation of S_n). *The irreducible representations of S_n are labelled by partitions of n, i.e. is vectors λ of n components such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$ and $\sum_{i=1}^n \lambda_i = n$.*

A pictorial way to indicate a partition is the Young diagram Y_λ, as seen in Fig. 4.1, where rows of λ_i boxes are arranged in non-increasing order from top to bottom. The length of the diagram $l(Y)$ is the number of rows, the size of the diagram $|Y|$ is the number of boxes. The length and the size of a partition are defined acccordingly. A Young tableau is obtained by filling a Young diagram of size n with distinct elements of $[n]$. Standard Young tableaus are Young tableaus obtained by filling the boxes with the rule that the integers are strictly decreasing in each row from left to right and strictly decreasing in each column from top to bottom. Semistandard Young tableau are Young tableaus obtained by filling the boxes with the rule that the integers are weakly decreasing in each row from left to right and strictly decreasing in each column from top to bottom.

The irreducible representations can be constructed from the algebra $\mathbb{C}[S_n]$ and Young symmetrizer [Sag01], as follows: for a Young tableau T, the horizontal permutations $H(T)$ are defined as the permutation that leave the elements of the rows invariant, and vertical permutations $V(T)$ are those that leave columns invariant. The horizontal symmetrizer is $c_T^H := \sum_{\sigma \in H(T)} \sigma \in \mathbb{C}[S_n]$, the vertical symmetrizer $c_T^V := \sum_{\sigma \in V(T)} \text{sgn}(\sigma)\sigma \in \mathbb{C}[S_n]$, and the symmetrizer is $c_T = c_T^H c_T^V$. The subspace $\mathbb{C}[S_n]c_T := \{xc_T | x \in \mathbb{C}[S_n]\}$



a)

16	10	8	6	4
13	9	7		
11	5	2		
3	1			

b)

4	3	3	2	2
4	2	1		
3	2	1		
2	1			

c)

Figure 4.1: Examples of a) Young diagram, b) Standard Young tableau, c) Semistandard Young tableau of length 4 and size 13.

is an irreducible representation space for the action of the group S_n by left multiplication. The irreducible representation is uniquely identified by the partition λ (it does not depend on the choice of the tableau), and its Hilbert space is indicated as $\mathcal{V}_\lambda(S_n)$. All the irreducible representation of S_n are obtainable in this way, and they are also known as Specht modules. The dimension of $\mathcal{V}_\lambda(S_n)$ is equal to the number of standard Young tableau of shape λ. In the following sections, we will denote the character of the irreducible representation λ of S_n as $\chi_\lambda(\mu)$, where μ is a partition. The dimension of the irreducible representation λ is $\omega_\lambda := \chi_\lambda((1,1,...,1))$.

4.1.4 Representations of $\mathrm{SU}(d)$

$\mathrm{SU}(d)$ is a compact group of great importance in quantum theory, as it describes the set of unitary channels for a Hilbert space of dimension d. The irreducible representations of $\mathrm{SU}(d)$ can be constructed from the representations of its Lie algebra $\mathfrak{su}(d)$, therefore we take the opportunity to introduce Lie algebras.

A real (complex) linear space V is called a real (complex) Lie algebra when there is a map $[\cdot,\cdot] : V \times V \to V$, called a commutator, that satisfies the following conditions:

- bilinearity: $[aX_1 + bX_2, Y] = a[X_1, Y] + b[X_2, Y]$,
- skew-symmetry: $[X, Y] = -[Y, X]$,
- Jacobi law: $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$.

Basic examples: the set $\mathfrak{gl}(V)$ of linear maps on V with the matrix commutator, the set $\mathfrak{u}(\mathcal{H})$ of skew-Hermitian matrices on $\mathcal{H}$. Consider now a subgroup G of $\mathrm{GL}(d)$, parametrized by l real variables $\theta_1, ..., \theta_l$, which is topologically closed. The Lie algebra $\mathfrak{g}$ of G is defined as

$$\mathfrak{g}(d) := \{X : e^{tX} \in G, \forall t \in \mathbb{R}\}, \tag{4.7}$$

where the matrix exponential is defined as $e^X := \sum_{n=1}^{\infty} \frac{X^n}{n!}$.

From the Lie product formula

$$\lim_{m\to\infty}\left(e^{tY/m}e^{tX/m}\right)^m = e^{t(X+Y)}, \tag{4.8}$$

since G is closed, it follows that $aX + bY \in \mathfrak{g}(d)$ for any $X, Y \in \mathfrak{g}(d)$ and $a, b \in \mathbb{R}$. Moreover, $e^{tX}Ye^{-tX} \in \mathfrak{g}$ since $e^{e^{tX}Ye^{-tX}} = e^{tX}e^{Y}e^{-tX} \in G$. Then we have

$$\lim_{t\to 0}\frac{e^{tX}Ye^{-tX} - Y}{t} = [X, Y], \tag{4.9}$$

therefore $[X, Y] \in \mathfrak{g}$ since $\mathfrak{g}$ is closed. This makes $\mathfrak{g}(d)$ a real Lie algebra with the commutator $[\cdot, \cdot]$.

A representation of an algebra is an homomorphism from the algebra to $\mathfrak{gl}(\mathcal{H})$. A representation $\mathbf{f}$ of a continuous group induces a representation of its Lie algebra, defined as

$$\mathbf{f}(X) := \lim_{t\to 0}\frac{\mathbf{f}(e^{tX}) - I}{t}. \tag{4.10}$$

Since $U \in \mathrm{U}(d)$ can be written as $U = e^A$ for a some anti-hermitian matrix $A = -A^\dagger$, for any finite dimensional unitary representation of the group G one has a an anti-Hermitian representation of the Lie algebra. A representation of an algebra is irreducible if it has no non-trivial invariant subspace, that is a subspace $\mathcal{K} \subseteq \mathcal{H}$ such that $\mathbf{f}(X)u \in \mathcal{K}$, $\forall\, u \in \mathcal{K}$, reducible otherwise. If $\mathbf{f}$ is irreducible as a representation of G, it is also irreducible as a representation of $\mathfrak{g}$.

Any matrix $U \in \mathrm{SU}(d)$ can be written as $U = e^A$ for a some anti-hermitian traceless matrix $A = -A^\dagger$, and indeed $\det e^A = e^{\mathrm{tr}\, A}$. Therefore, the real vector space of anti-hermitian traceless matrices coincides with $\mathfrak{su}(d)$. A basis of this vector space is given by the following matrices

$$F^x_{j.l} := \frac{i}{2}(|l\rangle\langle j| + |j\rangle\langle l|) \qquad 1 \le j < l \le d; \tag{4.11}$$

$$F^y_{j,l} := \frac{1}{2}(|j\rangle\langle l| - |l\rangle\langle j|) \qquad 1 \le j < l \le d; \tag{4.12}$$

$$F^z_j := i(|j\rangle\langle j| - |j+1\rangle\langle j+1|) \qquad j = 1, ..., r-1, \tag{4.13}$$

and a subspace is given by $\mathfrak{sd}(d)$, spanned by $\{F^z_j\}$. For an irreducible representation $\mathbf{f}$ of $\mathfrak{su}(d)$, we define

$$\mathbf{E}_j := i\mathbf{f}(F_j^z) \qquad j = 1, ..., d-1; \tag{4.14}$$

$$\mathbf{K}_{j,l\pm} := -i\mathbf{f}(F_{j,l}^x) \mp \mathbf{f}(F_{j,l}^y) \qquad 1 \leq j < l \leq d. \tag{4.15}$$

Defining the vectors $\boldsymbol{\alpha}_{j,l}$ with components $\alpha_{j,l,j'} = \delta_{j',j} - \delta_{j',l} - \delta_{j'+1,j} + \delta_{j'+1,l}$ the following commutation relations hold:

$$[\mathbf{E}_j, \mathbf{E}_{j'}] = \delta_{j,j'}; \tag{4.16}$$

$$[\mathbf{E}_{j'}, \mathbf{K}_{j,l,\pm}] = \pm\alpha_{j,l,j'}\mathbf{K}_{j,l,\pm}; \tag{4.17}$$

$$[\mathbf{K}_{j,l,+}, \mathbf{K}_{j,l,-}] = \mathbf{E}_j + ... + \mathbf{E}_{l-1} \tag{4.18}$$

From these commutation relations it follows that the representation space of $\mathbf{f}$ can be written as a direct sum of subspaces indexed by lists $\boldsymbol{m} = [m_1, .., m_{d-1}]$ of eigenvalues of simultaneous eigenvectors of $\{\mathbf{E}_j\}_{j=1,...,d-1}$. These lists of eigenvalues are called weights, and the subspace with weight $\boldsymbol{m}$ is indicated as $\mathcal{H}_{\boldsymbol{m}}$. Due to Eq. (4.17), for $v \in \mathcal{H}_{\boldsymbol{m}}$ we have $\mathbf{K}_{j,l,\pm}v \in \mathcal{H}_{\boldsymbol{m}\pm\boldsymbol{\alpha}_{j,l}}$ or $\mathbf{K}_{j,l,\pm}v = 0$. In fact, one obtains:

Theorem 4.1.2 (Irreducible representations of $\mathfrak{su}(d)$). *For any finite dimensional irreducible representation of $\mathfrak{su}(d)$ there exists a unique weight $\boldsymbol{w}$ of non-negative integers such that*

- $\mathcal{H}_{\boldsymbol{w}}$ *is one-dimensional,*
- $\mathbf{K}_{j,l,-}v = 0$ *for any* $v \in \mathcal{H}_{\boldsymbol{w}}$.

Irreducible representations of $\mathfrak{su}(d)$ with the same weight are equivalent, and labeled as $\mathbf{f}_{\boldsymbol{w}}$.

The representation space of $\mathbf{f}_{\boldsymbol{w}}$ is generated by the vectors $\mathbf{K}_{j,l,-}^{n_{j,l}}v$ for $v \in \mathcal{H}_{\boldsymbol{w}}$, therefore all the other weights of the irreducible representation are integer vectors. We can obtain irreducible representations of $\mathrm{SU}(d)$ from these irreducible representations of $\mathfrak{su}(d)$:

Proposition 4.1.2 (Irreducible representations of $\mathrm{SU}(d)$ and $\mathfrak{su}(d)$). *The irreducible representations of $\mathrm{SU}(d)$ can be obtained from the irreducible finite dimensional representations of $\mathfrak{su}(d)$, according to the definition*

$$\mathbf{f}_{\boldsymbol{w}}(U) := e^{\mathbf{f}_{\boldsymbol{w}}(A)}, \qquad U = e^A, \qquad A \in \mathfrak{su}(d). \tag{4.19}$$

This is a well defined representation, since for any matrix U there exists a traceless anti-hermitian matrix A such that $U = e^A$, and $\mathrm{SU}(d)$ is simply connected [Hal15].

Expanding $e^{\mathbf{f}(tA)} = e^{t\mathbf{f}(A)}$ at first order in t, if an invariant space exists for $e^{\mathbf{f}(tA)}$ then it is also an invariant space for $\mathbf{f}(A)$.

From this discussion it follows that irreducible representation of SU(d) can also be labeled by maximum weights. We will need representations with maximum weights that can be obtained from partitions. The maximum weight of the partition $(\lambda_1, ..., \lambda_d)$ is $[\lambda_1 - \lambda_2, \lambda_2 - \lambda_3, ..., \lambda_d - \lambda_{d-1}]$. We label these representations as $\mathcal{U}_\lambda(\mathrm{SU}(d))$.

The character of the irreducible representation corresponding to a partition λ is a function of the eigenvalues of $U \in \mathrm{SU}(d)$ and is given by the Schur polynomial:

$$s_\lambda(x_1, ..., x_d) := \frac{\det(x_i^{\lambda_j+\delta_j})}{\det(x_i^{\delta_j})} \tag{4.20}$$

where $x_i^{\lambda_j}$ denotes a $d \times d$ matrix with such entries and $\delta = (d-1, d-2, ..., 1, 0)$. The dimension of the Hilbert space $\mathcal{U}_\lambda(\mathrm{SU}(d))$ is given by $\omega_\lambda^{(d)} := s_\lambda(1, ..., 1)$, and corresponds to the number of semistandard Young tableau of shape λ and boxes filled with the integers $\{1, ..., d\}$.

4.1.5 A special case: SU(2)

The case of SU(2) is much easier to handle, with many explicit formulas available [VMK88]. In this case, the highest weight is given by an integer, and irreducible representations are often labeled by half the maximum weight, i.e. by an half integer j. The dimension of the irreducible representation j is $2j + 1$. The representation theory of SU(2) is of fundamental importance in physics, since it describes transformations of particles with spin, and irreducible representations of SU(2) with integer j are also irreducible representations of $SO(3)$, which describe rotations in space and appears in the description of the Hilbert space of particles in three dimensions. In our case SU(2) is interesting as it describes unitary transformations of qubits. Moreover, when two pure state in generic dimensions are involved, they effectively span the space of a qubit and the representation theory of SU(2) becomes relevant. We denote the representation matrix elements as $D^j_{m,n}(U)$, where j is the irreducible representation label and m and n are the weights, in this case integers between $-j$ and j. For SU(2), the decomposition in Eq. (4.5) simplifies considerably as

$$\mathcal{U}_{j_1} \otimes \mathcal{U}_{j_2} \cong \bigoplus_{|j_1 - j_2| \leq j \leq j_1 + j_2} \mathcal{U}_j, \tag{4.21}$$

since the multiplicities $C^j_{j_1,j_2}(\mathrm{SU}(2))$ either 1 if $|j_1 - j_2| \leq j \leq j_1 + j_2$ and 0 otherwise. Given that the multiplicity of j in the tensor product of j_1 and j_2 is at most one,

the Clebsch-Gordan coefficents are the matrix elements of the unitary transformation connecting the basis $\{|j,m\rangle\}$ and $\{|j_1,m_1,j_2,m_2\rangle\}$, where $|j_1,m_1,j_2,m_2\rangle := |j_1,m_1\rangle \otimes |j_2,m_2\rangle$.

$$C^{j,m}_{j_1,m_1,j_2,m_2} := \langle j,m|j_1,m_1,j_2,m_2\rangle \tag{4.22}$$

when $|j_1 - j_2| \leq j \leq j_1 + j_2$, and 0 otherwise. These coefficients can be chosen to be real. For the Wigner matrices, orthogonality relations Eq. (4.3) read

$$\int dU \overline{D^j_{m,n}(U)} D^{j'}_{m',n'}(U) = \frac{\delta_{j,j'}\delta_{m,m'}\delta_{n,n'}}{2j+1} \tag{4.23}$$

Using Clebsch-Gordan coefficents and orthogonality relations, we can transform any integral of products of elements of Wigner matrices into a contraction of Clebsch-Gordan coefficients. For example, using

$$D^{j_1}_{m_1,n_1}(U) D^{j_2}_{m_2,n_2}(U) = \sum_{|j_1-j_2|\leq j\leq j_1+j_2} C^{j,m_1+m_2}_{j_1,m_1,j_2,m_2} D^{j}_{m_1+m_2,n_1+n_2}(U) C^{j,n_1+n_2}_{j_1,n_1,j_2,n_2} \tag{4.24}$$

we can compute

$$\int dU \overline{D^j_{m,n}(U)} D^{j_1}_{m_1,n_1}(U) D^{j_2}_{m_2,n_2}(U) = C^{j,m}_{j_1,m_1,j_2,m_2} C^{j,n}_{j_1,n_1,j_2,n_2} \frac{\delta_{j,j_2+j_2}\delta_{m,m_1+m_2}\delta_{n,n_1+n_2}}{2j+1}. \tag{4.25}$$

By viewing Clebsch-Gordan coefficients as tensor with three decorated legs, where the decoration is the irreducible representation associated with the leg, one can pictorially represent these contractions, and there exists a graphical calculus which enables to simplify the resulting symbols [VMK88]. A general contraction may have free legs, that is legs which are not contracted. We will make use of an identity which is obtainable from this graphical calculus in Chapter 6, but we won't treat the rules of this graphical calculus, for which we refer to [VMK88; MD19].

4.2 Symmetries and optimal measurements

4.2.1 Covariant and invariant measurements

One of the most important application of representation theory in this thesis is to characterize optimal measurement in presence of symmetries. We now state two general results that will be used several times in this thesis. A wider treatment of these ideas can be found in [Hay17a].

For Θ homogeneous space under the action of G, a *covariant family* with respect to a unitary representation $\mathbf{f}$ of G is a family of states $\{\rho_\theta\}$ such that

$$\mathbf{f}(g)\rho_\theta \mathbf{f}(g)^\dagger = \rho_{g\theta} \tag{4.26}$$

We also assume that $\rho_\theta = \rho_{g\theta_0}$ for some $g \in G$ and fixed $\theta_0 \in \Theta$. The first results concerns *covariant measurements*, i.e. POVM of the form $\{\mathbf{f}(g)E_0\mathbf{f}(g)^\dagger\}_{g\in G}$, which estimate θ as $\tilde{\theta} = g\theta_0$ for some $\theta_0 \in \Theta$. A measure $d\mu(\theta)$ on Θ can be obtained from the left and right invariant measure dg of G. Assuming such measure exists, a generic POVM to estimate θ is $\{M_{\tilde{\theta}}\}$, $\int_\Theta d\mu(\tilde{\theta})M_{\tilde{\theta}} = I$, and we measure the accuracy of the estimation with some cost function $l(\tilde{\theta}, \theta) \geq 0$, with the property that $l(\tilde{\theta}, \theta) = l(g\tilde{\theta}, g\theta)$ for any $g \in G$. The set of covariant POVMs is denoted as $\mathcal{M}^{\text{cov}}(\Theta)$. We have that

Theorem 4.2.1 (Optimality of covariant measurements). *For a covariant family $\{\rho_\theta\}$ and a group G with Haar measure dg a measurement that minimizes the worst case cost*

$$\min_{M\in\mathcal{M}(\Theta)} \max_{\theta\in\Theta} \int_\Theta d\mu(\tilde{\theta}) \operatorname{tr}\big[M_{\tilde{\theta}}\rho_\theta\big] l(\tilde{\theta}, \theta), \tag{4.27}$$

and the average (Bayesian) cost

$$\min_{M\in\mathcal{M}(\Theta)} \int_\Theta d\mu(\theta) \int_\Theta d\mu(\tilde{\theta}) \operatorname{tr}\big[M_{\tilde{\theta}}\rho_\theta\big] l(\tilde{\theta}, \theta). \tag{4.28}$$

can be chosen to be covariant, i.e.

$$M_{\tilde{g}} = \mathbf{f}(\tilde{g})E_0\mathbf{f}(\tilde{g})^\dagger \qquad \int_G d\tilde{g}\mathbf{f}(\tilde{g})E_0\mathbf{f}(\tilde{g})^\dagger = I \tag{4.29}$$

with an estimator $\tilde{\theta} := \tilde{g}\theta_0$.

Proof. Given a POVM $\{M_{\tilde{\theta}}\}$, we define its covariant counterpart $\{M^{\text{cov}}_{\tilde{g}}\} = \mathbf{f}(\tilde{g})E_{\theta_0}\mathbf{f}(\tilde{g})^\dagger$, $E_{\theta_0} = \int_G dg\mathbf{f}(g)^\dagger M_{g\theta_0}\mathbf{f}(g)$

$$\begin{aligned}
&\int_\Theta d\mu(\theta) \int_\Theta d\mu(\tilde{\theta}) \operatorname{tr}\Big[M_{\tilde{\theta}}\rho_\theta l(\tilde{\theta}, \theta)\Big] = \int_\Theta d\mu(\tilde{\theta}) \int_G dg \operatorname{tr}\Big[M_{\tilde{\theta}}\mathbf{f}(g)\rho_{\theta_0}\mathbf{f}(g)^\dagger\Big] l(\tilde{\theta}, g\theta_0) \\
&= \int_G d\tilde{g} \int_G dg \operatorname{tr}\Big[M_{\tilde{g}\theta_0}\mathbf{f}(g)\rho_{\theta_0}\mathbf{f}(g)^\dagger\Big] l(\tilde{g}\theta_0, g\theta_0) \\
&= \int_G d\tilde{g} \int_G dg' \int_G dg \operatorname{tr}\Big[M_{\tilde{g}\theta_0}\mathbf{f}(\tilde{g}g'^{-1}g)\rho_{\theta_0}\mathbf{f}(\tilde{g}g'^{-1}g)^\dagger\Big] l(\tilde{g}\theta_0, \tilde{g}g'^{-1}g\theta_0) \\
&= \int_G d\tilde{g} \int_G dg' \int_G dg \operatorname{tr}\Big[M_{\tilde{g}\theta_0}\mathbf{f}(\tilde{g}g'^{-1}g)\rho_{\theta_0}\mathbf{f}(\tilde{g}g'^{-1}g)^\dagger\Big] l(g'\theta_0, g\theta_0) \\
&= \int_G dg' \int_G dg \operatorname{tr}\Big[\mathbf{f}(g'^{-1})^\dagger E_{\theta_0}\mathbf{f}(g'^{-1})\mathbf{f}(g)\rho_{\theta_0}\mathbf{f}(g)^\dagger\Big] l(g'\theta_0, g\theta_0) \\
&= \int_G d\tilde{g} \int_G dg \operatorname{tr}\Big[\mathbf{f}(\tilde{g})E_{\theta_0}\mathbf{f}(\tilde{g})^\dagger\mathbf{f}(g)\rho_{\theta_0}\mathbf{f}(g)^\dagger\Big] l(\tilde{g}\theta_0, g\theta_0), \qquad (4.30) \\
&= \int_\Theta d\mu(\theta) \int_G d\tilde{g} \operatorname{tr}\big[M^{\text{cov}}_{\tilde{g}}\rho_\theta\big] l(\tilde{g}\theta_0, \theta), \qquad (4.31)
\end{aligned}$$

where the first two equalities comes from the definition of the measure on Θ, the third from the left invariance of dg, the fourth from invariance of $l(\theta, \theta')$, the following from the definitions of the covariant counterpart of $\{M_{\tilde{\theta}}\}$.

This shows that in the Bayesian setting $\{M^{\mathrm{cov}}_{\tilde{g}}\}$ has the same performance of $\{M_{\tilde{\theta}}\}$, therefore we can restrict the optimization to covariant POVMs. The fact that covariant POVMs minimize the worst case cost follows since worst case cost is always higher than the average cost

$$\min_{M\in\mathcal{M}(\Theta)} \max_{\theta\in\Theta} \int_\Theta d\mu(\tilde{\theta}) \operatorname{tr}\big[M_{\tilde{\theta}}\rho_\theta\big] l(\tilde{\theta},\theta) \geq \min_{M\in\mathcal{M}^{\mathrm{cov}}(\Theta)} \int_\Theta d\mu(\theta) \int_\Theta d\mu(\tilde{\theta}) \operatorname{tr}\big[M_{\tilde{\theta}}\rho_\theta\big] l(\tilde{\theta},\theta), \tag{4.32}$$

the average cost of a POVM is equal to the average cost of its covariant counterpart, and the inequality above is saturated for a covariant POVM.

□

We note that for a covariant POVM, the total POVM element corresponding to θ_0 is $M_{\theta_0} = \int_H dh \mathbf{f}(h) E_{h\theta_0} \mathbf{f}(h)^\dagger$, therefore it commutes with $\mathbf{f}(h')$ for any $h' \in H$. The structure of covariant POVMs is thus simplified by applying Schur's lemma to M_{θ_0} according to irreducible representations of H.

An example of a covariant family is given by n copies of a pure state of $\mathbb{C}^d$, i.e. $|\psi\rangle\langle\psi|^{\otimes n} \in \mathbb{C}^{nd}$. In this case the group action is given by $\mathrm{SU}(d)$, and we choose as initial point a state $|0\rangle \in \mathbb{C}^d$. The set of this states is supported in the completely symmetric subspace of $\mathbb{C}^{nd}$, which hosts an irreducible representation of $\mathrm{SU}(d)$, of dimension $\binom{n+d-1}{d-1}$. A covariant POVM is given by $E_{U|0\rangle} = \binom{n+d-1}{d-1} U^{\otimes n} |0\rangle\langle 0|^{\otimes n} U^{\otimes n\dagger}$, and it is optimal to estimate $|\psi\rangle$ under various figures of merit [Hay97]. The states $U^{\otimes n}|0\rangle$ are an example of *coherent states*, obtained as the orbit of a maximum weight vector under the action of an irreducible representation [Per72].

Suppose now we have a collection of states $\{\rho_{\theta,\eta}\}$ such that it is covariant with respect to $\theta \in \Theta$ at each fixed $\eta \in \mathfrak{E}$, with a unitary representation $\mathbf{f}$ independent of η. We want to estimate η, while we are not interested in θ. We measure the accuracy of the estimation with some function $l(\tilde{\eta}, \eta) \geq 0$. We have that

Theorem 4.2.2 (Optimality of invariant measurements). *For a covariant family $\{\rho_{\theta,\eta}\}$ with respect to θ and a group G with Haar measure dg, a measurement that minimizes the worst case cost for the estimation of η*

$$\min_{M\in\mathcal{M}(\mathfrak{E})} \max_{\theta\in\Theta,\eta\in\mathfrak{E}} \int_{\mathfrak{E}} d\tilde{\eta} \operatorname{tr}[M_{\tilde{\eta}}\rho_{\theta,\eta}] l(\tilde{\eta},\eta) \tag{4.33}$$

and the worst case average (Bayesian) cost at fixed η

$$\min_{M\in\mathcal{M}(\mathfrak{E})}\max_{\eta\in\mathfrak{E}}\int_{\mathfrak{E}}d\tilde{\eta}\int_{\Theta}d\mu(\theta)\,\mathrm{tr}[M_{\tilde{\eta}}\rho_{\theta,\eta}]l(\tilde{\eta},\eta). \tag{4.34}$$

and the global Bayesian cost for a prior distribution

$$\min_{M\in\mathcal{M}(\mathfrak{E})}\int_{\mathfrak{E}}d\eta\,p(\eta)\int_{\mathfrak{E}}d\tilde{\eta}\int_{\Theta}d\mu(\theta)\,\mathrm{tr}[M_{\tilde{\eta}}\rho_{\theta,\eta}]l(\tilde{\eta},\eta). \tag{4.35}$$

can be chosen to be invariant, i.e.

$$M_\eta=\mathbf{f}(g)M_\eta\mathbf{f}(g)^\dagger,\qquad\forall g\in G. \tag{4.36}$$

In particular, the minimum worst case cost and of the minimum worst case Bayesian cost at fixed η are the same, and attained by the same POVM.

Proof. For any POVM $\{M_{\tilde{\eta}}\}$ we denote its corresponding invariant POVM as $\{M^{\mathrm{inv}}_{\tilde{\eta}}\}$ with $M^{\mathrm{inv}}_{\tilde{\eta}}=\int_G dg\mathbf{f}(g)M_{\tilde{\eta}}\mathbf{f}(g)^\dagger$. We have that

$$\begin{aligned}
&\int_{\mathfrak{E}}d\tilde{\eta}\int_{\Theta}d\mu(\theta)\,\mathrm{tr}[M_{\tilde{\eta}}\rho_{\theta,\eta}]l(\tilde{\eta},\eta)\\
&=\int_{\mathfrak{E}}d\tilde{\eta}\int_{G}dg\,\mathrm{tr}\Big[M_{\tilde{\eta}}\mathbf{f}(g)\rho_{\theta_0,\eta}\mathbf{f}(g)^\dagger\Big]l(\tilde{\eta},\eta)\\
&=\int_{\mathfrak{E}}d\tilde{\eta}\int_{G}dg'\int_{G}dg\,\mathrm{tr}\Big[M_{\tilde{\eta}}\mathbf{f}(g'g)\rho_{\theta_0,\eta}\mathbf{f}(g'g)^\dagger\Big]l(\tilde{\eta},\eta)\\
&=\int_{\mathfrak{E}}d\tilde{\eta}\,\mathrm{tr}\Big[M^{\mathrm{inv}}_{\tilde{\eta}}\mathbf{f}(g)\rho_{\theta_0,\eta}\mathbf{f}(g)^\dagger\Big]l(\tilde{\eta},\eta).
\end{aligned}\tag{4.37}$$

Therefore $\{M^{\mathrm{inv}}_{\tilde{\eta}}\}$ has the same performance of $\{M_{\tilde{\eta}}\}$ in the Bayesian case. The same is true for the global Bayesian cost, by the same derivation. We have that

$$\begin{aligned}
\max_{\theta\in\Theta}\int_{\mathfrak{E}}d\tilde{\eta}\,\mathrm{tr}[M_{\tilde{\eta}}\rho_{\theta,\eta}]l(\tilde{\eta},\eta)&\geq\int_{\mathfrak{E}}d\tilde{\eta}\int_{\Theta}d\mu(\theta)\,\mathrm{tr}[M_{\tilde{\eta}}\rho_{\theta,\eta}]l(\tilde{\eta},\eta)\\
&=\int_{\mathfrak{E}}d\tilde{\eta}\int_{\Theta}d\mu(\theta)\,\mathrm{tr}\big[M^{\mathrm{inv}}_{\tilde{\eta}}\rho_{\theta,\eta}\big]l(\tilde{\eta},\eta)
\end{aligned}\tag{4.38}$$

where the first inequality is saturated when $M^{\mathrm{inv}}_{\tilde{\eta}}=M_{\tilde{\eta}}$. This means

$$\max_{\theta\in\Theta,\eta\in\mathfrak{E}}\int_{\mathfrak{E}}d\tilde{\eta}\,\mathrm{tr}[M_{\tilde{\eta}}\rho_{\theta,\eta}]l(\tilde{\eta},\eta)\geq\max_{\eta\in\mathfrak{E}}\int_{\mathfrak{E}}d\tilde{\eta}\int_{\Theta}d\mu(\theta)\,\mathrm{tr}\big[M^{\mathrm{inv}}_{\tilde{\eta}}\rho_{\theta,\eta}\big]l(\tilde{\eta},\eta) \tag{4.39}$$

which is saturated when $M^{\mathrm{inv}}_{\tilde{\eta}}=M_{\tilde{\eta}}$. □

An example of a family $\{\rho_{\theta,\eta}\}$ covariant with respect to θ is the family of i.i.d states $\rho^{\otimes n}$ of dimension d, since $\rho = U\rho_{\text{diag}}U^\dagger$ for some $U \in \text{SU}(d)$ and ρ_{diag} positive-semidefinite diagonal $d \times d$ matrix. Here η can be the vector of eigenvalues of ρ. In fact, an optimal measurement for quantum tomography in sample complexity can be obtained by first estimating η with an invariant measurement, and then estimating U with a covariant measurement [Key06; OW16; OW17; Haa+17]. We recall the specific symmetries of i.i.d. states in the next section. We also mention that estimation problems in presence of symmetries can also be framed in the context of the resource theory of asymmetry [GS08; MS13; MS14b; MS14a].

Finally, we note that the theorems discussed in this section apply also to the discrete case of hypothesis testing, where the space of alternatives is discrete and a finite group acts on it.

4.2.2 Schur-Weyl duality

In this section we review a deep representation theory result that connects representations of the compact group $\text{SU}(d)$ and the symmetric group S_n, and it is of fundamental importance for quantum statistical inference. Consider the state space of n, d-dimensional systems, $\mathcal{H}_d^{\otimes n}$. This space naturally hosts unitary representations of these two groups, therefore a unitary represantation of the product $\text{SU}(d) \times S_n$.

Specifically, $\text{SU}(d)$ and S_n act on a basis $\{|i_1\rangle \otimes |i_2\rangle \otimes ... \otimes |i_n\rangle\}_{i_1,i_2....,i_n}$ of $\mathcal{H}_d^{\otimes n}$ via unitary representations $\mathbf{u}_n : \text{SU}(d) \to \text{U}(\mathcal{H}_d^{\otimes n})$, $\mathbf{s}_n : S_n \to \text{U}(\mathcal{H}_d^{\otimes n})$ as follows

$$\begin{aligned}
\mathbf{u}_n(U)\,|i_1\rangle \otimes |i_2\rangle \otimes ... \otimes |i_n\rangle &= U^{\otimes n}\,|i_1\rangle \otimes |i_2\rangle \otimes ... \otimes |i_n\rangle \\
&= U\,|i_1\rangle \otimes U\,|i_2\rangle \otimes ... \otimes U\,|i_n\rangle\,, \quad \forall U \in \text{SU}(d) \qquad (4.40)\\
\mathbf{s}_n(\tau)\,|i_1\rangle \otimes |i_3\rangle \otimes ... \otimes |i_n\rangle &= \left|\tau^{-1}(i_1)\right\rangle \otimes \left|\tau^{-1}(i_2)\right\rangle \otimes ... \otimes \left|\tau^{-1}(i_n)\right\rangle, \forall \tau \in S_n.
\end{aligned}$$

In particular, observe that the two representations commute, i.e. $[\mathbf{u}_n(U), \mathbf{s}_n(\tau)] = 0$, $\forall U \in \text{SU}(d)$, and $\forall \tau \in S_n$. Already by Schur's lemma, we are able to tell that $\mathbf{u}_n(U)$ should be block diagonal according to the decomposition into irreducible representations of S_n, and $\mathbf{s}_n(\tau)$ should be block diagonal according to the decomposition into irreducible representations of $\text{SU}(d)$. In fact, something stronger can be said. As we saw in previous sections, we associate a pair of irreducible representations of $\text{SU}(d)$ and S_n to any partition of size n and length at most d, with associated Young diagrams. Let $Y_{n,d}$ denote be the set of integer partitions of n in at most d parts. $\lambda \in Y_{n,d}$ can then also be written as a vector $\lambda = (\lambda_1, \lambda_2, ..., \lambda_d)$ with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d \geq 0$. We have

Theorem 4.2.3 (Schur-Weyl duality [Wey47; GW09]). *$\mathcal{H}_d^{\otimes n}$ can be decomposed as*

$$\mathcal{H}_d^{\otimes n} \cong \bigoplus_{\lambda \in Y_{n,d}} \mathcal{U}_\lambda(\mathrm{SU}(d)) \otimes \mathcal{V}_\lambda(S_n), \tag{4.41}$$

where the unitary irreducible representation (irrep) $\mathbf{u}_\lambda$ of $\mathrm{SU}(d)$ acts non trivially on the factor $\mathcal{U}_\lambda(\mathrm{SU}(d))$ of dimension $\omega_\lambda^{(d)}$ and the irrep $\mathbf{s}_\lambda$ of S_n acts non trivially on the factor $\mathcal{V}_\lambda(S_n)$ of dimension ω_λ.

A proof of the theorem can be found in [GW09], while [Hay17b; Hay17a] collects several applications in quantum information theory. The use of the congruence sign in Eq. (4.41) indicates that this block decomposition is accomplished by a unitary transformation; in the case considered here this unitary is the Schur transform. One can then ask how hard is to implement this change of basis on a quantum computer in the gate model, and efficient circuits for the Schur transform have been found [BCH06; Har05a; Kro19].

A state $\rho^{\otimes n} \in \mathcal{D}(\mathcal{H}_d^{\otimes n})$ is invariant under the action of $s_n(\sigma)$ for any σ. By Schur's lemma, $\rho^{\otimes n}$ can be decomposed in block diagonal form according to the isomorphism in Eq. (4.41).

$$\rho^{\otimes n} = \sum_{\lambda \in Y_{n,d}} \mathrm{SW}_\rho^n(\lambda) \rho_\lambda \otimes \frac{I_\lambda}{\omega_\lambda}, \tag{4.42}$$

where $\mathrm{SW}_\rho^n(\lambda)$ is a probability distribution over the Young diagrams. It is immediate to see that $\mathrm{SW}_\rho^n(\lambda)$ depends only on the number of copies n and on the spectrum of ρ, and ρ_λ are $\omega_\lambda^{(d)}$-dimensional states. In particular, one can compute SW_ρ^l from the characters of $\mathrm{SU}(d)$ as $\mathrm{SW}_\rho^n(\lambda) = s_\lambda(\eta_1, ..., \eta_d)$, where $\eta_1, ..., \eta_d$ are the eigenvalues of ρ. This is due to the fact that $\mathbf{u}_n(U)$ is a polynomial representation of $\mathrm{SU}(d)$, meaning that the entries of $\mathbf{u}_n(U)$ are polynomials in the entries of U: one can explicitly construct these polynomials from the Schur transform. Therefore, applying the Schur transform to $\rho^{\otimes n}$, the entries of $\mathrm{SW}_\rho^l(\lambda)\rho_\lambda$ will be a polynomial function of the entries of ρ, where the polynomial function is the same of $\mathbf{u}_n$. In particular, $\mathrm{tr}\big[\mathrm{SW}_\rho^n(\lambda)\rho_\lambda\big] = \mathrm{SW}_\rho^n(\lambda)$ can be computed with a unique polynomial function of eigenvalues, the same that gives the character of $\mathbf{u}_\lambda$ from the eigenvalues of U. For a more detailed discussion of this feature we refer to [Har05b; Chr06], where this fact is discussed by noting that Schur-Weyl duality holds also for the action of the larger group $\mathrm{GL}(d)$.

The projective measurement given by projectors on $\mathcal{U}_\lambda(\mathrm{SU}(d)) \otimes \mathcal{V}_\lambda(S_n)$, $\{\Pi_\lambda\}_{\lambda \in Y_{n,d}}$ is called weak Schur sampling [Har05a; Kro19], and it can be executed with gate complexity $O(l, \log d, \log 1/\delta)$, where δ is the precision of the implementation. From Theorem 4.2.2 and the discussion just made it should be clear that it is the optimal measurement to estimate the spectrum of a state. Indeed, the following important result holds [ARS88; KW01; HM02; CM06]:

Theorem 4.2.4 (Spectrum estimation). *We have*

$$P(D(\bar{\lambda}||s) > \epsilon) := \sum_{\lambda: D(\bar{\lambda}||s) > \epsilon} \mathrm{SW}_\rho^n(\lambda) \leq (n+1)^{d(d+1)/2} \exp(-\epsilon n) \tag{4.43}$$

In particular, with $O(d^2/\epsilon^2)\log(d/\epsilon)\log(1/\delta)$ copies of ρ the estimate $\bar{\lambda}$ obtained from the outcome λ of weak Schur sampling satisfies $D(\bar{\lambda}||s) \leq \epsilon$ with probability at least $1-\delta$. For state of rank r, one can substitute d with r.

From this large deviation bound sufficient sample complexities for estimating any unitarily invariant quantity can be obtained. This result can be seen as a quantum equivalent of the tail bounds for classical distributions based on the method of types [CT05], and it was used for constructing universal quantum source coding protocols [HM02], universal entanglement concentration [MH04], universal quantum Stein's lemma [Hay02] and universal classical-quantum channel coding [Hay08]. Recent works have attacked directly the estimation of the spectrum [OW15] and entropies [AKG19] studying the bias and the variance of suitable estimators based on weak Schur sampling, and obtaining upper and lower bounds on the sample complexities.

A general observable that can be obtained by post-processing of weak Schur sampling, can be written as $\mathcal{O} = \sum_{\lambda \in Y_{n,d}} O_\lambda \Pi_\lambda$. This is always the case for linear combinations of permutations, which are also invariant under conjugation with permutations, i.e. $\mathcal{O} = \sum_{\tau \in S_n} c_\tau \mathbf{s}_n(\tau)$, $\mathbf{s}_n(\sigma)\mathcal{O}\mathbf{s}_n(\sigma)^\dagger = \mathcal{O}$ for every $\sigma \in S_n$.

Finally, for any decomposition $\mathcal{H}_d^{\otimes n} = \otimes_{i=1}^N \mathcal{H}_d^{\otimes m_i}$ (where $\sum_{i=1}^N m_i = n$), one can define a family of weak Schur sampling projectors for each factor, $\{\Pi_\lambda^{(i)}\}_{\lambda \in Y_{m_i,d}}$, which give commuting projective measurements which we call local weak Schur sampling. Nonetheless, we can also consider the global weak Schur sampling measurement given by $\{\Pi_\lambda\}_{\lambda \in Y_{n,d}}$. Since $\{\Pi_\lambda\}_{\lambda \in Y_{n,d}}$ are invariant under local permutations, by Schur's lemma they commute with the projectors $\{\otimes_{i=1}^N \Pi_{\lambda_i}^{(i)}\}_{\lambda_i \in Y_{m_i,d}}$, therefore local and global weak Schur sampling can be done with a unique projective measurement. The probabilities of the outcomes do not depend on the order in which local and global measurements are executed. Therefore, the composition of global and local weak Schur samplings define a further projective measurement which we call nested weak Schur sampling. In particular, this measurement is also efficient in gate complexity.

Estimation of unitarily invariant properties of set of states

We will be interested in estimating unitarily invariant quantities of sets of states. In this section, we make some general observations on this problem, using the representation theory results recalled in this chapter.

One has the following recoupling of two $\mathrm{SU}(d)$ irreducible representations:

$$\mathcal{U}_\mu(\mathrm{SU}(d)) \otimes \mathcal{U}_\nu(\mathrm{SU}(d)) = \bigoplus_\lambda \mathcal{U}_\lambda(\mathrm{SU}(d)) \otimes \mathbb{C}^{C^\lambda_{\mu,\nu}} \tag{4.44}$$

with $C^\lambda_{\mu,\nu}$ called the Little-Richardson coefficient. Combining this fact with (4.41) we have

$$\mathcal{H}_d^{\otimes(h+k)} \cong \bigoplus_{\mu,\nu} \mathcal{U}_\mu(\mathrm{SU}(d)) \otimes \mathcal{V}_\mu(S_h) \otimes U_\nu(\mathrm{SU}(d)) \otimes \mathcal{V}_\nu(S_k) \tag{4.45}$$

$$\cong \bigoplus_{\mu,\nu,\lambda} \mathcal{U}_\lambda(\mathrm{SU}(d)) \otimes \mathbb{C}^{C^\lambda_{\mu,\nu}} \otimes \mathcal{V}_\mu(S_h) \otimes \mathcal{V}_\nu(S_k). \tag{4.46}$$

We also have of course

$$\mathcal{H}_d^{\otimes(h+k)} \cong \bigoplus_{\lambda} \mathcal{U}_\lambda(\mathrm{SU}(d)) \otimes \mathcal{V}_\lambda(S_{h+k}). \tag{4.47}$$

We will now use the abbreviations A to indicate the subsystem $\mathcal{H}_d^{\otimes(h)}$, B for the subsystem $\mathcal{H}_d^{\otimes(k)}$ and AB for the complete system $\mathcal{H}_d^{\otimes(h+k)}$. Defining the projectors Π^{AB}_λ on $\mathcal{U}_\lambda(\mathrm{SU}(d)) \otimes \mathcal{V}_\lambda(S_{h+k})$, Π^A_μ on $\mathcal{U}_\mu(\mathrm{SU}(d)) \otimes \mathcal{V}_\mu(S_h)$, Π^B_ν on $\mathcal{U}_\nu(\mathrm{SU}(d)) \otimes \mathcal{V}_\nu(S_k)$, it follows that for a state $\rho^{\otimes h} \otimes \sigma^{\otimes k}$ one has the following block diagonal decomposition

$$\rho^{\otimes h} \otimes \sigma^{\otimes k} = \sum_{\mu,\nu,\lambda} p_{\mu,\nu}(\rho,\sigma)\psi(\rho,\sigma)_{\mu,\nu} \otimes \frac{I_\mu}{\omega_\mu} \otimes \frac{I_\nu}{\omega_\nu}, \tag{4.48}$$

with $\psi(\rho,\sigma)_{\mu,\nu} \in \Sigma(\oplus_\lambda \mathcal{U}_\lambda(\mathrm{SU}(d)) \otimes \mathbb{C}^{C^\lambda_{\mu,\nu}})$, and $\frac{I_\mu}{\omega_\mu} \in \Sigma(\mathcal{V}_\mu(S_h))$, $\frac{I_\nu}{\omega_\nu} \in \Sigma(\mathcal{V}_\nu(S_k))$ with $p_{\mu,\nu}(\rho,\sigma) = \mathrm{tr}\big[(\Pi^A_\mu \otimes \Pi^B_\nu)\rho^{\otimes h} \otimes \sigma^{\otimes k}\big]$. Note that $[\Pi^{AB}_\lambda, (\Pi^A_\mu \otimes \Pi^B_\nu)] = 0$

Suppose now we are interested in estimating a unitarily invariant property of the pair of states (ρ,σ), that is a function $f(\rho,\sigma)$ such that $f(\rho,\sigma) = f(U\rho U^\dagger, U\sigma U^\dagger)$. The family of pair of states $\{\rho,\sigma\}$ is covariant under the action of $(\rho,\sigma) \to (U\rho U^\dagger, U\sigma U^\dagger)$. From Theorem 4.2.2, we can estimate $f(\rho,\sigma)$ optimally with an invariant measurement. Using the invariance of the measurement, the probability distribution of the outcome of a measurement are the same if the measurement is done on $\int dU (U\rho U^\dagger)^{\otimes h} \otimes (U\sigma U^\dagger)^{\otimes k}$ instead of $\rho^{\otimes h} \otimes \sigma^{\otimes k}$. One obtains

$$\int dU (U\rho U^\dagger)^{\otimes h} \otimes (U\sigma U^\dagger)^{k} = \sum_{\mu,\nu,\lambda} p_{\mu,\nu,\lambda}(\rho,\sigma) \frac{I_\lambda}{\omega^{(d)}_\lambda} \otimes \tilde{\psi}(\rho,\sigma)_{\mu,\nu,\lambda} \otimes \frac{I_\mu}{\omega_\mu} \otimes \frac{I_\nu}{\omega_\nu} \tag{4.49}$$

with $\tilde{\psi}(\rho,\sigma)_{\mu,\nu,\lambda} \in \Sigma(\mathbb{C}^{C^{\lambda}_{\mu,\nu}})$, $p_{\mu,\nu,\lambda} = \mathrm{tr}\big[\Pi^{AB}_{\lambda}(\Pi^{A}_{\mu}\otimes\Pi^{B}_{\nu})\rho^{\otimes h}\otimes\sigma^{\otimes k}\big]$, $\frac{I_\lambda}{\omega_\lambda^{(d)}} \in \Sigma(\mathcal{U}_\lambda(\mathrm{SU(d)}))$. Since we can always estimate $f(\rho,\sigma)$ at arbitrary accuracy for h and k large enough, the following fact, which we state as a theorem, is then evident:

Theorem 4.2.5 (Estimation of unitarily invariant properties). *Provided that h and k are sufficiently large, all unitarily invariant properties $f(\rho,\sigma)$ can be estimated at any precision given $\rho^h\otimes\sigma^k$ by projecting with $\Pi^{AB}_{\lambda}(\Pi^{A}_{\mu}\otimes\Pi^{B}_{\nu})$, which extracts a triple (μ,ν,λ) according to $p_{\mu,\nu,\lambda}$, and by some measurement on $\mathcal{D}(\mathbb{C}^{C^{\lambda}_{\mu,\nu}})$.*

In the case of $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$ pure in Eq. (4.49), a drastic simplification arises.

Theorem 4.2.6. *For two pure states $|\psi\rangle$ and $|\phi\rangle$ with overlap $|\langle\psi|\phi\rangle|^2 = c$, we have*

$$\int_{\mathrm{SU}(d)} \mathrm{d}U\, U^{\otimes h+k}\left[(|\psi\rangle\langle\psi|)^{\otimes h}\otimes(|\phi\rangle\langle\phi|)^{\otimes k}\right]U^{\dagger\otimes(h+k)} \tag{4.50}$$

$$= \sum_J P_{h,k}(J|c)\,\frac{I_{\lambda_J}}{\omega_{\lambda_J}^{(d)}}\otimes|J_{h,k}\rangle\langle J_{h,k}|\,. \tag{4.51}$$

Here $\frac{I_{\lambda_J}}{\omega_{\lambda_J}^{(d)}}$ is the completely mixed state in $\mathcal{U}_{\lambda_J}(\mathrm{SU}(d))$, $\lambda_J := (\frac{h+k}{2}+J, \frac{h+k}{2}-J, 0, ..., 0)$, $|J_{h,k}\rangle\langle J_{h,k}| \in \Sigma(\mathcal{V}_{\lambda_J}(S_{h+k}))$ is independent of $|\psi\rangle$ and $|\phi\rangle$, and $P_{h,k}(J|c)$ is a probability distribution in J dependent only on c.

Proof. First of all, since $|\psi\rangle^{\otimes n}$ and $|\phi\rangle^{\otimes n}$ are invariant under permutations, the sum over μ and ν is restricted to Young diagrams of one row, corresponding to the the invariant representation of S_h and S_k, which have dimension 1. Moreover, using the invariance of the Haar measure we can insert for free an average over $V^{\otimes h+k}$, where V sampled from the Haar measure acts non trivially in the two-dimensional subspace $\mathcal{E}$ spanned by $|\psi\rangle$ and $|\phi\rangle$. By Schur-Weyl duality, we have $\mathcal{E}^{\otimes h+k} \cong \bigoplus_J \mathcal{U}_J(\mathrm{SU}(2))\otimes\mathcal{V}_{\lambda_J}(S_{h+k})$, where now we label SU(2) irreducible representations by the total angular momentum J, and the corresponding representations $\mathcal{V}_{\lambda_J}(S_{h+k})$ are those with Young diagram of two rows $\lambda_J = (\frac{h+k}{2}+J, \frac{h+k}{2}-J, 0, ..., 0)$, and J ranges from $|h-k|/2$ to $(h+k)/2$. Since the multiplicities in the couplings of two SU(2) irreducible representations are always zero or one, the state $\tilde{\psi}(\rho,\sigma)_{\mu,\nu,\lambda}$ in Eq. (4.49) is a trivial one-dimensional state and we can write:

$$\Big(\int_{\mathrm{SU}(2)} \mathrm{d}V\, V^{\otimes(N+M)}\left[(|\psi\rangle\langle\psi|)^{\otimes N}\otimes(|\phi\rangle\langle\phi|)^{\otimes M}\right]V^{\dagger\otimes(N+M)}\Big)$$
$$= \sum_J P_{h,k}(J|c)\,\frac{I_J}{2J+1}\otimes|J_{h,k}\rangle\langle J_{h,k}|\,, \tag{4.52}$$

where $|J_{h,k}\rangle\langle J_{h,k}|$ is a pure state in $\Sigma(\mathcal{V}_{\lambda_J}(S_{h+k}))$ which does not depend on $|\psi\rangle$ and $|\phi\rangle$, and $\frac{I_J}{2J+1}$ is the completely mixed state in $\mathcal{U}_J(\mathrm{SU}(2))$, and $P_{h,k}(J|c)$ is a probability

distribution on J, computed in Appendix A.1. Since $\mathcal{E}^{\otimes h+k}$ is an invariant subspace under the action of S_{h+k}, by Schur's lemma it can be block diagonalized according to irreducible representations of S_{h+k} on the whole space $\mathbb{C}^{d^{\otimes h+k}}$, and Schur-Weyl duality applied to $\mathcal{E}^{\otimes h+k}$ gives this decomposition. Therefore we have that $\sum_J P_{h,k}(J|c)\frac{I_J}{2J+1}\otimes|J_{h,k}\rangle\langle J_{h,k}|$ is also block diagonal according to the global Schur-Weyl duality decomposition, $\mathcal{H}_d^{\otimes(h+k)} = \oplus_\lambda \mathcal{U}_\lambda(\mathrm{SU}(d))\otimes\mathcal{V}_\lambda(S_{h+k})$, with $\frac{I_J}{2J+1}\in\Sigma(\mathcal{U}_{\lambda_J}(\mathrm{SU}(d)))$ and $|J_{h,k}\rangle\langle J_{h,k}|\in\Sigma(\mathcal{V}_{\lambda_J}(S_{h+k}))$. Therefore, a further integration over $\mathrm{SU}(d)$ gives

$$\begin{aligned}&\int_{\mathrm{SU}(d)}\mathrm{d}U\int_{\mathrm{SU}(2)}\mathrm{d}V\,UV^{\otimes(N+M)}\left[(|\psi\rangle\langle\psi|)^{\otimes N}\otimes(|\phi\rangle\langle\phi|)^{\otimes M}\right](V^\dagger U^\dagger)^{\otimes(N+M)}\\&=\sum_J P_{h,k}(J|c)\frac{I_{\lambda_J}}{\omega_{\lambda_J}^{(d)}}\otimes|J_{h,k}\rangle\langle J_{h,k}|\,,\end{aligned}\tag{4.53}$$

as stated. □

If one restricts the attention to $p_{\mu,\nu,\lambda}$, there is still an interesting class of quantities that can be estimated. Indeed, we already know that μ and ν converge to the spectra of respectively ρ and σ. Instead, if $h = pn$ and $k = (1-p)n$, λ converges to the spectrum of $p\rho + (1-p)\sigma$ when $n\to\infty$. An implicit argument for this fact is contained in Matthias Christandl's thesis [Chr06]. Here we give a tail bound for the estimate, for the more general case of convex combinations of an arbitrary number of states.

We define

$$\bar\rho := \left(\sum_{i=1}^m \frac{k_i\rho_i}{n}\right), \qquad \sigma := \bigotimes_{i=1}^m \rho_i^{\otimes k_i}\tag{4.54}$$

with the constraint $n = \sum_{i=1}^m k_i$.

Global weak Schur sampling applied to σ gives a partition λ with probability $p_\lambda(\{\rho_i,k_i\}) := \mathrm{tr}\left[\Pi_\lambda\bigotimes_{i=1}^m\rho_i^{\otimes k_i}\right]$. We find that the estimate $\bar\lambda := \frac{\lambda}{n}$ converges to the spectrum of $\sum_{i=1}^m\frac{k_i\rho_i}{n}$, which we indicate as $s(\{\rho_i,k_i\})$:

Theorem 4.2.7 (Spectral estimation of a convex combination of states: tail bound). *The probability distribution of the random variable $\bar\lambda$ obtained by weak Schur sampling on $\bigotimes_{i=1}^m\rho_i^{\otimes k_i}$ satisfies*

$$P(D(\bar\lambda/n||s(\{\rho_i,k_i\}))>\epsilon) := \sum_{\lambda:D(\bar\lambda/n||s(\{\rho_i,k_i\}))>\epsilon} p_\lambda(\{\rho_i,k_i\})\tag{4.55}$$

$$\leq (n+1)^{d(d-1)/2+m}\exp(-\epsilon n)\tag{4.56}$$

In particular, with $O((d^2+m)/\epsilon^2)\log\big((d^2+m)/\epsilon\big)\log(1/\delta))$ copies of ρ the estimate $\bar{\lambda}$ obtained from weak Schur sampling satisfies $D(\bar{\lambda}/n||s(\{\rho_i,k_i\}))\leq\epsilon$ with probability at least $1-\delta$. If all the states have rank less than r, one can substitute d with rm.

Proof. The following matrix inequality holds for any probability distribution p on $[m]$ with associated vector $\vec{p}$, expanding the product

$$\left(\sum_{i=1}^{m}\frac{k_i\rho_i}{n}\right)^{\otimes n}\geq\mathcal{B}(n,\vec{p},\vec{k})\frac{1}{n!}\sum_{\tau\in S_n}\mathbf{s}_n(\tau)\left(\bigotimes_{i=1}^{m}\rho_i^{\otimes k_i}\right)\mathbf{s}_n(\tau)^\dagger \tag{4.57}$$

where [CT05]

$$\mathcal{B}(n,\vec{p},\vec{k}):=\binom{n}{k_1,\,k_2,\,....,k_m}\prod_{i=1}^{l}p_i^{k_i}\geq\frac{1}{(n+1)^m}e^{-nD(\vec{k}/n||\vec{p})}, \tag{4.58}$$

so that we have

$$\mathcal{B}(n,\frac{\vec{k}}{n},\vec{k})\geq\frac{1}{(n+1)^m}. \tag{4.59}$$

It follows that

$$\left(\sum_{i=1}^{m}\frac{k_i\rho_i}{n}\right)^{\otimes n}\geq\frac{1}{(n+1)^m}\frac{1}{n!}\sum_{\tau\in S_n}\mathbf{s}_n(\tau)\left(\bigotimes_{i=1}^{m}\rho_i^{\otimes k_i}\right)\mathbf{s}_n(\tau)^\dagger. \tag{4.60}$$

Finally, we have that

$$\begin{aligned}p_\lambda(\{\rho_i,k_i\})&=\operatorname{tr}\left[\Pi_\lambda\bigotimes_{i=1}^{m}\rho_i^{\otimes k_i}\right]=\operatorname{tr}\left[\Pi_\lambda\frac{1}{n!}\sum_{\tau\in S_n}\mathbf{s}_n(\tau)\left(\bigotimes_{i=1}^{m}\rho_i^{\otimes k_i}\right)\mathbf{s}_n(\tau)^\dagger\right]=\\&\leq(n+1)^m\operatorname{tr}\left[\Pi_\lambda\left(\sum_{i=1}^{m}\frac{k_i\rho_i}{n}\right)^{\otimes n}\right]\\&=(n+1)^m\mathrm{SW}_{\vec{\rho}}^n(\lambda).\end{aligned} \tag{4.61}$$

and the inequality of the theorem follow by combining the last inequality with Theorem 4.2.4. □

A class of closeness measurements obtainable from the spectra of states and of their convex combination are the Jensen divergences [BR82; BH09], which include the Jensen-Shannon [Lin91; MLP05] divergence $QJS(\rho,\sigma):=S(\frac{\rho+\sigma}{2})-\frac{1}{2}(S(\rho)+S(\sigma))$ and the Hilbert-Schmidt distance $\operatorname{tr}\big[(\rho-\sigma)^2\big]=-\operatorname{tr}\big[(\rho+\sigma)^2\big]+2\operatorname{tr}\big[\rho^2\big]+2\operatorname{tr}\big[\sigma^2\big]$ as a particular case. These divergences possess metric properties [ES03; Lam+08; BH09]. The fact that Jensen-Shannon divergence and its Tsallis variation for order $\alpha\in(0,2]$ are square of metrics is a recent breakthrough [Sra21; Vir21]. The Jensen-Shannon divergence has

a special place, as it is a special case of the Holevo quantity, therefore it also has an operational significance. The possibility to estimate $-\operatorname{tr}\left[(\sum_i p_i\rho_i)^2\right] + \sum_i p_i \operatorname{tr}\left[(\rho_i)^2\right]$ with weak Schur sampling, from copies of states ρ_i will be central in the construction of the measurement in Chapter 7.

4.3 Pauli channels

In this section we concentrate on an important subclass of random unitary channels: Pauli channels, which describe random bit flip and phase flip errors in qubits. Explicit error correcting codes are devised to be able to correct Pauli errors [Got10]. We also consider the generalization of Pauli channels for qudits. As we mentioned in Chapter 3, these channels exhibit non-additive quantum and private information, and it remains an important open problem to characterize their quantum and private capacities. However, their particular symmetries make many computations accessible, and we will take advantage of this fact to find refined upper bounds on their quantum and private capacities in Chapter 8. The following treatment of generalized Pauli channels follows the phase-space description of finite dimensional quantum mechanics [Woo87; App05; Gro06; GE08; dBe13; GNW21].

4.3.1 Qubit Pauli group

The Pauli group of one qubit has the following elements:

$$\mathcal{P} := \{\pm I, \pm iI, \pm X, \pm iX, \pm Y, \pm iY, \pm Z, \pm iZ\}, \tag{4.62}$$

where

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and the product is given by the matrix multiplication. Note that all the elements of $\mathcal{P}$ are both unitary and hermitian. From $\mathcal{P}$, one can construct the Pauli group of n qubits obtained as the tensor product of n copies of $\mathcal{P}$, $\mathcal{P}^n := \{\otimes_{j=1}^n \omega_j | \omega_j \in \mathcal{P}\}$. For our purposes it suffices to consider $\mathsf{P}^n := \mathcal{P}^n / C^n$, the quotient of the Pauli group with its center $C^n := \{\pm I^{\otimes n}, \pm iI^{\otimes n}\}$. Each element of P^n can be identified by a pair of n bit-strings $x = (q, p)$ according to the definition

$$P_{(q,p)} := i^{-p\cdot q} \otimes_{j=1}^n Z^{p_j} X^{q_j}. \tag{4.63}$$

It is then immediate to see that for any two $x = (q, p), y = (q', p')$ we have $P_x P_y = (-1)^{\langle x,y\rangle} P_y P_x$, where

$$\langle x, y\rangle = p \cdot q' - q \cdot p' \mod 2\,. \tag{4.64}$$

In particular, two Pauli unitaries either commute or anticommute. Two very useful properties are

$$\frac{1}{2^n}\operatorname{Tr}[P_x] = \delta_{x,0},$$
$$\frac{1}{2^n}\sum_{(q,p)\in\mathbb{Z}_2^{2n}} P_{(q,p)}\rho P_{(q,p)} = \frac{I}{2^n} \qquad \forall\rho\in\Sigma(\mathbb{C}^{2n}). \tag{4.65}$$

4.3.2 Qudit Pauli group

The generalization to qudits is straightforward. The generalization of the Pauli group for one qudit, is the group $\mathcal{W}_d$ generated by τI ($\tau := e^{\frac{(d^2+1)\pi i}{d}}$), and the Weyl-Heisenberg operators X, Z acting as

$$X|j\rangle = |j+1\rangle \mod d, \qquad Z|j\rangle = e^{j\frac{2\pi i}{d}}|j\rangle \qquad j = 0,...,d-1. \tag{4.66}$$

In this case, these matrices are unitaries but not generally hermitian. For several qudits, likewise we set $\mathcal{W}_d^n := \{\otimes_{j=1}^n \omega_j | \omega_j \in \mathcal{W}_d\}$. The center of this group is still a set of multiples of the identity $C_d^n = \{\tau^j I^{\otimes n} : j = 0,...,D-1\}$, where $D = d$ if d is odd and $D = 2d$ if d is even; we define $\mathsf{W}_d^n := \mathcal{W}_d^n / C_d^n$. Each element of W_d^n can be identified by a pair of n Dit-strings $x = (q,p) \in \mathbb{Z}_D^{2n}$ according to the definition

$$W_{(q,p)} := e^{-\frac{(d^2+1)\pi i}{d}(p\cdot q)} \otimes_{j=1}^n Z^{p_j} X^{q_j}. \tag{4.67}$$

The commutation relations in these case read

$$W_x W_y = e^{\frac{2\pi i}{d}\langle x,y\rangle} W_y W_x, \tag{4.68}$$

where now

$$\langle x,y\rangle = p\cdot q' - q\cdot p' \mod D. \tag{4.69}$$

Moreover, for any $x, z \in \mathbb{Z}_D^{2n}$ we have

$$W_{x+dz} = (-1)^{(d+1)\langle x,z\rangle} W_x, \tag{4.70}$$

meaning that even when d is even we can restrict to $x \in \mathbb{Z}_d^{2n}$ if we are interested in listing all the Pauli unitary channels, as the $W_{x+d(1,1,...,1,1)}$ will give the same unitary channel of W_x.

As in the qubit case, we have the following important properties

$$\frac{1}{d^n}\operatorname{Tr}[W_x] = \delta_{x,0},$$
$$\frac{1}{d^n}\sum_{x\in\mathbb{Z}_d^{2n}} P_x\rho P_x = \frac{I}{d^n} \qquad \forall\rho\in\Sigma(\mathbb{C}^{dn}). \tag{4.71}$$

4.3.3 Pauli channels

Pauli channels are defined as convex combinations of Pauli unitaries, that is:

$$\Phi_{\mathbf{w}}[\rho] = \sum_{x \in \mathbb{Z}_d^{2n}} w_x W_x \rho W_x^\dagger, \tag{4.72}$$

where now it suffices to sum over $\mathbb{Z}_d^{2n}$ instead of $\mathbb{Z}_D^{2n}$ because of Eq. (4.70), and $\mathbf{w}(x) = w_x$. is a probability distribution over $\mathbb{Z}_d^{2n}$.

A crucial property of Pauli channels is that they commute with Pauli unitaries,

$$\Phi_{\mathbf{w}}[W_x \rho W_x^\dagger] = W_x \Phi_{\mathbf{w}}[\rho] W_x^\dagger. \tag{4.73}$$

An important Pauli channel on one qudit is the depolarizing channel, which is associated with a probability distribution $\mathbf{w}_p(x) = (0, ..., 0) = \frac{p}{d^2} + \delta_{0,x}(1-p)$.

$$\begin{aligned} \Phi_p^{(d)}[\rho] &:= (1 - \frac{d^2-1}{d^2}p)\rho + \frac{p}{d^2} \sum_{x \in \mathbb{Z}_d^2 \setminus \{0\}} W_x \rho W_x^\dagger \\ &= (1-p)\rho + p\frac{I}{d}. \end{aligned} \tag{4.74}$$

The depolarizing channel implements the mixture between the input signal and the completely mixed state.

4.4 Gaussian channels

In this section we introduce on Gaussian states and channels, highlighting on the group theoretic structures that describe them. Physically, Gaussian states coincide with thermal states of quadratic hamiltonians, which explain their importance in physics. In fact, they are the simplest class of continuous variable states to produce in the lab [Ser17]. Gaussian channels are channels that preserve Gaussian states. They are an important class of realistic noise models for electromagnetic waves in vacuum or fiber, with phase-insensitive Gaussian channels having a particular importance in this regard. While the classical capacity of phase-insensitive channels can be exactly computed [GHGP15], their quantum and private capacities are still open. We will bound the latter in Chapter 8. We follow [Ser17] in the presentation of Gaussian states and channels and [Hol19] for the properties of Gaussian channels for communication.

4.4.1 Displacements

In this section we consider infinite dimensional Hilbert spaces $L^2(\mathbb{R}^m)$, together with the representation of the Lie algebra of canonical commutation relations:

$$[\hat{x}_i, \hat{p}_j] = i\delta_{ij} \qquad [\hat{x}_i, \hat{x}_j] = 0 \quad [\hat{p}_i, \hat{p}_j] = 0 \tag{4.75}$$

where we set $\hbar = 1$, and, as we will do when denoting non-constant operators on $L^2(\mathbb{R}^m)$, we used the hat notation. This space describes m modes of the radiation field. The position and momentum operators, or quadratures, can be grouped into a vector

$$\hat{\mathbf{r}} := \begin{pmatrix} \hat{x}_1 \\ \hat{p}_1 \\ \vdots \\ \hat{x}_m \\ \hat{p}_m \end{pmatrix} \tag{4.76}$$

For which the commutation relations become

$$[\hat{r}_i, \hat{r}_j] = i\Omega_{ij} I, \tag{4.77}$$

where

$$\Omega := \bigoplus_{i=1}^{m} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.78}$$

The operators $\hat{\mathbf{r}}$ generate the *displacements*

$$\hat{D}(\mathbf{s}) = \exp\left(\mathrm{i}\mathbf{s}^{\mathsf{T}}\Omega\hat{\mathbf{r}}\right), \qquad \mathbf{s} \in \mathbb{R}^{2m}, \tag{4.79}$$

which act on $\hat{\mathbf{r}}$ by shifts:

$$\hat{D}(\mathbf{s})\hat{\mathbf{r}}\hat{D}(\mathbf{s})^{\dagger} = \hat{\mathbf{r}} + \mathbf{s} \tag{4.80}$$

The displacements are a subgroup of the unitary operators on $L^2(\mathbb{R}^m)$, satisfying

$$\hat{D}(\mathbf{s}_1 + \mathbf{s}_2) = \hat{D}(\mathbf{s}_1)\hat{D}(\mathbf{s}_2)e^{\mathrm{i}\mathbf{s}_1^{\mathsf{T}}\Omega\mathbf{s}_2/2}. \tag{4.81}$$

and

$$\mathrm{tr}\Big[\hat{D}(\mathbf{s}_1)\hat{D}(-\mathbf{s}_2)\Big] = \delta(\mathbf{s}_1 - \mathbf{s}_2). \tag{4.82}$$

In fact, the set of displacements constitutes an orthogonal complete operator set and allows the definition of the characteristic function of a trace class operator $\hat{\rho} \in \Sigma(L^2(\mathbb{R}^m))$ as [Ser17]

$$\phi_{\hat{\rho}}(\mathbf{s}) := \mathrm{tr}\Big[\hat{\rho}\hat{D}(-\mathbf{s})\Big], \tag{4.83}$$

which uniquely identifies $\hat{\rho}$.

4.4.2 Gaussian states

Gaussian states are those states whose characteristic function is Gaussian, i.e.

$$\phi_{\hat{\rho}}(\mathbf{s}) = \exp\left(-\frac{1}{4}\mathbf{s}^{\mathsf{T}}\Omega^{\mathsf{T}}V\Omega\mathbf{s} + \mathrm{i}\mathbf{s}^{\mathsf{T}}\Omega\bar{\mathbf{r}}\right), \tag{4.84}$$

with

$$\bar{\mathbf{r}} := \mathrm{tr}[\hat{\mathbf{r}}\hat{\rho}] \tag{4.85}$$

and (defining $\{A, B\} := AB + BA$

$$\boldsymbol{\sigma} := \mathrm{tr}\big[\{(\hat{\mathbf{r}} - \bar{\mathbf{r}}), (\hat{\mathbf{r}} - \bar{\mathbf{r}})^{\mathrm{T}}\}\hat{\rho}\big] \tag{4.86}$$

being the associated statistical mean vector and covariance matrix respectively. It can be shown that

$$\boldsymbol{\sigma} \geq \pm i\Omega, \qquad \boldsymbol{\sigma} > 0. \tag{4.87}$$

The creation and annihilation operators $\hat{a}_i^\dagger$, $\hat{a}_i$ are defined as

$$\hat{a}_i = \frac{\hat{x}_i + \hat{p}_i}{\sqrt{2}}, \tag{4.88}$$

They satisfy $[a_i^\dagger, a_j] = 1$ and can be grouped as vector

$$\hat{\mathbf{a}} := \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_m \end{pmatrix}. \tag{4.89}$$

On $L^2(\mathbb{R})$, the number operator is defined as

$$\hat{n} := \hat{a}^\dagger\hat{a}, \tag{4.90}$$

$\hat{n} \geq 0$ and its spectrum is $\mathbb{N}$, with multiplicities 1. The eigenvectors $|n\rangle$ form the Fock basis. The expectation value $\hat{n}$ is called the energy, and a thermal state of energy N is defined as

$$\hat{\rho}_N := \frac{1}{N+1} \sum_{i=1}^{n} \left(\frac{N}{N+1} \right)^i |i\rangle \langle i| \tag{4.91}$$

On $L^2(\mathbb{R}^m)$, we can define

$$\hat{n}_i := \hat{a}_i^\dagger a_i \tag{4.92}$$

and the total number operator is defined as:

$$\hat{N} := \sum_{1=1}^{m} \hat{a}_i^\dagger a_i = \sum_{1=1}^{m} \hat{n}_i. \tag{4.93}$$

$\hat{N} \geq 0$. The eigenvectors of $\hat{N}$ can be labeled as $\{|n_1, ..., n_m\rangle\}$ and they are also called Fock basis. We call m-mode thermal the states of $L^2(\mathbb{R}^m)$ states that can be written as product of thermal states of $L^2(\mathbb{R})$.

Writing the position and momentum operators in terms of the creation and annihilation operators, we can give an alternative parametrization of displacement operators, $D(\vec{\alpha}) = e^{\sum_i \alpha a_i - \alpha_i^* a_i^\dagger}$, where $\vec{\alpha}$ is a vector of complex variables. An important subset of Gaussian states are *coherent states*, which are obtained applying displacements to the vacuum state:

$$|\vec{\alpha}\rangle = D(\vec{\alpha}) |0\rangle . \tag{4.94}$$

Note that $|\vec{\alpha}\rangle$ are always product states. Coherent states form a resolution of the identity

$$\int \frac{d^2\vec{\alpha}}{\pi^m} |\vec{\alpha}\rangle\langle\vec{\alpha}| = I, \tag{4.95}$$

therefore they also define a POVM, which is called heterodyne measurement.

4.4.3 Symplectic transformations and the structure of Gaussian states

Gaussian channels are channels that map Gaussian states to Gaussian states, and *Gaussian unitaries* are unitary Gaussian channels, which form a subgroup of the unitary

operators on $L^2(\mathbb{R}^m)$. Gaussian unitary also include a representation of the symplectic group $\mathrm{Sp}(2m, \mathbb{R})$, defined as the $2m \times 2m$ invertible matrices S that satisfy

$$S\Omega S^{\mathsf{T}} = \Omega. \tag{4.96}$$

The operators $\hat{U}_S$ act on the quadrature implementing S on the quadratures

$$\hat{U}_S \hat{\mathbf{r}} \hat{U}_S^\dagger = S\hat{\mathbf{r}}, \tag{4.97}$$

and we call them symplectic unitaries.

They act on displacements as

$$\hat{U}_S \hat{D}(\mathbf{r}) \hat{U}_S^\dagger = D(S\hat{\mathbf{r}}). \tag{4.98}$$

The orthogonal symplectic matrices are a subgroup of $\mathrm{Sp}(2m, \mathbb{R})$ which is isomorphic to $\mathrm{U}(m)$. They are represented by energy-preserving Gaussian unitaries, i.e. commuting with $\hat{N}$. An important $U(1)$ subgroup of energy preserving unitary are those of the form $\hat{U}(\theta) = e^{-i\theta\hat{N}}$, generated by $\hat{N}$. They act on $\hat{\mathbf{r}}$ as

$$\hat{U}(\theta) \hat{\mathbf{r}} \hat{U}(\theta)^\dagger = R(\theta)^{\otimes m} \hat{\mathbf{r}}, \tag{4.99}$$

where

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{4.100}$$

A building block for non-energy preserving Gaussian unitaries are single mode squeezing, defined as

$$\hat{S}(r_i) = \exp\Big(\frac{r_i}{2}(\hat{a}_i^2 - \hat{a}_i^{\dagger 2})\Big), \tag{4.101}$$

and acting non-trivially only on the i-th quadrature operators as

$$\hat{S}(r_i) \hat{x}_i \hat{S}(r_i)^\dagger = e^{r_i} \hat{x}_i \qquad \hat{S}(r_i) \hat{p}_i \hat{S}(r_i)^\dagger = e^{-r_i} \hat{p}_i \tag{4.102}$$

for $r_i \in \mathbb{R}$. It can be proved that every symplectic unitary can be written as $\hat{U}_S = \hat{U}_u \hat{S}(r_1) ... \hat{S}(r_m) \hat{U}_v$ for some $u, v \in \mathrm{U}(m)$ and $r_i \in \mathbb{R}$, and that any Gaussian unitary can be written as a composition of a symplectic unitary and a displacement.

Moreover, any Gaussian state can be generated applying Gaussian unitaries to a suitable m-mode thermal state. The Gaussian unitary can be obtained by the symplectic diagonalization of the covariance matrix of the state, which is always possible by Williamson's theorem [Ser17]:

Theorem 4.4.1 (Williamson's theorem). *Let M be a $2n \times 2n$ strictly positive matrix. There exists $S \in \mathrm{Sp}(2\mathrm{n}, \mathbb{R})$ such that*

$$SMS^{\mathsf{T}} = D \qquad D = (\lambda_1, \lambda_1, \lambda_2, \lambda_2, ..., \lambda_n, \lambda_n) \tag{4.103}$$

with $d_i > 0$ for each i.

$(\lambda_1, ..., \lambda_n)$ are the symplectic eigenvalues of M, and they are also equal to the eigenvalues of $i\Omega M$, which appear in pairs with positive and negative sign. From Williamson's theorem, the symplectic eigenvalues of $\boldsymbol{\sigma}$ determine the spectrum of the associated state. In particular, the spectrum of a gaussian state always coincides with the spectrum of a product of thermal states, and the von Neumann entropy can be computed from the symplectic eigenvalues of $\boldsymbol{\sigma}$ as:

$$S(\hat{\rho}) = \sum_{i=1}^{m} h(\lambda_i) \qquad h(x) := \frac{x+1}{2} \log \frac{x+1}{2} - \frac{x-1}{2} \log \frac{x-1}{2}. \tag{4.104}$$

We conclude the section by defining *gauge-invariant* states.

Definition 4.4.1 (Gauge-invariant states). Gauge-invariant states are states which satisfy

$$\hat{U}(\theta)\hat{\rho}\hat{U}(\theta)^{\dagger} = \hat{\rho}. \tag{4.105}$$

Consequently, their first and second moments satisfy

$$\bar{\mathbf{r}} = 0 \tag{4.106}$$

$$R(\theta)^{\otimes m} \boldsymbol{\sigma} R(\theta)^{\otimes m\dagger} = \boldsymbol{\sigma}. \tag{4.107}$$

The non-Gaussian channel

$$\Phi_m[\hat{\rho}] := \frac{1}{2\pi} \int_0^{2\pi} d\theta \, \hat{U}(\theta)\hat{\rho}\hat{U}(\theta)^{\dagger}, \tag{4.108}$$

outputs gauge-invariant states.

4.4.4 Gaussian channels

Since coherent states are a resolution of the identity and they are Gaussian states, a Gaussian channel is identified by its action on Gaussian states. This action can be expressed as an action on the first and second moments. The following characterization can be proved [Ser17]:

Theorem 4.4.2 (Classification of Gaussian channels). *Any Gaussian channel $\mathcal{N}$ from m modes to m modes acts on the first and second moments of Gaussian states as*

$$\begin{aligned} \bar{\mathbf{r}} &\to X\bar{\mathbf{r}} + \mathbf{s} \\ \boldsymbol{\sigma} &\to X\boldsymbol{\sigma}X^{\mathsf{T}} + Y \end{aligned} \tag{4.109}$$

with X, Y real $m \times m$ matrices, $\mathbf{s}$ real vector of m real variables, and

$$Y + i\Omega \geq iX\Omega X^{\mathsf{T}}. \tag{4.110}$$

The action on a general state can then be expressed on the characteristic function as

$$\chi_{\hat{\rho}}(\mathbf{s}) \to \chi_{\mathcal{N}[\hat{\rho}]}(\mathbf{s}) = \chi_{\hat{\rho}}(\Omega^{\mathsf{T}}X\Omega\mathbf{s})e^{-\frac{1}{4}\mathbf{s}^{\mathsf{T}}\Omega^{\mathsf{T}}Y\Omega\mathbf{s}}. \tag{4.111}$$

An important subset of Gaussian channels are those having a peculiar symmetry property:

$$\mathcal{N}[e^{-i\theta\hat{N}}\rho e^{i\theta\hat{N}}] = e^{-i\theta\hat{N}}\mathcal{N}[\rho]e^{i\theta\hat{N}}, \tag{4.112}$$

which is called *gauge-covariance* [Hol19]. We will also refer to these channels as *phase-insensitive* channels. For these channels the Holevo information is additive [Gio+14; Hol19]. However, no such simplification is known for the coherent information. In particular, the quantum capacity of a class of single-mode phase-insensitive Gaussian channels describing attenuation, amplification and mixing with environmental noise is not known. We now introduce these channels, and we will give bounds on their quantum capacities in Chapter 8.

Thermal attenuators $\mathcal{E}_{\eta,N}$, describe attenuation of signals, according to a parameter $0 \leq \eta \leq 1$, in presence of a thermal environment of average photon number $N \geq 0$. The action on Gaussian states is defined by the mapping [CGH06]

$$\bar{\mathbf{r}} \xrightarrow{\mathcal{E}_{\eta,N}} \bar{\mathbf{r}}' = \sqrt{\eta}\bar{\mathbf{r}}\,, \tag{4.113}$$

$$\boldsymbol{\sigma} \xrightarrow{\mathcal{E}_{\eta,N}} \boldsymbol{\sigma}' = \eta\boldsymbol{\sigma} + (1-\eta)(2N+1)I_2\,, \tag{4.114}$$

with I_2 being the two dimensional identity. The physical interpretation of the channel is appreciated via the Stinespring representation [HG12; Wee+12; Ser17] which describes this transformation as a beam splitter coupling the system with an extra enviromental mode E initialized in a thermal (Gaussian) state, purified as a two mode squeezed state.

Indeed, labelling with A the system mode, and indicating with $\hat{\rho}_A$ its input state we can write

$$\mathcal{E}_{\eta,N}[\hat{\rho}_A] := \mathrm{Tr}_E[\hat{U}_\eta(\hat{\rho}_A \otimes (\hat{\tau}_N)_E)\hat{U}_\eta^\dagger]. \tag{4.115}$$

In this expression $\hat{U}_\eta$ ($0 \leq \eta \leq 1$) is an energy preserving unitary two-mode unitary operator that transforms $\hat{\mathbf{r}}$ according to

$$\hat{U}_\eta \hat{\mathbf{r}} \hat{U}_\eta^\dagger = \begin{pmatrix} \sqrt{\eta} I_2 & \sqrt{1-\eta} I_2 \\ -\sqrt{1-\eta} I_2 & \sqrt{\eta} I_2 \end{pmatrix} \hat{\mathbf{r}}. \tag{4.116}$$

The thermal state $\hat{\tau}_N$ entering in (4.115) is purified by a two-mode squeezed state $|\tau_N\rangle$, which has

$$\begin{aligned} \bar{\mathbf{r}}_{|\tau_N\rangle} &= (0,0,0,0) \\ \boldsymbol{\sigma}_{|\tau_N\rangle} &= \begin{pmatrix} (2N+1) I_2 & 2\sqrt{N(N+1)}\sigma_3 \\ 2\sqrt{N(N+1)}\sigma_3 & (2N+1) I_2 \end{pmatrix}, \end{aligned} \tag{4.117}$$

with

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.118}$$

We shall also consider single-mode thermal amplifiers $\Phi_{g,N}$. In this case the input state interacts with a thermal bath through a two mode squeezing operator with parameter $g \geq 1$ inducing the mapping

$$\bar{\mathbf{r}} \xrightarrow{\Phi_{g,N}} \bar{\mathbf{r}}' = \sqrt{g}\bar{\mathbf{r}}\,, \tag{4.119}$$

$$\boldsymbol{\sigma} \xrightarrow{\Phi_{g,N}} \boldsymbol{\sigma}' = g\boldsymbol{\sigma} + (g-1)(2N+1) I_2\,. \tag{4.120}$$

In this case we have the following physical representation

$$\Phi_{g,N}[\hat{\rho}_A] := \mathrm{Tr}_E[\hat{S}_g(\hat{\rho}_A \otimes (\hat{\tau}_N)_E)\hat{S}_g^\dagger], \tag{4.121}$$

where $\hat{S}_k$ is defined by

$$\hat{S}_g \hat{\mathbf{r}} \hat{S}_g^\dagger = \begin{pmatrix} \sqrt{g} I_2 & \sqrt{g-1}\sigma_3 \\ \sqrt{g-1}\sigma_3 & \sqrt{g} I_2 \end{pmatrix} \hat{\mathbf{r}}. \tag{4.122}$$

Finally, the single-mode additive gaussian noise channel Λ_β can be expressed as

$$\Lambda_\beta[\hat{\rho}] := \frac{\beta}{2\pi} \int_{\mathbb{R}^2} \mathrm{d}\mathbf{r} e^{-\frac{\beta}{2}\mathbf{r}^\intercal \mathbf{r}} \hat{D}(\mathbf{r}) \hat{\rho} \hat{D}(\mathbf{r})^\dagger, \tag{4.123}$$

where $\beta > 0$. The action on first and second moments is

$$\bar{\mathbf{r}} \xrightarrow{\Lambda_\beta} \bar{\mathbf{r}}' = \bar{\mathbf{r}} \,, \tag{4.124}$$

$$\boldsymbol{\sigma} \xrightarrow{\Lambda_\beta} \boldsymbol{\sigma}' = \boldsymbol{\sigma} + 2I_2/\beta \,, \tag{4.125}$$

from which we understand the name of this channel: it adds uniform noise to the state, with an intensity proportional to $1/\beta$, and it is the quantum counterpart of white noise.

Additive Gaussian noise is an example of *classical mixing channels*, [Ser17], characterized as follows. For any $n \times n$ square matrix $Y \geq 0$ with eigenvalues $\lambda_1, ..., \lambda_n$, let us indicate the support of Y as $S(Y)$, $\det_+ Y = \prod_{i:\lambda_i>0} \lambda_i$, and pseudoinverse of Y as $Y^{\ominus 1}$. Classical mixing channels have the form:

$$\Lambda_Y[\hat{\rho}] := \int_{S(Y)} \mathrm{d}\mathbf{r} \frac{e^{-\mathbf{r}^\intercal Y^{\ominus 1}\mathbf{r}}}{\sqrt{\pi}^{\dim S(Y)}\sqrt{\det_+ Y}} \hat{D}(\mathbf{r})\hat{\rho}\hat{D}(\mathbf{r})^\dagger \,. \tag{4.126}$$

Channels of this type are Gaussian and the action on the first and second moments is

$$\mathbf{m} \xrightarrow{\Lambda_Y} \mathbf{m}' = \mathbf{m} \,, \qquad V \xrightarrow{\Lambda_Y} V' = V + Y \,. \tag{4.127}$$

We will need the following characterization of the coherent information of degradable Gaussian channels, following Propositions 10.27 and 12.40 of [Hol19]

Theorem 4.4.3 (Coherent information of degradable Gaussian channels). *For a degradable Gaussian channel $\mathcal{N}$, $\mathcal{N}^c = \Gamma \circ \mathcal{N}$, with Γ a Gaussian channel, Gaussian states maximize the coherent information $I_c(\mathcal{N}, \hat{\rho})$ among states with fixed first and second moments. In particular, the coherent information is maximized on Gaussian states. If $\mathcal{N}$ is phase-insensitive, the coherent information is maximized on gauge-invariant Gaussian states.*

The last statement of the theorem is a direct consequence of the concavity of coherent information, Theorem 3.3.3. Indeed, for a phase-covariant degradable channel

$$I_c(\mathcal{N}, \Phi_m[\hat{\rho}]) \geq \frac{1}{2\pi}\int_0^{2\pi} d\theta I_c(\mathcal{N}, \hat{U}(\theta)\hat{\rho}\hat{U}(\theta)) = I_c(\mathcal{N}, \hat{\rho}), \tag{4.128}$$

the inequality due to concavity and the equality due to phase covariance, simply writing $I_c(\mathcal{N}, \hat{\rho}) = S(\mathcal{N}[\hat{\rho}]) - S(\mathcal{N} \otimes \mathcal{I}[|\rho\rangle\langle\rho|])$, where $|\rho\rangle$ is a purification of $\hat{\rho}$. This means that we can maximize among gauge-invariant states, and therefore gauge-invariant Gaussian states by the first part of Theorem 4.4.3. Moreover, a more general formulation of gauge-covariance is sufficient to have the equality, and we will need it in Chapter 8:

Definition 4.4.2 (Generalized gauge-covariance). A channel from m modes to m' modes has the generalized gauge-covariance property if for any θ

$$\mathcal{N}[\hat{U}(\theta)\hat{\rho}\hat{U}(\theta)^\dagger] = \hat{U}'(\theta)\mathcal{N}[\hat{\rho}]\hat{U}'(\theta)^\dagger \tag{4.129}$$

for some unitary operators $\hat{U}'(\theta)$.

From the same argument, degradable Gaussian channels which have generalized gauge-covariance have coherent information maximized on gauge-invariant Gaussian states. For channels of one mode input, gauge-invariant states coincide with thermal states. However, we need another step to perform the maximization over the energy of the thermal states. This actually follows from a very basic property of Gaussian channels, which can be immediately obtained from the characterization of Theorem 4.4.2:

$$\mathcal{N}[\hat{D}(\mathbf{s})\hat{\rho}\hat{D}(\mathbf{s})^\dagger] = \hat{D}(X\mathbf{s})\mathcal{N}[\hat{\rho}]\hat{D}(X\mathbf{s})^\dagger, \tag{4.130}$$

for some X real matrix. Using concavity and unitary invariance, and the expression Eq. 4.123 for the additive noise channel, we obtain

$$I_c(\mathcal{N}, \hat{\rho},) \leq I_c(\mathcal{N}, \Lambda_\beta[\hat{\rho}]). \tag{4.131}$$

Therefore, since additive noise channels increase arbitrarily the energy of the input state and preserve gauge-invariant states, the coherent information of the channel can be evaluated in the infinite energy limit.

Chapter 5

Learning machines for quantum state discrimination

This chapter is largely based on:

- Marco Fanizza, Andrea Mari, and Vittorio Giovannetti. "Optimal Universal Learning Machines for Quantum State Discrimination". In: *IEEE Transactions on Information Theory* 65.9 (2019), pp. 5931–5944. DOI: 10.1109/TIT.2019.2916646. arXiv: 1805.03477.

5.1 Introduction

In this chapter we compute optimal probabilities of error for several scenarios of programmable state discrimination [BH05], a particular setting of the quantum state discrimination task recalled in Chapter 2. As in the binary discrimination problem, we want to correctly classify a quantum state which is known to be initialized in one of two possible states. However, we assume that this task should be performed by a quantum machine which does not have at its disposal a complete classical description of the two template states, but it has access to n states prepared in the first template state and by m more states prepared in the second template state. The problem has been studied in both the unambiguous [FDF02; FD04; DB02; BH05; HHH05b; BFH06; ZYQ06; HB07; HB08; Sed+07; Sed+09; Bar+08; Sen+10; Col12; Zho14], and minimum error setting [HHH05a; GK10; Sen+10; AH11; Sen+12], as well as discrimination with an error margin [Sen+13]. We focus on the the minimum error case, extending previous results in a variety of scenarios. In the minimum error case, for large n and m one expects to recover the optimal probability of error for binary state discrimination, given by the Holevo-Helstrom theorem, Eq. (2.6). Therefore, the interest is to evaluate finite size

correction, when n and m are finite.

Programmable state discrimination attracted a renewed interest, as it represents a genuine instance of supervised quantum machine learning [DB18]. Indeed, machine learning (ML) studies how to instruct a computer to solve a specific task by feeding it with a collection of training data from which it could learn how to proceed. This approach finds applications in a variety of practical pattern recognition, decision and clustering problems where a definite classification of the various alternatives are not directly accessible [Vap98; SSBD13; LBH15]. There is a vast effort in trying to understand if quantum computer can be useful for classical machine learning tasks, and conversely to use classical machine learning algorithms for quantum problems, which we do not report here (see for example the reviews [Wit14; SSP14; Bia+17; Cil+17; DB18]). However, the problem we address in Chapter 6 is partially motivated by improving a quantum estimation subroutine of several proposed quantum algorithms for classical machine learning. Here we concentrate on a purely quantum generalization of machine learning tasks, which take quantum inputs as resources to perform some quantum task [ABG06]. Other examples of this paradigm are quantum template matching [SCJ01; SC02], learning how to perform a unitary transformation [Bis+10; SBZ19; MC19], change point detection [AH11; Sen+16; SCMT17; SMVMT18], programmable unitary discrimination [Hil+10; SSM18; SSM21], unsupervised classification of quantum states [Sen+19]. An interesting question which is brought up by these works is whether optimal strategies admit a semiclassical separation with a learning phase which measures part of the data (e.g. a training set) and outputs a candidate operation to apply on the remaining data (e.g. a test set). This is not the most general strategy, but can be optimal in some scenarios [Bis+10; Sen+12; MSW16].

Our analysis can be framed in the context of parameter estimation of covariant families, for which invariant/covariant measurement are optimal, and the optimal measurement in the worst case scenario can be also determined optimizing a Bayesian cost (see Theorems 4.2.1, 4.2.2). Indeed, we directly address the problem in the Bayesian scenario, giving the agent a prior distribution of the template states which is invariant under the action of the same unitary operator on both states. We instead assume that some information on the purities, on the distribution of the purities, and on the distance between the states are given. Note that these are all unitarily invariant properties. These scenarios naturally emerge when, for instance, the training and the target data are affected by some unavoidable deteriorating processes which the agent is aware of, or when the different templates are affected by uncertainties arising from the absence of a common, shared reference frame [BRS07].

The chapter has the following structure: the notation, the model and the results are introduced in Sec. 5.2. We derive the general form of optimal measurement in Sec. 5.3,

while our results are explicitly derived in three dedicated subsections. Specifically in Sec. 5.3.1 we study the case of an optimal universal machines which is trained to discriminate between two qubit density matrices of fixed but different purities. In Sec. 5.3.2 instead we focus on the case where the training data are two generic (possibly) mixed quantum systems. In Sec. 5.3.3 we discuss the scenario where the training data are pure with fixed relative overlap, but otherwise unknown, and the results of this analysis are valid for a generic dimension. Finally in Sec. 5.3.4 we compare the optimal machines that leads to the optimal probability thresholds for the three scenarios, commenting about their compatibility. Sec. 5.4 presents an implementation of optimal machines obtained by exploiting the QISKit software development kit [Abr+19].

5.2 The model

In a classical supervised learning classification problem a training set of labelled data is provided to the machine, and the machine produces a classifier which can be used to predict the label of new unlabelled data. In a probabilistic setting one can assume that the dataset, consisting of pairs (x, y) of data $x \in X$ and labels $y \in Y$ is sampled with a probability distribution $P : X \times Y \to [0, 1]$. A classifier is a labelling rule, which is not necessarily deterministic, which obeys some conditional distribution $p(y|x)$. We are interested in minimizing the average case probability of error: a good learning algorithm which has access to samples from P should obtain a classifier with a misclassification probability close to the optimal, as the training dataset becomes large, and with the fewest assumptions on the distribution P. The assumptions on P can also be described probabilistically as a prior probability distribution G over the possible P. Given this prior G, one can say that an algorithm is optimal for a training set of fixed size if it attains the lowest probability on average. This average is done over all the possible distributions P, assuming they are distributed according to G. A straightforward way to generalise classical probabilistic task is to substitute probability distributions with quantum states: in the problem considered in this chapter, the conditional probabilities $P(X|Y)$ are replaced with quantum states, distributed according to a classical prior.

In the general case considered in the literature, the agent is provided with n_1 copies of ρ_1, n_2 copies of ρ_2, m copies of the test state. Calling A the n d-level systems initialized in ρ_1, B the m d-level system initialized in ρ_2, and X the system of the state to classify, we have two possible alternatives:

- $\rho_X = \rho_1$, $\tau_1^{n_1,n_2,m} := {\rho_1^{\otimes n_1}}_A \otimes {\rho_1^{\otimes m}}_X \otimes {\rho_2^{\otimes n_2}}_B$,
- $\rho_X = \rho_2$, $\tau_2^{n_1,n_2,m} := {\rho_1^{\otimes n_1}}_A \otimes {\rho_2^{\otimes m}}_X \otimes {\rho_2^{\otimes n_2}}_B$.

We focus our attention to the case $n_1 = n_2$, and $m = 1$ with one copy of the quantum

state to be tested available. Moreover, the state can be initialized in one of the two alternatives ρ_1 and ρ_2 with equal probability 1/2. We will thus avoid writing superscripts on $\tau_{1/2}^{n_1,n_2,m}$.

Chosen a two-outcome POVM $\mathcal{M} \equiv \{E_1, E_2\}$ that acts globally on the full system AXB, the average probability of error reads:

$$P_{err}^{(n)} = \int d\mu(\rho_1, \rho_2) \frac{\mathrm{Tr}[\tau_1 E_2] + \mathrm{Tr}[\tau_2 E_1]}{2}, \tag{5.1}$$

where $\mathrm{Tr}[\tau_i E_j]$ is the probability of finding outcome i when the state is in the j-th configuration, while $d\mu(\rho_1, \rho_2)$ is a probability measure that gauges the initial ignorance of the agent about ρ_1 and ρ_2. Exploiting then the completeness relation of $\mathcal{M}$ this can be finally recast into

$$P_{err}^{(n)} = \frac{1}{2} - \frac{1}{4}\mathrm{Tr}[\Theta(E_1 - E_2)] \,, \tag{5.2}$$

where Θ is the trace-null, Hermitian operator

$$\Theta = \alpha^{(n)} - \beta^{(n)} \,, \tag{5.3}$$

given by the difference between the following density matrices of AXB,

$$\begin{aligned} \alpha^{(n)} &\equiv \int d\mu(\rho_1, \rho_2) {\rho_1^{\otimes n}}_A \otimes \rho_{1X} \otimes {\rho_2^{\otimes n}}_B, \\ \beta^{(n)} &\equiv \int d\mu(\rho_1, \rho_2) {\rho_1^{\otimes n}}_A \otimes \rho_{2X} \otimes {\rho_2^{\otimes n}}_B. \end{aligned} \tag{5.4}$$

The Holevo-Helstrom theorem, Eq. (2.6) applies to this minimization problem by seeing it as a binary state discrimination between two average states $\alpha^{(n)}$ and $\beta^{(n)}$. The minimum in Eq. (5.2) can be obtained by choosing an optimal POVM $\mathcal{M}$ which has components E_1, E_2 respectively projecting on the positive and the negative eigenspaces of Θ, giving:

$$P_{err,min}^{(n)} = \frac{1}{2} - \frac{1}{4}\|\Theta\|_1 \,. \tag{5.5}$$

Some general properties of $P_{err,min}^{(n)}$ can be determined by simple arguments. First of all since the agent can always discard part of the ancillary states before attempting to identify Q, by data processing, for all possible choices of the measure $d\mu(\rho_1, \rho_2)$, $P_{err,min}^{(n)}$ has to fulfil the inequality

$$P_{err,min}^{(n)} \leq \frac{1}{2} - \frac{1}{4}\| \int d\mu(\rho_1, \rho_2)(\rho_1 - \rho_2)\|_1 \,, \tag{5.6}$$

and being a decreasing function of n, i.e.

$$P_{err,min}^{(n)} \geq P_{err,min}^{(n+1)} \ . \tag{5.7}$$

Furthermore, by exploiting the convexity of the trace-norm, the following lower bound can be established

$$P_{err,min}^{(n)} \geq \frac{1}{2} - \frac{1}{4} \int d\mu(\rho_1, \rho_2) \|\rho_1 - \rho_2\|_1 =: \bar{P}_H \ , \tag{5.8}$$

for all n integers. The term on the right-hand-side of this inequality corresponds to the average Helstrom error probability $\bar{P}_H$, i.e. the average minimum error probability obtainable with a full classical description of the template states: under this condition in fact, for each couple of density matrices ρ_1 and ρ_2, the agent can taylor a specific POVM on X that it is optimized to distinguish them. Invoking a full tomographic reconstruction of ρ_1 and ρ_2 (assuming ρ_1 and ρ_2 are states of a finite-dimensional Hilbert space), the gap between $P_{err,min}^{(n)}$ and $\bar{P}_H$ (optimal excess risk function [Sen+10]), can be shown to nullify in the asymptotic regime $n \to \infty$, i.e.

$$\lim_{n\to\infty} P_{err,min}^{(n)} = \frac{1}{2} - \frac{1}{4} \int d\mu(\rho_1, \rho_2) \|\rho_1 - \rho_2\|_1. \tag{5.9}$$

Apart from the above results explicit expressions for $P_{err,min}^{(n)}$ are known only for a limited set of configurations. In ref. [HHH05a] the authors focus on the case where both ρ_1 and ρ_2 are pure states of a finite dimensional Hilbert space, extracted independently from the Haar measure, and give the solution of the optimal probability of error as a finite sum. In [Sen+10], the authors provide the formal solution for the minimum error case under the assumption that ρ_1 and ρ_2 are qubit density matrices having the same assigned purity, and under several priors on the purity. They also compute the optimal probability of error for pure random qubit states, for arbitrary m, n_1, n_2. They compute several limits and asymptotic expansions of the probability of error. For qubit mixed states, $m = 1$, $n_1 = n_2 = n \to \infty$, they provide an asymptotic expansion of the error probability at order $O(\frac{1}{n})$. For pure qubit states, they provide unambiguous and minimum error probabilities in the limits $m \to \infty$, $n_1 = n_2$ fixed, an asymptotic expansion of the minimum error probabilities for m fixed, $n_1 = n_2 = n \to \infty$ at $O(\frac{1}{n})$, and an asymptotic expansion of the minimum error probabilities for $m = n_1 = n_2 = n \to \infty$, at order $O(\frac{1}{n})$. In [GK10] the optimal performances for qubit states of separable strategies, where the training data are used to estimate a candidate POVM, is obtained with the formalism of local asymptotic normality. [Sen+12] shows that an optimal strategy for the case of pure qubits $n_1 = n_2$, $m = 1$ is of the separable form, and that separable POVMs are asymptotically optimal for mixed qubits. It is not clear if these results survive for higher dimensional models such as the one we consider for pure states. The case of pure states in dimension d with

arbitrary n_1, n_2, m and fixed overlap has been formally solved in [AH11], under the name of change point problem. In this paper the authors also evaluate the error exponent of the optimal probability of error in the limit $n_1 = n_2 = \alpha m \to \infty$, where $\alpha > 0$ is a fixed proportionality constant.

Our contribution, in the setting $m = 1$, $n_1 = n_2 = n \to \infty$, establishes further asymptotic expansions of $P^{(n)}_{err,min}$ at large n. We do not have a rigorous estimation of the remainder terms in the asymptotic expansions, but we can verify that the proposed expansions are numerically accurate.

Specifically, we consider the following cases:

i) ρ_1 and ρ_2 qubit states having different assigned purities but being otherwise arbitrary;

ii) ρ_1 and ρ_2 being completely arbitrary (not necessarily pure) qubit states;

iii) ρ_1 and ρ_2 being arbitrary pure qudit states having assigned overlap.

Let us fix some notation, before stating the results. Adopting the Bloch sphere representation we express the template states ρ_1 and ρ_2 in terms of their associated Bloch vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ via the mapping

$$\rho_1 = \frac{I + \mathbf{r}_1 \cdot \sigma}{2}, \qquad \rho_2 = \frac{I + \mathbf{r}_2 \cdot \sigma}{2}\ , \tag{5.10}$$

with $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ being the Pauli vector. Any pair of qubit states can then be written as

$$\rho_1 = U_1 \frac{I + r'_1 \sigma_z}{2} U_1^\dagger, \qquad \rho_2 = U_2 \frac{I + r'_2 \sigma_z}{2} U_2^\dagger\ , \tag{5.11}$$

for some $U_1, U_2 \in \mathrm{SU}(2)$. Our prior distribution $\mu(\rho_1, \rho_2)$ can then be written as a distribution over U_1, U_2, r'_1, r'_2. Our different scenarios are distinguished by the prior distribution. As usual, we denote dU the Haar measure of SU(2). $d\mu(r') = 3r'^2 dr'$ is the hard sphere prior, while $\delta(r' - r)$ constrains the purity to a fixed value and $\delta(U - U_0)$ constrains U to be U_0. In case iii) we use the following parametrization for pairs of states at fixed overlap: defining $U_0 = \exp(-i\sigma_y(\pi - \theta)/2)$ we have that any pair of states $U_1 |0\rangle$ and $U_1 U_0 |0\rangle$ with U_1 unitary have overlap $|\langle 0| U_0 |0\rangle|^2 = \sin^2 \theta/2$ and that any pair of states $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$ with overlap $|\langle\psi_1|\psi_2\rangle|^2 = \sin^2\theta/2$ can be written as $|\psi_1\rangle\langle\psi_1| = U_1 |0\rangle\langle 0| U_1^\dagger$ and $|\psi_2\rangle\langle\psi_2| = U_1 U_0 |0\rangle\langle 0| U_0^\dagger U_1^\dagger$ for some unitary U_1.

We establish the following results:

Proposition 5.2.1 (Scenario i)). *For a prior distribution*

$$d\mu(\rho_1, \rho_2) = dU_1 dU_2 \delta(r'_1 - r_1)\delta(r'_2 - r_2) dr'_1 dr'_2, \tag{5.12}$$

the minimum error probability is

$$\begin{aligned} P_{err,min}^{(n\gg 1)} &= \frac{1}{2} - \frac{1}{24}\frac{(r_1+r_2)^3 - |r_1-r_2|^3}{r_1 r_2} + \frac{5}{24\,n}\frac{(r_1+r_2)^3 + |r_1-r_2|^3}{r_1^2 r_2^2} \\ &- \frac{1}{24\,n}\frac{(r_1+r_2)^5 - |r_1-r_2|^5}{r_1^3 r_2^3} + o\left(\frac{1}{n}\right). \end{aligned} \tag{5.13}$$

Proposition 5.2.2 (Scenario ii)). *For a prior distribution*

$$d\mu(\rho_1,\rho_2) = dU_1 dU_2 d\mu(r_1') d\mu(r_2') dr_1' dr_2', \tag{5.14}$$

the minimum error probability is

$$P_{err,min}^{(n\gg 1)} = \frac{17}{70} + \frac{18}{35n} + o\left(\frac{1}{n}\right) . \tag{5.15}$$

Proposition 5.2.3 (Scenario iii)). *For a prior distribution*

$$d\mu(\rho_1,\rho_2) = dU_1 \delta(U_2 - U_1 U_0)\delta(r_1'-1)\delta(r_2'-1), \qquad U_0 = \exp(-i\sigma_y(\pi-\theta)/2), \tag{5.16}$$

the minimum error probability is

$$P_{err,min,d}^{(n\gg 1)} = \frac{1}{2}\left(1 - |\cos\tfrac{\theta}{2}|\right) + \tfrac{3+\cos\theta}{8\sqrt{2}\sqrt{1+\cos\theta}}\frac{1}{n} + \tfrac{1-60\cos\theta-5\cos 2\theta}{128\sqrt{2}(1+\cos\theta)^{3/2}}\frac{1}{n^2} + o\left(\frac{1}{n^2}\right).$$

The result is valid also for arbitrary finite dimension d, when dU_1 is substituted by the Haar measure of $\mathrm{SU}(d)$.

5.3 Symmetries of average states

We now set out to determine the eigenvalues $\{\lambda_\ell\}_\ell$ of the operator Θ defined in Eq. (5.3), in order to rewrite Eq. (5.5) as

$$P_{err,min}^{(n)} = \frac{1}{2}\left(1 - \sum_\ell{}^{+}\lambda_\ell\right) , \tag{5.17}$$

the sum being restricted on the positive part of the spectrum. This derivation overlaps with those in [HHH05a; Sen+10; AH11].

Since $\Theta = \alpha^{(n)} - \beta^{(n)}$, we first focus on diagonalizing $\alpha^{(n)}$ and $\beta^{(n)}$ exploiting their symmetry properties. Then, by noticing the common symmetries of $\alpha^{(n)}$ and $\beta^{(n)}$, one can reduce the problem to a diagonalization of 2×2 matrices. This procedure is an explicit derivation of the fact that we can choose an invariant measurement for minimizing an

average case cost, as the family of states τ_1 and τ_2 are covariant with respect to the action of $\mathrm{SU}(2) \times S_n \times S_n$ in the sense considered by Theorem 4.2.2. Indeed, the average states $\alpha^{(n)}$ and $\beta^{(n)}$, and therefore Θ, are invariant under the action of the same group:

- $[\Theta, \mathbf{s}_n^{(A)}(\tau)] = [\Theta, \mathbf{s}_n^{(B)}(\tau)] = 0$ for every $\mathbf{s}_n^{(A)}(\tau), \mathbf{s}_n^{(B)}(\tau)$ qubit permutations acting respectively on $\mathcal{H}_A$ and $\mathcal{H}_B$.
- $[\Theta, U^{\otimes 2n+1}] = 0$ for every $U \in \mathrm{SU}(2)$.

By Schur's lemma, Θ will be block diagonal according to the decomposition into copies of irreducible representations of $\mathrm{SU}(2) \times S_n \times S_n$.

By Schur-Weyl duality, reviewed in Sec. 4.2.2, restricting to $d = 2$, we can decompose the Hilbert space of n qubits as follows

$$\mathcal{H} = \bigoplus_{n/2-\lfloor n/2 \rfloor \leq j \leq n/2} (\mathcal{U}_j \otimes \mathcal{V}_{j,n}), \tag{5.18}$$

where $\mathcal{U}_j$ hosts an irreducible representation of $\mathrm{SU}(2)$ with total angular momentum j, and $\mathcal{V}_{j,n}$ hosts an irreducible representation of S_n associated to the partition $(\frac{n}{2}+j, \frac{n}{2}-j, 0, ..., 0)$. The dimension of $\mathcal{U}_j$ is $2j+1$ and the dimension of $\mathcal{V}_{j,n}$ is

$$\omega(j,n) \quad = \quad \frac{n!\,(2j+1)}{\left(\frac{n-2j}{2}\right)!\left(\frac{n+2j}{2}+1\right)!}\,. \tag{5.19}$$

The Hilbert space of the systems AXB can be thus decomposed as

$$\mathcal{H}_{AXB} = \bigoplus_s \bigoplus_t \mathcal{U}_s^A \otimes \mathcal{U}_{1/2}^X \otimes \mathcal{U}_t^B \otimes \mathcal{V}_{s,n}^A \otimes \mathcal{V}_{t,n}^B, \tag{5.20}$$

Where $\mathrm{SU}(2)$ acts non-trivially on $\mathcal{U}_s^A \otimes \mathcal{U}_{1/2}^X \otimes \mathcal{U}_t^B$, as a product of the associated irreducible representations, the permutations of the qubits in system A act non-trivially on $\mathcal{V}_{s,n}^A$, and the permutations on qubits in system B act non-trivially on $\mathcal{V}_{t,n}^B$.

Moreover, using the recoupling theory of $\mathrm{SU}(2)$, recalled in Sec. 4.1.5 we can couple $\mathcal{U}_s^A \otimes \mathcal{U}_{1/2}^X$ obtaining $\mathcal{U}_s^A \otimes \mathcal{U}_{1/2}^X = \mathcal{U}_{s+1/2}^{AX} \bigoplus \mathcal{U}_{s-1/2}^{AX}$, and coupling with $\mathcal{U}_t^B$ we obtain a basis of $\mathcal{U}_s^A \otimes \mathcal{U}_{1/2}^X \otimes \mathcal{U}_t^B = \bigoplus_q \mathcal{U}_q^{AXB} \otimes \mathbb{C}^{g(s,t,q)}$, where $g(s,t,q)$ are multiplicities to determine, which can be either zero, one or two. A basis of the whole space can be indexed as

$$\{\big|s' = s \pm 1/2, t; q, m\big\rangle_{i,k}\} \tag{5.21}$$

where q and m are a labels for a basis of $\mathcal{U}_q^{AXB}$: q is a total angular momentum label and m is an eigenvalue for J_{tot}^z, $\vec{J}_{tot}^2 |s' = s \pm 1/2, t; q, m\rangle = q(q+1) |s' = s \pm 1/2, t; q, m\rangle$, $J_{tot}^z |s' = s \pm 1/2, t; q, m\rangle = m |s' = s \pm 1/2, t; q, m\rangle$. The label q span from $||t-s| - 1/2|$ to $t + s + 1/2$, while m runs from q to $-q$. We stress that in the above construction, and in the remainder of this section, it is implicit assumed that $|s \pm 1/2, t; q, m\rangle_{i,k}$ is null whenever the parameters s, t and q do not fit the necessary angular momentum selection rules. i, k are additional labels associated to two bases of $\mathcal{V}_{s,n}^A$ and $\mathcal{V}_{t,n}^B$. This allows us to identify four different scenarios:

a) $q = s + t + \frac{1}{2}$;

b) $q = t - s - \frac{1}{2}$ and $t > s$;

c) $q = s - t - \frac{1}{2}$ and $s > t$;

d) all s, t, q fitting the selection rules which are not included in the previous cases.

In the first three cases, only one of the elements of the pairs $\{|s \pm 1/2, t; q, m\rangle\}$ survives: specifically the $s + 1/2$ element for *a)* and *b)*, while the $s - 1/2$ element for *c)*. This establishes the multiplicities $g(s, t, q)$, and a basis of the multiplicity space is indexed by the allowed values of $s \pm 1/2$ and t at fixed q.

The symmetry under $\mathrm{SU}(2) \times S_n \times S_n$ forces Θ to be block diagonalized as

$$\Theta = \bigoplus_{q,s,t} I_q^{AXB} \otimes \Theta^{(s,t,q)} \otimes I_{s,n}^A \otimes I_{t,n}^B, \tag{5.22}$$

where now $I_{s,n}^A \otimes I_{t,n}^B$ is the identity operator on $V_{s,n}^A \otimes \mathcal{V}_{t,n}^B$, I_q^{AXB} is the identity operator on $\mathcal{U}_q^{AXB}$. The operator $\Theta^{(s,t,q)}$ is thus acting on the multiplicity spaces $\mathbb{C}^{g(s,t,q)}$, which have dimension at most 2. In cases *a)* and *b)* and *c)*, $\Theta^{(s,t,q)}$ is a 1×1 matrix, with eigenvalues $\lambda_{s,t,q}^{(n)}$ that we can formally compute as

$$\Theta_{++}^{(s,t,q)} =_{i,k} \langle s + 1/2, t; q, m| \, \Theta \, |s + 1/2, t; q, m\rangle_{i,k} \ , \tag{5.23}$$

for the case cases *a)* and *b)*, and

$$\Theta_{--}^{(s,t,q)} =_{i,k} \langle s - 1/2, t; q, m| \, \Theta \, |s - 1/2, t; q, m\rangle_{i,k} \ , \tag{5.24}$$

for the *c)* case. The corresponding multiplicity is determined instead by Eq. (5.22), and it evaluates to

$$M_{s,t,q}^{(n)} = (2q + 1) \, \omega(s, n) \, \omega(t, n) \ . \tag{5.25}$$

In the scenario *d)* instead both the elements of the couple $\{|s \pm 1/2, t; q, m\rangle\}$ survive and the symmetry of the problem forces Θ to be described by 2×2 block diagonal terms

$\Theta|_{i,k}^{s,t,q,m}$ of the form,

$$\Theta|_{i,k}^{s,t,q,m} \equiv \begin{bmatrix} \Theta_{++}^{(s,t,q)} & \Theta_{+-}^{(s,t,q)} \\ \Theta_{-+}^{(s,t,q)} & \Theta_{--}^{(s,t,q)} \end{bmatrix} , \tag{5.26}$$

with $\Theta_{++}^{(s,t,q)}$ and $\Theta_{--}^{(s,t,q)}$ as in Eq. (5.23) and Eq. (5.24) and with

$$\Theta_{+-}^{(s,t,q)} = [\Theta_{-+}^{(s,t,q)}]^* = \tag{5.27}$$
$$_{i,k}\left\langle s+1/2,t;q,m\right| \Theta \left|s-1/2,t;q,m\right\rangle_{i,k} \quad .$$

Accordingly we get a further set of eigenvalues identified with the functions

$$\lambda_{s,t,q}^{(n)}(\pm) = \left(\frac{\Theta_{--}^{(s,t,q)}+\Theta_{++}^{(s,t,q)}}{2}\right) \pm \sqrt{\left(\frac{\Theta_{--}^{(s,t,q)}-\Theta_{++}^{(s,t,q)}}{2}\right)^2 + |\Theta_{+-}^{(s,t,q)}|^2} , \tag{5.28}$$

again characterized by multiplicities $M_{s,t,q}^{(n)}$ defined as in Eq. (5.25). The corresponding eigenvectors are instead provided by the superpositions

$$|\psi_{s,t;q,m}^{(\pm)}\rangle_{i,k} \;=\; A^{(s,t,q)} \left|s+1/2,t;q,m\right\rangle_{i,k} + B_{s,t,q}^{(n)}(\pm) \left|s-1/2,t;q,m\right\rangle_{i,k} \ , \tag{5.29}$$

with amplitudes $A^{(s,t,q)} = \Theta_{+-}^{(s,t,q)}$ and

$$B_{s,t,q}(\pm) = \left(\frac{\Theta_{--}^{(s,t,q)}-\Theta_{++}^{(s,t,q)}}{2}\right) \pm \sqrt{\left(\frac{\Theta_{--}^{(s,t,q)}-\Theta_{++}^{(s,t,q)}}{2}\right)^2 + |\Theta_{+-}^{(s,t,q)}|^2} , \tag{5.30}$$

which, for ease of notation we present in a non-normalized form.

5.3.1 Scenario i): mixed states with fixed purity

In this scenario we can write

$$\rho_1 = U_1 \frac{I + \mathbf{r}_1 \cdot \sigma}{2} U_1^\dagger, \qquad \rho_2 = U_2 \frac{I + \mathbf{r}_2 \cdot \sigma}{2} U_2^\dagger \, , \tag{5.31}$$

where $r_1 \equiv |\mathbf{r}_1|$ and $r_2 \equiv |\mathbf{r}_2|$ are constant, and average over all possible orientations of $\mathbf{r}_1$ and $\mathbf{r}_2$ with two independent copies of the Haar measure.

Accordingly we rewrite Eq. (5.4) as

$$\alpha^{(n)} \;=\; \int dU_1 \left(U_1 \rho_1 U_1^\dagger\right)^{\otimes n+1} \otimes \int dU_2 \left(U_2 \rho_2 U_2^\dagger\right)^{\otimes n} , \tag{5.32}$$

$$\beta^{(n)} \;=\; \int dU_1 \left(U_1 \rho_1 U_1^\dagger\right)^{\otimes n} \otimes \int dU_2 \left(U_2 \rho_2 U_2^\dagger\right)^{\otimes n+1} . \tag{5.33}$$

With this choice both $\alpha^{(n)}$ and $\beta^{(n)}$, as well as their difference Θ, become explicitly invariant under unitaries acting in the same way on each qubit, i.e. $U^{\otimes 2n+1}$.

In addition of the invariance under $U^{\otimes 2n+1}$, $\alpha^{(n)}$ and $\beta^{(n)}$ are also invariant under separate rotations of partitions of the system, in particular AX/B for $\alpha^{(n)}$ and A/XB for $\beta^{(n)}$.

By Schur-Weyl duality, for ρ with Bloch vector of modulus r one has the identity

$$\int dU \left(U\rho U^\dagger\right)^{\otimes n} = \bigoplus_j f_j^{(n)}(r) I_j \otimes I_{j,n}, \tag{5.34}$$

where I_j is the identity opertor on $\mathcal{U}_j$ and $I_{j,n}$ is the identity operator on $\mathcal{V}_{j,n}$

We compute $f_j^{(n)}$ in Appendix A.2, obtaining

$$f_j^{(n)}(r) \;=\; \frac{1}{2j+1}\left(\frac{1-r^2}{4}\right)^{\frac{n}{2}-j} \frac{\left(\frac{1+r}{2}\right)^{2j+1} - \left(\frac{1-r}{2}\right)^{2j+1}}{r}\,, \tag{5.35}$$

This allows us to cast Eq. (5.32) in the following form

$$\alpha^{(n)} = \bigoplus_{s',t} f_{s'}^{(n+1)}(r_1) f_t^{(n)}(r_2) {I_{s'}}^{AX} \otimes {I_t}^B \otimes {I_{s,n}}^{AX} \otimes {I_{t,n}}^B\,, \tag{5.36}$$

where ${I_{s'}}^{AX} \otimes {I_t}^B \otimes {I_{s,n}}^{AX} \otimes {I_{t,n}}^B$ is the identity operator on ${\mathcal{U}_{s'}}^{AX} \otimes {\mathcal{U}_t}^B \otimes {\mathcal{V}_{s,n}}^{AX} \otimes {\mathcal{V}_{t,n}}^B$, and ${\mathcal{U}_{s'}}^{AX}$ is the irreducible representations of dimension $2s'+1$ in $\mathcal{U}_{(s)}^A \otimes \mathcal{U}_{(1/2)}^X$, with $s' = s \pm 1/2$. Adopting the basis $\{|s' = s \pm 1/2, t; q, m\rangle_{i,k}\}_{q,m}$, defined in Eq. (5.21) we can then use Eq. (5.36) to decompose $\alpha^{(n)}$ as a direct sum of independent contributions acting on the subspaces $\mathcal{H}_{A_i X B_k}^{(s,t)}$, i.e.

$$\alpha^{(n)} = \bigoplus_{s,t} \bigoplus_{i,k} \left(\bigoplus_{q,m} \alpha^{(n)}|_{i,k}^{s,t,q,m}\right)\,, \tag{5.37}$$

where, for each s, t, i and k we exploited the fact that each term further decompose into a direct sum of either 1×1 or 2×2 blocks of the form

$$\begin{aligned}\alpha^{(n)}|_{i,k}^{s,t,q,m} &= f_{s+1/2}^{(n+1)}(r_1) f_t^{(n)}(r_2) \left|s+1/2, t; q, m\right\rangle_{i,k} \left\langle s+1/2, t; q, m\right| \\ &+ f_{s-1/2}^{(n+1)}(r_1) f_t^{(n)}(r_2) \left|s-1/2, t; q, m\right\rangle_{i,k} \left\langle s-1/2, t; q, m\right|\,,\end{aligned} \tag{5.38}$$

where as already mentioned it is implicit assumed that the vectors $|s \pm 1/2, t; q, m\rangle_{i,k}$ nullify whenever the parameters s, t and q do not fit the angular momentum selection rules. In a similar fashion we have that

$$\beta^{(n)} = \bigoplus_{s,t'} f_s^{(n)}(r_1) f_{t'}^{(n+1)}(r_2) {I_s}^A \otimes {I_{t'}}^{XB} \otimes {I_{s,n}}^A \otimes {I_{t',n}}^{XB}\,, \tag{5.39}$$

where ${I_s}^A \otimes {I_{t'}}^{XB} \otimes {I_{s,n}}^A \otimes {I_{t',n}}^{XB}$ is the identity operator on ${\mathcal{U}_s}^A \otimes {\mathcal{U}_{t'}}^{XB} \otimes {\mathcal{V}_{s,n}}^A \otimes {\mathcal{V}_{t',n}}^{XB}$, and ${\mathcal{U}_{t'}}^{XB}$ is the irreducible representations of dimension $2t'+1$ in $\mathcal{U}^X_{(1/2)} \otimes \mathcal{U}^B_{(t)}$, with $t' = t \pm 1/2$. Again this yields the following decomposition

$$\beta^{(n)} = \bigoplus_{s,t} \bigoplus_{i,k} \left(\bigoplus_{q,m} \beta^{(n)}|^{s,t,q,m}_{i,k} \right) , \tag{5.40}$$

where now

$$\begin{aligned} \beta^{(n)}|^{s,t,q,m}_{i,k} &= f^{(n)}_s(r_1) f^{(n+1)}_{t+1/2}(r_2) \left|s, t+1/2; q, m\right\rangle_{i,k} \left\langle s, t+1/2; q, m\right| \\ &+ f^{(n)}_s(r_1) f^{(n+1)}_{t-1/2}(r_2) \left|s, t-1/2; q, m\right\rangle_{i,k} \left\langle s, t-1/2; q, m\right| \ . \end{aligned} \tag{5.41}$$

In this expression the elements

$$\{ \left|s, t' = t \pm 1/2; q, m\right\rangle_{i,k} \} , \tag{5.42}$$

are obtained by coupling ${\mathcal{U}_{1/2}}^X$ and ${\mathcal{U}_t}^B$ and, as usual, we assume they nullify whenever s, t and q do not fulfil the necessary selection rules. These vectors form a new basis connected with $\{|s' = s \pm 1/2, t; q, m\rangle_{i,k}\}$ via a unitary transformation in the multiplicity space $\mathbb{C}^{g(s,t,q)}$, expressed by the following four amplitude probabilities

$$\begin{aligned} C^{(s,t,q)}_{++} &\equiv {}_{i,k}\left\langle s+\tfrac{1}{2}, t; q, m \middle| s, t+\tfrac{1}{2}; q, m \right\rangle_{i,k} \ , \\ C^{(s,t,q)}_{+-} &\equiv {}_{i,k}\left\langle s+\tfrac{1}{2}, t; q, m \middle| s, t-\tfrac{1}{2}; q, m \right\rangle_{i,k} \ , \\ C^{(s,t,q)}_{-+} &\equiv {}_{i,k}\left\langle s-\tfrac{1}{2}, t; q, m \middle| s, t+\tfrac{1}{2}; q, m \right\rangle_{i,k} \ , \\ C^{(s,t,q)}_{--} &\equiv {}_{i,k}\left\langle s-\tfrac{1}{2}, t; q, m \middle| s, t-\tfrac{1}{2}; q, m \right\rangle_{i,k} \ , \qquad , \end{aligned} \tag{5.43}$$

relating the two different recouplings Eq. (5.21) and Eq. (5.42) of the irreducible representations $s, t, \frac{1}{2}$. This is exactly the information that the Wigner 6j symbols [VMK88] of SU(2) encode, and indeed $C^{(s,t,q)}_{\pm\pm}$ can be written as

$$C^{(s,t,q)}_{\pm\pm} \quad = \quad (-1)^{\pm\frac{1}{2}\pm\frac{1}{2}} \sqrt{(2s \pm 1 + 1)(2t \pm 1 + 1)} \begin{Bmatrix} t \pm \frac{1}{2} & t & \frac{1}{2} \\ s \pm \frac{1}{2} & s & q \end{Bmatrix} , \tag{5.44}$$

which for the particular case at hand gives a closed analytic expression. Notice that $C^{(s,t,q)}_{\pm\pm}$ do not depend on m, as requested by Schur's lemma.

From Eq. (5.37) and (5.40) it now follows that a similar decomposition holds also for Θ,

$$\Theta = \bigoplus_{s,t} \bigoplus_{i,k} \left(\bigoplus_{q,m} \Theta|^{s,t,q,m}_{i,k} \right) , \tag{5.45}$$

where for assigned s, t, i and k, $\Theta|_{i,k}^{s,t,q,m}$ are the following 1×1 or 2×2 matrices

$$\Theta|_{i,k}^{s,t,q,m} = \alpha^{(n)}|_{i,k}^{s,t,q,m} - \beta^{(n)}|_{i,k}^{s,t,q,m} \ . \tag{5.46}$$

Invoking the convention established when introducing Eq. (5.21) we notice that 1×1 blocks occur explicitly in the scenarios detailed in the introductory part of the section: *a)* $q = s + t + \frac{1}{2}$, *b)* $q = t - s - \frac{1}{2}$ and $t > s$, and *c)* $q = s - t - \frac{1}{2}$ and $s > t$, yielding the eigenvalues

$$\lambda_{s,t,q}^{(n)} \ = \ f_s^{(n)}(r_1) f_t^{(n)}(r_2) \, \Lambda_{s,t,q}^{(n)} \ , \tag{5.47}$$

with

$$\Lambda_{s,t,q}^{(n)} = \begin{cases} R_{s,+}^{(n)}(r_1) - R_{t,+}^{(n)}(r_2) & \text{case } a), \\ R_{s,+}^{(n)}(r_1) - R_{t,-}^{(n)}(r_2) & \text{case } b), \\ R_{s,-}^{(n)}(r_1) - R_{t,+}^{(n)}(r_2) & \text{case } c), \end{cases} \tag{5.48}$$

where we introduced the functions

$$R_{j,\pm}^{(n)}(r) \equiv \frac{f_{j\pm 1/2}^{(n+1)}(r)}{f_j^{(n)}(r)} \ . \tag{5.49}$$

For s, t, and q belonging to the remaining case *d)* instead, (5.46) is a 2×2 matrix of the form (5.26)

$$\begin{bmatrix} \Theta_{++}^{(s,t,q)} & \Theta_{+-}^{(s,t,q)} \\ \Theta_{-+}^{(s,t,q)} & \Theta_{--}^{(s,t,q)} \end{bmatrix},$$

with eigenvalues as in (5.28) with the following identifications

$$\Theta_{++}^{(s,t,q)} = f_s^{(n)}(r_1) f_t^{(n)}(r_2) \left[R_{s,+}^{(n)}(r_1) - R_{t,+}^{(n)}(r_2)(C_{++}^{(s,t,q)})^2 - R_{t,-}^{(n)}(r_2)(C_{+-}^{(s,t,q)})^2 \right],$$
$$\Theta_{--}^{(s,t,q)} = f_s^{(n)}(r_1) f_t^{(n)}(r_2) \left[R_{s,-}^{(n)}(r_1) - R_{t,+}^{(n)}(r_2)(C_{-+}^{(s,t,q)})^2 - R_{t,-}^{(n)}(r_2)(C_{--}^{(s,t,q)})^2 \right],$$

and

$$\Theta_{+-}^{(s,t,q)} = -f_s^{(n)}(r_1) f_t^{(n)}(r_2) \left[R_{t,+}^{(n)}(r_2) C_{++}^{(s,t,q)} C_{-+}^{(s,t,q)} + R_{t,-}^{(n)}(r_2) C_{+-}^{(s,t,q)} C_{--}^{(s,t,q)} \right],$$

where we used the coefficients $C_{\pm\pm}^{(s,t,q)}$ in Eq. (5.43) to express the elements of $\beta^{(n)}|_{i,k}^{s,t,q,m}$ into the basis $\{|s' = s \pm 1/2, t; q, m\rangle_{i,k}\}$. The corresponding eigenvalues can also be expressed as in the rescaled form in Eq. (5.47) with

$$\Lambda_{s,t,q}^{(n)}(\pm) = a_{s,t}^{(n)} \pm b_{s,t,q}^{(n)} \ , \tag{5.50}$$

the functions $a_{s,t}^{(n)}$ and $b_{s,t,q}^{(n)}$ being defined as

$$a_{s,t}^{(n)} \equiv \frac{R_{s,+}^{(n)}(r_1)+R_{s,-}^{(n)}(r_1)-R_{t,+}^{(n)}(r_2)-R_{t,-}^{(n)}(r_2)}{2} , \tag{5.51}$$

$$b_{s,t,q}^{(n)} \equiv \frac{\sqrt{[G_s(r_1)-G_t(r_2)]^2-4G_s(r_1)G_t(r_2)(C_{++}^{(s,t,q)})^2}}{2} , \tag{5.52}$$

where for ease of notation we introduced

$$G_j(r) \equiv f_{j+1/2}^{(n+1)}(r) - f_{j-1/2}^{(n+1)}(r) . \tag{5.53}$$

For future reference we observe that from Eq. (5.44) the following inequality can be determined

$$b_{s,t,q}^{(n)} \geq b_{s,t,q=s+t-1/2}^{(n)} , \tag{5.54}$$

which in turn can be used to establish useful bounds for the eigenvalues (5.50), i.e.

$$\Lambda_{s,t,q}^{(n)}(+) \geq \Lambda_{s,t,q=s+t-1/2}^{(n)}(+) , \tag{5.55}$$

$$\Lambda_{s,t,q}^{(n)}(-) \leq \Lambda_{s,t,q=s+t-1/2}^{(n)}(-) . \tag{5.56}$$

Replacing all this into Eq. (5.17) we can finally write

$$P_{err,min}^{(n)} = \frac{1}{2} - \frac{1}{2} \sum_{s,t,q,\ell}{}^{+} f_s^{(n)}(r_1) f_t^{(n)}(r_2)\, M_{s,t,q}^{(n)}\, \Lambda_{s,t,q}^{(n)}(\ell) , \tag{5.57}$$

with $M_{s,t,q}^{(n)}$ being the multiplicity factor defined in Eq. (5.25), the index ℓ assuming the values $\pm$ for the case *d)*, and where the subscript $^+$ indicates that only the positive values of $\Lambda_{s,t,q}^{(n)}(\ell)$ are allowed into the sum. In order to get an asymptotic expansion of Eq. (5.57) we now notice that for large n the following expansion holds,

$$f_s^{(n)}(r)\omega(s,n) \approx \frac{1+r}{r} \frac{1}{1+\frac{n}{2}+s} B(n, \frac{1+r}{2}, n/2+s) \tag{5.58}$$

where $B(n, \frac{1+r}{2}, n/2+s)$ is a binomial distribution for the variable $n/2+s$, and the neglected terms give an exponentially suppressed contribution as n goes to infinity. The mean of $\frac{s}{n}$ is $\frac{r}{2}$ and the variance is $\frac{1-r^2}{4n}$, the next moments give contribution $O(n^{-2})$. The sum over s goes from zero or $1/2$ to $n/2$, therefore if r is sufficiently greater than 0 we are neglecting in the sum a region where the binomial distribution is small and the total contribution of the region to the sum is exponentially suppressed. The second useful observation is that the eigenvalues and the term multiplying the binomial distribution in Eq. (5.58), expanded in the variables $\frac{s}{n}$ and $\frac{t}{n}$ around their means, show series coefficients that do not increase in powers of n as one goes to higher terms. Therefore to get the

leading and next to leading term one needs the expansion only at second order in these variables. We will see this kind of technique to produce an asymptotic expansion also in Chapter 6, therefore we provide in Appendix A.3 a lemma A.3.1 which permit a more formal treatment, which allows to control in principle the remainder terms.

The expansion in $\frac{s}{n}, \frac{t}{n}$ around their means let us also determine the sign of the eigenvalues in the relevant region for the sum. In particular for the four cases analyzed so far we have:

a) $\Lambda^{(n)}_{s,t,q=s+t+1/2} = \frac{r_1-r_2}{2} + O\left(|\frac{s}{n} - \frac{r_1}{2}| + |\frac{t}{n} - \frac{r_2}{2}| + |\frac{1}{n}|\right)$,

b) $\Lambda^{(n)}_{s,t,q=t-s-1/2} = \frac{r_1+r_2}{2} + O\left(|\frac{s}{n} - \frac{r_1}{2}| + |\frac{t}{n} - \frac{r_2}{2}| + |\frac{1}{n}|\right)$,

c) $\Lambda^{(n)}_{s,t,q=s-t-1/2} = -\frac{r_1+r_2}{2} + O\left(|\frac{s}{n} - \frac{r_1}{2}| + |\frac{t}{n} - \frac{r_2}{2}| + |\frac{1}{n}|\right)$,

d) $\Lambda^{(n)}_{s,t,q}(+) \geq \frac{\sqrt{(r_1-r_2)^2}}{2} + O\left(|\frac{s}{n} - \frac{r_1}{2}| + |\frac{t}{n} - \frac{r_2}{2}| + |\frac{1}{n}|\right)$,
$\Lambda^{(n)}_{s,t,q}(-) \leq -\frac{\sqrt{(r_1-r_2)^2}}{2} + O\left(|\frac{s}{n} - \frac{r_1}{2}| + |\frac{t}{n} - \frac{r_2}{2}| + |\frac{1}{n}|\right)$,
as $\frac{s}{n} \to \frac{r_1}{2}$, $\frac{t}{n} \to \frac{r_2}{2}$, and $n \to \infty$.

where in deriving the last two inequalities we used (5.55) and (5.56). The above expressions allows us to identify the positive terms which, in the limit of large n, contribute to the sum (5.57): for instance taking $r_1 > r_2$ we noticed that the positive eigenvalues are those associated with case *a)* and the first of case *d)*, while the case *b)*, which is also positive, can be ignored because $t > s$ is not in the relevant region of the sum over s, t. With this information, the sum over q can now be performed at the relevant order with the second order of the Euler-MacLaurin expansion (the details are available in the supplementary Mathematica [Wol18] notebooks, available at [Git]):

$$\sum_{i=a}^{b} f(i) \approx \int_a^b f(x)dx + \frac{f(a)+f(b)}{2}. \tag{5.59}$$

We noticed that the following orders in the Euler-MacLaurin expansion do not contribute at the order of our asymptotic expansion. However, we did not bound rigorously the remainder term, and the validity of the approximation is confirmed by the numerical check.

The final result, which takes into account also the case $r_1 < r_2$, is

$$\begin{aligned} P^{(n\gg 1)}_{err,min} &\simeq \frac{1}{2} - \frac{1}{24}\frac{(r_1+r_2)^3 - |r_1-r_2|^3}{r_1 r_2} + \frac{5}{24\,n}\frac{(r_1+r_2)^3 + |r_1-r_2|^3}{r_1^2 r_2^2} \\ &- \frac{1}{24\,n}\frac{(r_1+r_2)^5 - |r_1-r_2|^5}{r_1^3 r_2^3}. \end{aligned} \tag{5.60}$$

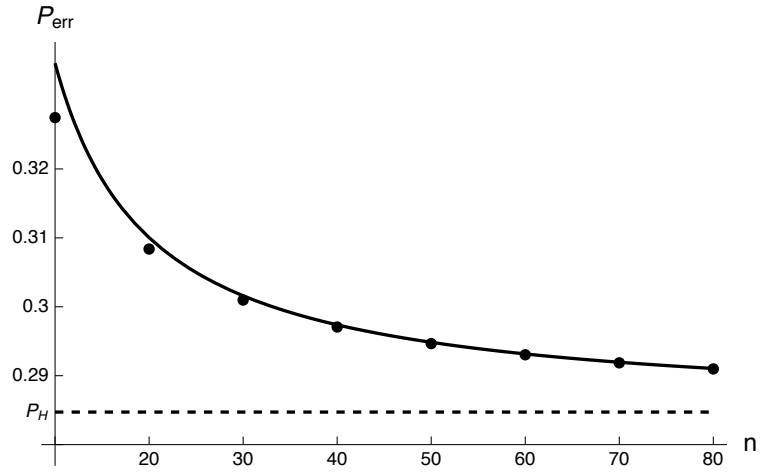


Figure 5.1: Scenario i) Minimal probability of error as a function of n, with $r_1 = \frac{3}{4}$ and $r_2 = \frac{1}{2}$: exact values (dots), asymptotic expansion Eq. (5.60) (solid line), Helstrom probability (dashed line).

which for $r_1 = r_2$ reproduce correctly the result of [Sen+10], and which in agreement with Eq. (5.9) exhibits a leading order that corresponds to the average of the Helstrom probabilities, i.e.

$$\bar{P}_H = \frac{1}{2} - \frac{1}{4}\int \sin\theta d\theta \frac{\sqrt{(r_1 - r_2\cos\theta)^2 + r_2\sin^2\theta}}{2} = \frac{1}{2} - \frac{1}{24}\frac{(r_1+r_2)^3 - |r_1 - r_2|^3}{r_1 r_2} . \tag{5.61}$$

In Fig. 5.1 we show the comparison between the exact values of $P^{(n)}_{err,min}$ of Eq. (5.57) and the asymptotic expansion Eq. (5.60).

5.3.2 Scenario ii): mixed states with hard sphere prior

In the scenario ii) we are interested in considering the case where ρ_1 and ρ_2 are arbitrary (possibly) mixed density matrices. As already stated, this corresponds to the choice

$$d\mu(\rho_1, \rho_2) = dU_1 3r_1^2 dr_1 dU_2 3r_2^2 dr_2 , \tag{5.62}$$

where again dU represents the Haar measure of SU(2) while $d\mu(r)$ is a measure that gauges our ignorance about the purity of the template states, i.e. the length of their associated Bloch vectors. Accordingly the only difference with the previous paragraph is that now, in the expression of $\alpha^{(n)} = \bigoplus_{s,t}\bigoplus_{i,k}(\bigoplus_{q,m}\alpha^{(n)}|^{s,t,q,m}_{i,k})$ and $\beta^{(n)} = \bigoplus_{s,t}\bigoplus_{i,k}\left(\bigoplus_{q,m}\beta^{(n)}|^{s,t,q,m}_{i,k}\right)$ given in Eq. (5.38) and (5.41) we have now to replace the functions $f_j^{(n)}(r)$ with their averaged values, i.e.

$$f_j^{(n)}(r) \to f_j^{(n)} \equiv \int d\mu(r) f_j^{(n)}(r) , \tag{5.63}$$

such that

$$\alpha^{(n)}|_{i,k}^{s,t,q,m} = f_{s+1/2}^{(n+1)} f_t^{(n)} \left|s+1/2,t;q,m\right\rangle_{i,k} \left\langle s+1/2,t;q,m\right| + f_{s-1/2}^{(n+1)} f_t^{(n)} \left|s-1/2,t;q,m\right\rangle_{i,k} \left\langle s-1/2,t;q,m\right| \ , \tag{5.64}$$

and

$$\beta^{(n)}|_{i,k}^{s,t,q,m} = f_s^{(n)} f_{t+1/2}^{(n+1)} \left|s,t+1/2;q,m\right\rangle_{i,k} \left\langle s,t+1/2;q,m\right| + f_s^{(n)} f_{t-1/2}^{(n+1)} \left|s,t-1/2;q,m\right\rangle_{i,k} \left\langle s,t-1/2;q,m\right| \ . \tag{5.65}$$

Our choice for $d\mu(r)$ yields

$$f_j^{(n)} = 6\frac{\left(\frac{n}{2}-j\right)!\left(1+\frac{n}{2}+j\right)!}{(n+3)!} \ . \tag{5.66}$$

The associated eigenvalues of Θ can then be expressed as in Eq. (5.47) with the rescaled quantities $\Lambda_{s,t,q}^{(n)}$ such that the eigenvalues $\lambda_{s,t,q}^{(n)}$ are

$$\lambda_{s,t,q}^{(n)} = f_s^{(n)} f_t^{(n)} \ \Lambda_{s,t,q}^{(n)} \ ,$$

$\Lambda_{s,t,q}^{(n)}$ are obtained as in Eq. (5.48),(5.50), with the terms $R_{s,\pm}^{(n)}(r)$ being replaced by

$$R_{s,+}^{(n)} \equiv \frac{f_{s+1/2}^{(n+1)}}{f_s^{(n)}} = \frac{2+\frac{n}{2}+s}{n+4}, \qquad R_{s,-}^{(n)} \equiv \frac{f_{s-1/2}^{(n+1)}}{f_s^{(n)}} = \frac{1+\frac{n}{2}-s}{n+4} \ , \tag{5.67}$$

and the same for $R_{t,\pm}^{(n)}(r)$.

As a result, for the cases *a)*, *b)*, *c)*, and *d)*, we get the following solutions,

a)

$$\Lambda_{s,t,q=s+t+\frac{1}{2}}^{(n)} = \frac{s-t}{n+4},$$

b)

$$\Lambda_{s,t,q=t-s-\frac{1}{2}}^{(n)} = \frac{1+s+t}{n+4},$$

c)

$$\Lambda_{s,t,q=s-t-\frac{1}{2}}^{(n)} = -\frac{1+s+t}{n+4}.$$

d)

$$\Lambda_{s,t,q}^{(n)}(\pm) = \pm\frac{\sqrt{3-4q(1+q)+8s(1+s)+8t(1+t)}}{2(n+4)},$$

which shows that only terms entering in the expression (5.57) for $P^{(n)}_{err,min}$ are those of *a)* with $s > t$, those of *b)*, and the $\Lambda^{(n)}_{s,t,q}(+)$ term of *d)*. Accordingly we can write

$$P^{(n)}_{err,min} = \frac{1 - S^{(n)}}{2} \ , \tag{5.68}$$

with

$$\begin{aligned} S^{(n)} \ = & \sum_{s>t} f^{(n)}_s f^{(n)}_t M^{(n)}_{s,t,s+t+\frac{1}{2}} \ \Lambda^{(n)}_{s,t,s+t+\frac{1}{2}} + \sum_{t>s} f^{(n)}_s f^{(n)}_t M^{(n)}_{s,t,t-s-\frac{1}{2}} \ \Lambda^{(n)}_{s,t,t-s-\frac{1}{2}} \\ & + \sum_{s,t} f^{(n)}_s f^{(n)}_t \sum_{q=|s-t|+\frac{1}{2}}^{s+t-\frac{1}{2}} M^{(n)}_{s,t,q} \ \Lambda^{(n)}_{s,t,q}(+), \end{aligned} \tag{5.69}$$

with $M^{(n)}_{s,t,q}$ the multiplicity factors of defined in Eq. (5.25) which allow for a simplification of the resulting formula thanks to the identity

$$f^{(n)}_s \ f^{(n)}_t M^{(n)}_{s,t,q} = \frac{36(2s+1)(2t+1)(2q+1)}{(n+1)^2(n+2)^2} \ . \tag{5.70}$$

To get to the final result at order $O\left(\frac{1}{n}\right)$ one can still exploit the Euler McLaurin formula (5.59) for each of the three sums, and the details are available in the supplementary Mathematica notebooks [Git]. The result is

$$P^{(n\gg 1)}_{err,min} \simeq \frac{17}{70} + \frac{18}{35n} \ , \tag{5.71}$$

which in $n \to \infty$ agrees with the average Helstrom probability $\bar{P}_H = 17/70$ that in the present case can be obtained by integrating Eq. (5.61) with respect to r_1 and r_2 with the corresponding hard sphere measures. In Fig. 5.2 we show the comparison between the exact values of $P^{(n)}_{err,min}$ of Eq. (5.68) and the asymptotic expansion in Eq. (5.71).

5.3.3 Scenario iii): pure states with fixed overlap

Scenario iii) considers the case where the templates states $\rho_1 = |\psi\rangle\langle\psi|$ and $\rho_2 = |\phi\rangle\langle\phi|$ are pure and characterized by a mutual overlap $|\langle\psi|\phi\rangle|^2 = \sin^2\frac{\theta}{2}$ which is known a priori, while no information about the absolute orientation of the pair of states is assumed. This task could be relevant in a scenario where for instance the machine is asked to discriminate between two possible configurations on the basis of templates generated by an external party, which does not share a common reference frame with the machine itself. Without loss of generality we can model this problem by setting

$$|\psi\rangle = U|0\rangle \ , \qquad |\phi\rangle = UU_0|0\rangle \ , \tag{5.72}$$

with a fixed unitary U_0, such that $|\langle 0|U_0|0\rangle|^2 = \sin^2\frac{\theta}{2}$, and U to be averaged over the Haar measure. This time we can consider a general finite dimension d, and still solve the

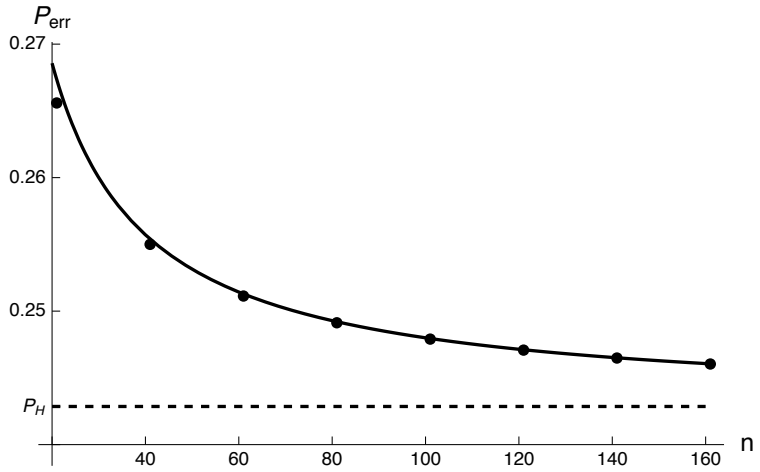


Figure 5.2: Scenario ii) Minimal probability of error as a function of n: exact values (dots), asymptotic expansion Eq. (5.71) (solid line), Helstrom probability (dashed line).

problem using the recoupling theory of SU(2), using the result of Theorem 4.2.6. Indeed by substituting $h \to n+1$, $k \to n$, $J \to q$, and expressing c in terms of θ, the states in Eq. (5.4) can be written as

$$\begin{aligned} \alpha^{(n)} &= \int dU \left(U |0\rangle \langle 0| U^\dagger\right)^{\otimes n+1} \otimes \left(UU_0 |0\rangle \langle 0| U_0^\dagger U^\dagger\right)^{\otimes n} \\ &= \textstyle\sum_q P_{n+1,n}(q|\theta) \frac{I_{\lambda_q}}{\omega_{\lambda_q}^{(d)}} \otimes |q_{n+1,n}\rangle\langle q_{n+1,n}| \end{aligned} \tag{5.73}$$

$$\begin{aligned} \beta^{(n)} &= \int dU \left(U |0\rangle \langle 0| U^\dagger\right)^{\otimes n} \otimes \left(UU_0 |0\rangle \langle 0| U_0^\dagger U^\dagger\right)^{\otimes n+1} \\ &= \textstyle\sum_q P_{n+1,n}(q|\theta) \frac{I_{\lambda_q}}{\omega_{\lambda_q}^{(d)}} \otimes |q_{n,n+1}\rangle\langle q_{n,n+1}| \end{aligned} \tag{5.74}$$

where $P_{h,k}(q|\theta)$ depends only on θ and not on d, I_{λ_q} is the identity operator on $\mathcal{U}_{\lambda_q}(SU(d))$, and $|q_{n+1,n}\rangle\langle q_{n+1,n}|$ and $|q_{n,n+1}\rangle\langle q_{n,n+1}|$ are two different pure states of $\mathcal{V}_{\lambda_q}(S_{2n+1})$. From these expressions, it is manifest that the trace distance between $\alpha^{(n)}$ and $\beta^{(n)}$ depends only on θ, therefore we can compute it for $d = 2$. $P_{h,k}(q|\theta)$ is computed in Appendix A.1, obtaining in the special case of $h = n+1, k = n$

$$P_{n+1,n}(q|\theta) = \sum_{h=-n/2}^{n/2} D^{\frac{n}{2}}_{h,\frac{n}{2}}(U_0(\theta) D^{\frac{n}{2}}_{\frac{n}{2},h}(U_0(\theta)^\dagger)\, C^{q,\frac{n+1}{2}+h}_{\frac{n+1}{2},\frac{n+1}{2},\frac{n}{2},h} C^{q,\frac{n+1}{2}+h}_{\frac{n+1}{2},\frac{n+1}{2},\frac{n}{2},h}, \tag{5.75}$$

where $U_0(\theta) = \exp(-i\sigma_y(\pi-\theta)/2)$, and the symbol $D^j_{mm'}(U)$ represent the matrix elements of the irreducible representations of $U \in \mathrm{SU}(2)$ with dimension $2j+1$, and $C^{q,l}_{j,m,j',m'}$ being the Clebsch-Gordan coefficients, as defined in Sec. 4.1.5.

We now rewrite $\alpha^{(n)}$ in terms of the decomposition used in this chapter. With respect to previous scenarios, $\alpha^{(n)}$ loses the invariance under independent $U_1^{\otimes n+1}, U_2^{\otimes n}$ $U_1, U_2 \in \mathrm{SU}(2)$ applied respectively to the system AX and B. However, since multi-copy pure states are supported in the completely symmetric subspace, corresponding to the Young diagram with only one row, we have the following decomposition

$$\alpha^{(n)} = \sum_q I_q{}^{AXB} \otimes \alpha^{(n)}_q(\theta) \otimes I^{AX}_{n/2+1/2,n+1} \otimes I^B_{n/2,n}\,, \tag{5.76}$$

where $\alpha^{(n)}_q(\theta)$ is an operator on $\mathbb{C}^{g(n/2+1,n/2,q)}$, $I^{AX}_{n/2+1,n+1} \otimes I^B_{n/2,n}$ is the identity operator on $\mathcal{V}^{AX}_{n/2+1,n} \otimes \mathcal{V}^B_{n/2,n}$, I^{AXB}_q is the identity operator on $\mathcal{U}^{AXB}_q$. $\mathcal{V}^A_{n/2+1,n+1} \otimes \mathcal{V}^B_{n/2,n}$ has dimension 1, therefore we do not need indices i,k to label a basis for the support of $\alpha^{(n)}_q(\theta)$. From these symmetries, one can obtain the eigenvectors of $\alpha^{(n)}$ as $\{\left|\frac{n+1}{2}, \frac{n}{2}; q, m\right\rangle\}$ and their respective eigenvalues are computed in A.1):

$$\alpha^{(n)} \left|\tfrac{n+1}{2}, \tfrac{n}{2}; q, m\right\rangle = \frac{P_{n+1,n}(q|\theta)}{2q+1} \left|\tfrac{n+1}{2}, \tfrac{n}{2}; q, m\right\rangle, \tag{5.77}$$

Analogous properties applies for $\beta^{(n)}$ when expressed into the basis in Eq. (5.42). Therefore as in the previous cases Θ can be expressed as a direct sum of 1×1 and 2×2 block matrices. In the present case, however due to the special restriction on s and t instead of the four possible cases observed in the previous section, only *a)* and *d)* may occur. It turns out that for the case *a)* the associated eigenvalue is always null. For *d)* instead we have

$$\lambda^{(n)}_{s=n/2,t=n/2,q}(\pm) = \pm\frac{P_{n+1,n}(q|\theta)}{2q+1}|C^{(s=n/2,t=n/2,q)}_{+-}|, \tag{5.78}$$

and the eigenvectors are the same that we obtain for $r_1 = r_2 = 1$ in the case of completely random orientations: for pure states, the optimal POVM in the fixed overlap case is the same. Therefore, writing the eigenvalues in a simpler notation as $\lambda^{(n)}_q(\pm)$, we have

$$\Theta^{(n)} = \sum_q \left(\lambda^{(n)}_q(+)\Pi_{q,+} + \lambda^{(n)}_q(-)\Pi_{q,-}\right), \tag{5.79}$$

where Π_{q+} and Π_{q-} are the projectors on eigenvectors with total angular momentum q and respectively positive and negative eigenvalues.

Replacing all this into Eq. (5.17) we can finally write

$$P_{err,min}^{(n)} = \frac{1}{2} - \frac{1}{2}\sum_{q}(2q+1)\ \lambda_q^{(n)}(+)\ , \tag{5.80}$$

where we used the fact that $M^{(n)}_{s=n/2,t=n/2,q} = 2q+1$ and that only the + elements of the pairs (5.78) are positive. We note that

$$D^{\frac{n}{2}}_{h,\frac{n}{2}}(U_0)D^{\frac{n}{2}}_{\frac{n}{2},h}(U_0^\dagger) = \tfrac{n!}{(\frac{n}{2}+h)!(\frac{n}{2}-h)!}\left(\cos^2\left(\tfrac{\pi-\theta}{2}\right)\right)^{\frac{n}{2}+h}\left(\sin^2\left(\tfrac{\pi-\theta}{2}\right)\right)^{\frac{n}{2}-h},$$

is a binomial distribution in the variable $\frac{n}{2}+h \in \{0,...,n\}$. We also note that

$$\left(C^{q,\frac{n+1}{2}+h}_{\frac{n+1}{2},\frac{n+1}{2},\frac{n}{2},h}\right)^2 = \frac{2\left(\frac{n}{2}-h\right)!\,(n+1)!}{\left(\frac{n}{2}+h\right)!}\frac{\left(\frac{n}{2}+h+q+\frac{1}{2}\right)!}{\left(q-\frac{1}{2}-\frac{n}{2}-h\right)!\left(n-q+\frac{1}{2}\right)!\left(n+q+\frac{3}{2}\right)!}, \tag{5.81}$$

is also a probability distribution in the variable $q \in \{\frac{n}{2}+h,...,n+\frac{1}{2}\}$. Then the terms entering in the sum of Eq. (5.80) rewrite explicitly as

$$\begin{aligned}(2q+1)\lambda_q^{(n)}(+) = \sum_h \frac{n!}{\left(\frac{n}{2}+h\right)!\left(\frac{n}{2}-h\right)!}\left(\cos^2\left(\frac{\pi-\theta}{2}\right)\right)^{\frac{n}{2}+h}\left(\sin^2\left(\frac{\pi-\theta}{2}\right)\right)^{\frac{n}{2}-h}\\ \times\frac{2\left(\frac{n}{2}-h\right)!\left(\frac{n}{2}+h+q+\frac{1}{2}\right)!\,(n+1)!}{\left(\frac{n}{2}-h+q-\frac{1}{2}\right)!\left(\frac{n}{2}+h\right)!\left(n-q+\frac{1}{2}\right)!\left(n+q+\frac{3}{2}\right)!}\frac{1}{2}\sqrt{\frac{2(3/2+q+n)(1/2-q+n)}{(n/2+1/2)(n+1)}}\ .\end{aligned} \tag{5.82}$$

As usual we focus on the limit of large $n \gg 1$ for $P^{(n)}_{err,min}$. In this case we notice that in order to get up to the order $O(\frac{1}{n^2})$ for the resulting expression, one can expand $|C^{(s=n/2,t=n/2,q)}_{+-}|$ around the mean of the distribution in q and consider contributions up to the fourth central moment (see Appendix A.3 and supplementary Mathematica notebooks [Git]), expand the result around the mean of the h distribution and calculate the contributions up to the relevant moment (not more than the fourth). The result is

$$P^{(n\gg1)}_{err,min} \simeq \frac{1}{2}\left(1-|\cos\tfrac{\theta}{2}|\right) + \frac{3+\cos\theta}{8\sqrt{2}\sqrt{1+\cos\theta}}\frac{1}{n} + \frac{1-60\cos\theta-5\cos 2\theta}{128\sqrt{2}(1+\cos\theta)^{3/2}}\frac{1}{n^2}\ , \tag{5.83}$$

where, as expected, the first contribution corresponds to the corresponding averaged Helstrom probability $\bar{P}_H$ – see also Fig. 5.3. We notice that for small deviations from orthogonality, one has

$$P^{(n\gg1)}_{err,min} \simeq \frac{\theta^2}{16} + \frac{1}{4n} - \frac{1}{8n^2}\left(1-\frac{\theta^2}{4}\right)\ , \tag{5.84}$$

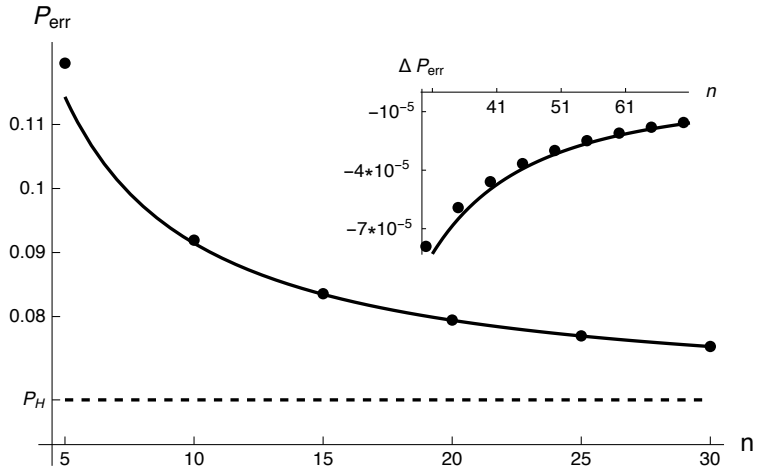


Figure 5.3: Scenario iii) Minimal probability of error as a function of n,with $\theta = \frac{\pi}{3}$: exact values (dots), asymptotic expansion Eq. (5.83) (solid line), Helstrom probability (dashed line). In the inset we show the second order correction.

The expansion around coincident states is instead singular, but the formula is still valid when the states are not coincident and $n(\pi - \theta) \gg 1$. Since the optimal POVM is the same of the totally random pure state scenario, averaging over θ before doing the asymptotic expansion gives the result of Eq. (5.60) when $r_1 = r_2 = 1$. Integrating over the probability distribution of the overlap $c = \sin^2 \frac{\theta}{2}$, which for Haar random $|\psi\rangle_1$ and $|\psi_2\rangle$ is known (e.g. [AG15]) and equal to $P(c) = (d-1)(1-c)^{d-2}$, gives also the same result up to first order, while the order n^{-2} is not integrable. This is not inconsistent: one can see that the averaged $P^{(n)}_{err,min}$ displays a $n^{-\frac{3}{2}}$ correction to Eq. (5.60) which is not recoverable from this expansion, which at fixed n works only in the region $n(\pi - \theta) \gg 1$. We obtain

$$P^{(n\gg 1)}_{err,min,d} \simeq \frac{1}{2} - \frac{d-1}{2d-1} + \frac{(d-1)^2}{3+4d(d-2)}\frac{1}{n}\,. \qquad . \tag{5.85}$$

which agrees with the zero-th order result in [HHH05a]. The asymptotic correction that we find can be also directly calculated by following the approach in [HHH05a], the interested reader can find the calculations in the supplementary Mathematica notebooks [Git].

5.3.4 Compatibility between optimal machines

In the previous subsections we have analysed three different scenarios, which in principle give rise to different optimal machines. However, additional symmetries make some

of the optimal machines compatible, in the sense that there exists a measurement that is optimal for different scenarios. In particular, if $\hat{S}_{AB}$ is the swap operator between $\mathcal{H}_A$ and $\mathcal{H}_B$, one can verify that $\hat{S}_{AB}\alpha^{(n)}\hat{S}_{AB}^{\dagger} = \beta^{(n)}$ in scenario (ii), (iii) and also (i) when $r_1 = r_2$. If this happens then $\hat{S}_{AB}\Theta\hat{S}_{AB}^{\dagger} = -\Theta$; it follows that if $|\lambda\rangle$ is an eigenvector of Θ with eigenvalue λ, then also $\hat{S}_{AB}|\lambda\rangle$ is an eigenvector, with eigenvalue $-\lambda$. Since $\hat{S}_{AB}|s+1/2,s;q,m\rangle_{i,k} = |s,s+1/2;q,m\rangle_{i,k}$, $\hat{S}_{AB}|s-1/2,s;q,m\rangle_{i,k} = |s,s-1/2;q,m\rangle_{i,k}$, in the spaces $\mathcal{H}^{(s,s)}_{A_iXB_k}$ spanned by $|s\pm1/2,s;q,m\rangle_{i,k}$ the eigenvectors are automatically determined as the orthogonal vectors $|\lambda_+\rangle$,$|\lambda_-\rangle$ in $\mathcal{H}^{(s,s)}_{A_iXB_k}$ such as $\hat{S}_{AB}|\lambda_+\rangle = |\lambda_-\rangle$.

In particular, since the the relevant subspace in scenario (iii) is only $\mathcal{H}^{(\frac{n}{2},\frac{n}{2})}_{A_iXB_k}$, the optimal machine for scenario (i) when $r_1 = r_2$, or the one for scenario (ii), are also optimal for scenario (iii).

5.4 Implementation of the optimal POVM

From the knowledge of the eigenvectors in Eq. (5.29) one can reconstruct the optimal POVM. Since it is a projective measurement, it can be realized by a change of basis from the the eigenvectors to the computational basis, followed by a local measurement. In the following we consider the implementation of the optimal machine of scenario iii), for the case $n = 1$. The change of basis is:

$$\begin{aligned}
|\psi_{\frac{1}{2},\frac{1}{2};\frac{3}{2},\frac{3}{2}}\rangle &\to |\uparrow\uparrow\uparrow\rangle \quad (C)\\
|\psi_{\frac{1}{2},\frac{1}{2};\frac{3}{2},\frac{1}{2}}\rangle &\to |\uparrow\uparrow\downarrow\rangle \quad (C)\\
|\psi^{(-)}_{\frac{1}{2},\frac{1}{2};\frac{1}{2},\frac{1}{2}}\rangle &\to |\downarrow\uparrow\uparrow\rangle \quad (B)\\
|\psi^{(+)}_{\frac{1}{2},\frac{1}{2};\frac{1}{2},\frac{1}{2}}\rangle &\to |\uparrow\downarrow\uparrow\rangle \quad (A)\\
|\psi^{(+)}_{\frac{1}{2},\frac{1}{2};\frac{1}{2},-\frac{1}{2}}\rangle &\to |\downarrow\uparrow\downarrow\rangle \quad (A)\\
|\psi^{(-)}_{\frac{1}{2},\frac{1}{2};\frac{1}{2},-\frac{1}{2}}\rangle &\to |\uparrow\downarrow\downarrow\rangle \quad (B)\\
|\psi_{\frac{1}{2},\frac{1}{2};\frac{3}{2},-\frac{1}{2}}\rangle &\to |\downarrow\downarrow\uparrow\rangle \quad (C)\\
|\psi_{\frac{1}{2},\frac{1}{2};\frac{3}{2},-\frac{3}{2}}\rangle &\to |\downarrow\downarrow\downarrow\rangle \quad (C)
\end{aligned} \tag{5.86}$$

where A (B) means that the result of the measurement is interpreted as $X = A$ ($X = B$), while for C we "flip a coin" to decide. In the computational basis the unitary rotation

reads

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & \frac{-3-\sqrt{3}}{6} & 0 & \frac{3-\sqrt{3}}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-3-\sqrt{3}}{6} & 0 & \frac{1}{3+\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{3}} & \frac{3-\sqrt{3}}{6} & 0 & \frac{-3-\sqrt{3}}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{3-\sqrt{3}}{6} & 0 & \frac{1}{-3+\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.87}$$

and the probability of error as a function of θ is

$$P^{(1)}_{err,min} = \frac{1}{2} - \frac{1+\cos\theta}{4\sqrt{3}}. \tag{5.88}$$

These kind of operations are suitable for all programmable devices which are based on the circuit model of quantum computation, as for example the recent quantum superconducting processors developed by IBM [Ibm]. By using the software development kit QISKit [Abr+19], we have determined a circuit that realises the POVM for the $n = 1$ case with input pure states and checked its performance with the IBM simulator. The number of gates of our implementation is 61 single qubit operations and 60 CNOT, with a depth of 43 operations. Given that the failure probability of a CNOT on real machines is about $5 \cdot 10^{-2}$, the failure probability of the circuit is at least $1 - 0.95^{60} \approx 0.954$. Indeed we tried to remotely perform the experiment on the physical chip, without any significant results. This fact underlines the importance of gate optimisation and error correction for the proper operation of future quantum computers. However, with the simulation tools of QISKit Aer, we were able to simulate the circuit with an error model consisting in depolarising errors (Fig. 5.4) and thermal relaxation errors (Fig. 5.5): decreasing the depolarising probability and increasing the relaxation times we can show how the circuit is sensitive to this kind of noises, and that we recover the expected behaviour for small noise.

5.5 Remarks

In this work we have discussed the performances of optimal universal learning quantum machines that aim at discriminating the states of a qudit starting from a collection of templates states in the hybrid, yet realistic scenario, where at least some global information on the training set is classically available. Like classical supervised learning is a fundamental tool with classical data, arguably quantum learning machines will be important for dealing with quantum data with quantum processors. Indeed, given that quantum tomography is very expensive in terms of resources, dealing with quantum data

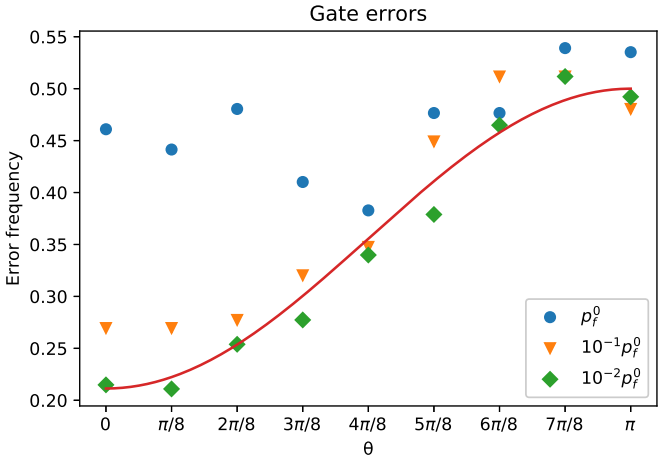


Figure 5.4: Simulation of the optimal machine with QISKit Aer, with depolarising error modeled after the gate average infidelity of each gate: p_f are the depolarising probability for the 16 qubit machine (Melbourne) if all the infidelity is due to a depolarising channel. Frequency of misclassification errors with 256 repetitions for each θ, compared with the predicted minimum error function (solid line).

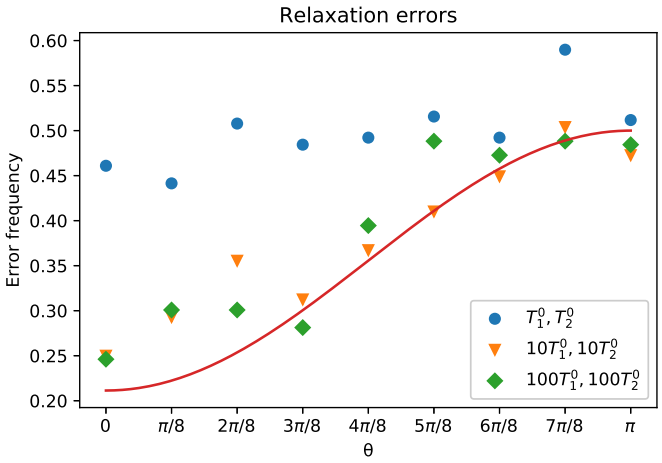


Figure 5.5: Simulation of the optimal machine with QISKit Aer, with thermal relaxation times $T_1 = T_2 = T$ equal for each qubit. Gate times are set to $200\,ns$ for 1 qubit gates and $800\,ns$ for CNOT. Frequency of misclassification errors with 256 repetitions for each θ, compared with the predicted minimum error function (solid line).

requires to study alternatives which need little information about the data, make use of the full power of quantum mechanics, and extract only the relevant information for the problem at hand. Our work extends the previous results considering more general scenarios. An interesting observation is that the optimal machine that does not assume any kind of information about the template state, scenario (ii), it is also optimal for scenario (iii), where the template states are assumed to be pure. It is therefore a very general machine, which can be seen as the most convenient learning algorithm. Several open question remain. The gate complexity of the implementation is not clear, and it would be important to determine if the separation between learning and testing phases which holds for qubits [Sen+12] is still valid in general dimension. The case of mixed states in general dimension is also not studied. While it could be hard to exactly evaluate the leading correction as in the qubit case, it would be interesting to understand how the number of copies of training states needed to get a probability of error which is arbitrarily close to the Holevo-Helstrom bound depends on the dimension, in the worst case. This would clarify quantitatively the advantage of programmable discriminators with respect to tomography.

Chapter 6

Optimal overlap estimation

This chapter is largely based on:

- M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti. "Beyond the Swap Test: Optimal Estimation of Quantum State Overlap". In: *Physical Review Letters* 124.6 (2020), p. 060503. DOI: 10.1103/PhysRevLett.124.060503. arXiv: 1906.10639.

6.1 Introduction

The overlap between two pure quantum states is an example of the unitarily invariant quantities treated in Sec. 4.2. In this chapter we consider the problem of estimating the overlap between two unknown pure states of a d-dimensional Hilbert space. This primitive attracted the attention in the quantum foundations community as an archetypical instance of estimation of relative information [BIMT06; LSB06; GI06; BRS07]. A simple way to estimate the overlap is the swap test (SWT) [Buh+01; Cin+18; Cha+18]: given two systems in the state $|\psi\rangle|\phi\rangle$, the probability of projecting it on its symmetric subspace or its orthogonal is determined by the overlap between $|\psi\rangle$ and $|\phi\rangle$. By repeating this measurement on several pairs of copies one can obtain a good estimate of this probability, and hence the overlap. The swap test is used in several quantum information processing tasks, such as quantum fingerprinting [Buh+01; dBe04; KDK17], entanglement estimation [Eke+02; Wal+07; MKB05; HM10], to quantum algorithms for classical machine learning tasks [Har+10; LMR13; RML14; Cha+18; Cin+18; ZFF19; WKS16; Hav+19; LR18]. The last application has attracted renewed interest on the overlap estimation problem, and its efficient implementation and generalization on near-term quantum computers have been discussed [Cha+18; Cin+18]. It is then a natural question to ask what are the limits to the accuracy of the estimation of the overlap for

a fixed number of copies (say N copies of $|\psi\rangle$ and M copies of $|\phi\rangle$), and how much an optimal measurement improves the performances with respect to the swap test, for the same number of copies.

The measurement optimizing the average information gain was identified in [BRS04], and it is easily obtained as the optimal measurement according to several others figure of merit, as a corollary of theorem 4.2.2, as we show in Sec. 6.2. This measurement is weak Schur sampling, according to the decomposition of Theorem 4.2.6, therefore it is also efficiently implementable [BCH06; Har05b; Kro19]. Regarding the explicit evaluation of the optimal performances, previous works have solved the minimization of the qubit average mean square error [BIMT06; LSB06], and also showed how to compute the optimal average mean square error in generic dimension, albeit without a closed form [GI06]. More recently, the estimation of the Hilbert-Schmidt distance between two unknown mixed states was used to obtain an algorithm for quantum state certification [BOW19]; in this paper, the authors determine minimum variance unbiased estimator for the Hilbert-Schmidt distance, which for pure states reduces to the overlap. We extend these results addressing the estimation problem in full generality. Our main results are an asymptotic expansion for the Fisher information of the family of averaged multi-copy states at fixed overlap, which gives a lower bound on the mean square error of any estimator, according to the Cramér Rao bound, Theorem 2.4.2. The answer we obtain is a function of the overlap, and therefore we refer to this setting as the local setting, in contrast to the global, Bayesian, setting, where we compute exactly the minimum average mean square error for d-dimensional unknown states, left open by [GI06]. We state these results in Sec. 6.2, and prove them respectively in Appendix A.4 and Sec. 6.7.

Moreover, we compare our results with the swap test and with two strategies based on estimating either one or both $|\psi\rangle$ and $|\phi\rangle$, see Fig. 6.1. Such strategies are useful in distributed scenarios where copies of $|\psi\rangle$ and $|\phi\rangle$ are produced in different and distant laboratories. In the limit of large $M+N$ and $|M-N|$ constant the optimal strategy displays a finite asymptotic gap with respect to all the others, as we recall in Sec. 6.3. Also note that the optimal measurement can be performed with the same accuracy even without assuming the states to be labeled, as an instance of unsupervised learning [Sen+19]. The same is not true for the other candidates. We also show that the optimal measurement is less invasive than the swap test (Sec. 6.4), and robust against single-qubit noise (Sec. 6.5). Although the optimal measurement is efficiently implementable in terms of gate complexity, the required number of gates can be still too high for near term implementations: it is important to address how the performance of the estimation is affected if error correction is not available, and we discuss this problem in Sec. 6.6.

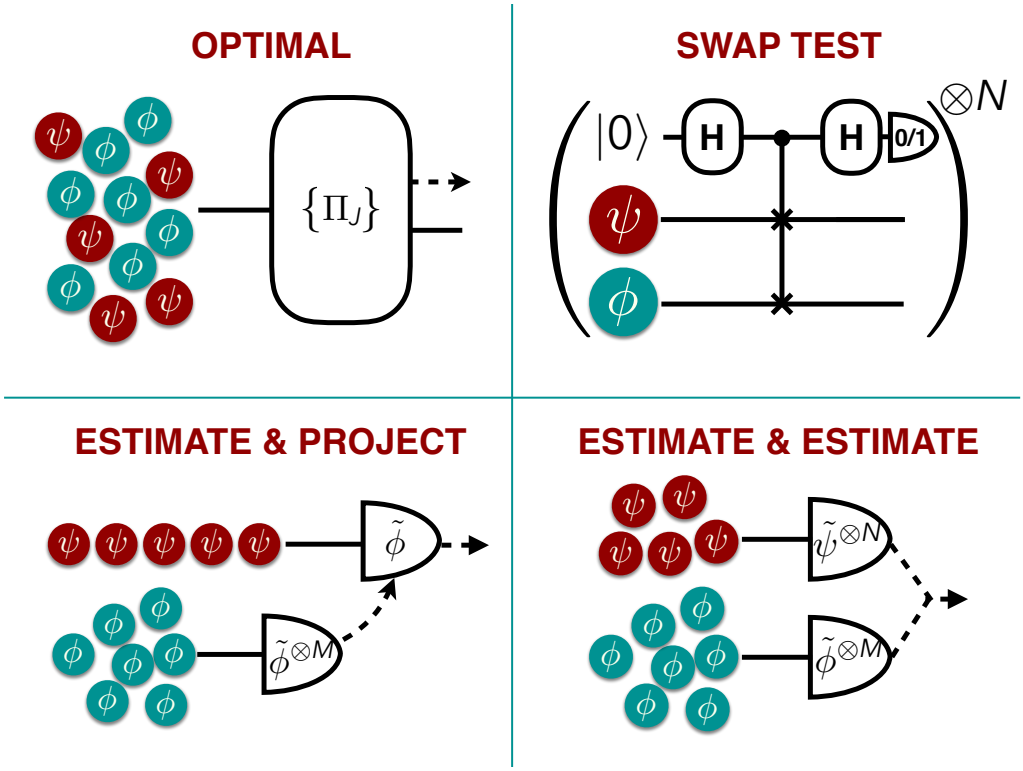


Figure 6.1: Sketch of the OvE (overlap estimation) strategies studied in the chapter. Top-left: optimal measurement, e.g. by Schur transform (see Ref. [Har05b] for the circuit implementation). Top-right: circuit for the SWT, to be repeated N times. Bottom-left: estimate $|\phi\rangle$ and project $|\psi\rangle$ on the estimated direction. Bottom-right: estimate both $|\phi\rangle$ and $|\psi\rangle$ and calculate the overlap.

6.2 Optimal measurements for overlap estimation

We study the estimation of the overlap between two unknown qudit states $|\psi\rangle$ and $|\phi\rangle$, given N copies of $|\psi\rangle$ and M copies of $|\phi\rangle$. The overlap is $c = |\langle\psi|\phi\rangle|^2$. This problem can be framed as in Theorem 4.2.2, since the input states are of the form

$$\rho(U, c) = U^{\otimes(N+M)} \, |\Psi_c\rangle\langle\Psi_c| \, U^{\dagger\otimes(N+M)}, \tag{6.1}$$

where $|\Psi_c\rangle = I^{\otimes N} \otimes W^{\otimes M} |0\rangle^{\otimes(N+M)}$, with $|\langle 0| W |0\rangle|^2 = c$. This is clearly a covariant family with respect to the action of $\mathrm{SU}(d)$ and Theorem 4.2.2 applies. For estimators $\tilde{c}$ of c, we consider as a cost function the mean square error $l(\tilde{c}, c) := (\tilde{c} - c)^2$. We are thus interested in measurements optimizing the worst case cost

$$\min_{M\in\mathcal{M}([0,1])}\max_{U\in\mathrm{SU(d)},c\in[0,1]}\int_0^1 d\tilde{c}\,\mathrm{tr}[M_{\tilde{c}}\rho_{U,c}](\tilde{c}-c)^2, \tag{6.2}$$

or, equivalently by Theorem 4.2.2, the worst case averaged cost at fixed c,

$$\min_{M\in\mathcal{M}([0,1])}\max_{c\in[0,1]}\int dU\int_0^1 d\tilde{c}\,\mathrm{tr}[M_{\tilde{c}}\rho_{U,c}](\tilde{c}-c)^2, \tag{6.3}$$

and the global Bayesian cost for some prior $p(c)$,

$$\min_{M\in\mathcal{M}([0,1])}\int_0^1 dc\,p(c)\int_0^1 d\tilde{c}\int dU\,\mathrm{tr}[M_{\tilde{c}}\rho_{U,c}](\tilde{c}-c)^2. \tag{6.4}$$

Specifying the optimality conditions to the problem at hand, we can state the following fact.

Proposition 6.2.1. *An estimator that optimizes the worst case cost, the worst case average cost, and the global Bayesian cost for the problem of estimating c from the covariant family* $\{\rho(U,c)\}$ *can be chosen to be an estimator* $\tilde{c}_J$ *from a projective measurement* $\{\Pi_J\}\bigcup\{I-\sum_J\Pi_J\}$*, where by Schur-Weyl duality* $\mathcal{H}_d^{\otimes N+M}=\bigoplus_\lambda\mathcal{U}_\lambda(\mathrm{SU}(d))\otimes\mathcal{V}_\lambda(S_{N+M})$*, and* Π_J *is the projector on* $\mathcal{U}_{\lambda_J}(\mathrm{SU}(d))\otimes\mathcal{V}_{\lambda_J}(S_{M+N}))$*,* $\lambda_J=(\frac{M+N}{2}+J,\frac{M+N}{2}-J,0,...,0)$.

Proof. From Theorem 4.2.2 we know that both the worst case average scenario at fixed c and the average case scenario for some prior probability on c are optimized by a POVM which is invariant, meaning that the POVM elements $E_{\tilde{c}}$ of the estimator $\tilde{c}$ are block diagonal according to the decomposition of $\mathcal{H}_d^{\otimes N+M}$ given by Schur-Weyl duality $\mathcal{H}_d^{\otimes N+M}=\bigoplus_\lambda\mathcal{U}_\lambda(\mathrm{SU}(d))\otimes\mathcal{V}_\lambda(S_{N+M})$. In particular, POVM elements should have the form $E_{\tilde{c}}=\sum_\lambda I_{\lambda_J}\otimes E^\lambda_{\tilde{c}}$, with I_{λ_J} is the identity operator in $\mathcal{U}_{\lambda_J}(\mathrm{SU}(d))$ and $E^\lambda_{\tilde{c}}$ is an operator on $\mathcal{V}_{\lambda_J}(S_{M+N})$.

The probability distribution of an invariant POVM on $\rho(U,c)$ is equal to the probability distribution of the same POVM acting on the averaged state

$$\rho(c)=\int \mathrm{d}U\,U^{\otimes(N+M)}\,|\Psi_0\rangle\langle\Psi_0|\,U^{\dagger\otimes(N+M)}. \tag{6.5}$$

From Theorem 4.2.6 we know that

$$\rho(c)=\sum_J P_{N,M}(J|c)\frac{I_{\lambda_J}}{\omega^{(d)}_{\lambda_J}}\otimes|J_{N,M}\rangle\langle J_{N,M}|\,, \tag{6.6}$$

where $\lambda_J := (\frac{M+N}{2} + J, \frac{M+N}{2} - J, 0, ..., 0)$, $\frac{I_{\lambda_J}}{\omega_{\lambda_J}^{(d)}}$ is the completely mixed state in $\mathcal{U}_{\lambda_J}(\mathrm{SU}(d))$, $|J_{N,M}\rangle\langle J_{N,M}| \in \Sigma(\mathcal{V}_{\lambda_J}(S_{M+N}))$ is independent of $|\psi\rangle$ and $|\phi\rangle$, and $P_{N,M}(J|c)$ is a probability distribution in J dependent only on c.

The probability distribution of the outcomes of $E_{\tilde{c}}^{\lambda}$ is $p(\tilde{c}|c) = \mathrm{tr}[E_{\tilde{c}}\rho(U,c)] = \sum_J P_{N,M}(J|c)\,\mathrm{tr}\Big[E_{\tilde{c}}^{\lambda_J}\,|J_{N,M}\rangle\langle J_{N,M}|\Big]$, with $\int_0^1 \mathrm{d}\tilde{c}\,\mathrm{tr}\Big[E_{\tilde{c}}^{\lambda_J}\,|J_{N,M}\rangle\langle J_{N,M}|\Big] = 1$. We have

$$\begin{aligned}
&\int_0^1 \mathrm{d}\tilde{c}\,p(\tilde{c}|c)(\tilde{c}-c)^2 = \sum_J P_{N,M}(J|c)\int_0^1 \mathrm{d}\tilde{c}\,\mathrm{tr}\Big[E_{\tilde{c}}^{\lambda_J}\,|J_{N,M}\rangle\langle J_{N,M}|\Big](\tilde{c}-c)^2 \\
&\geq \sum_J P_{N,M}(J|c)\left(\int_0^1 \mathrm{d}\tilde{c}\,\mathrm{tr}\Big[E_{\tilde{c}}^{\lambda_J}\,|J_{N,M}\rangle\langle J_{N,M}|\Big]\tilde{c}-c\right)^2 \\
&=: \sum_J P_{N,M}(J|c)(\tilde{c}(J)-c)^2 \\
&= \sum_J \mathrm{tr}[\Pi_J\rho(c)](\tilde{c}(J)-c)^2, \qquad (6.7)
\end{aligned}$$

where the inequality comes from the convexity of the figure of merit. This means that choosing a POVM $\{\Pi_J\}$ of orthogonal projections and a deterministic post-processing $\tilde{c}(J)$ is optimal. □

The last argument shows that remaining optimization has to be done on $\tilde{c}(J)$, an estimator of c which is sampled from the classical distribution $P_{N,M}(J|c)$. We can still apply the results on quantum estimation presented in Chapter 2 to this case, with the reduction to a purely classical problem. In particular, we consider bounding the mean square error of a worst case optimal estimator with the Cramér-Rao bound, Eq. (2.15),

$$v(c) := \mathrm{MSE}(\tilde{c}) \geq H(c)^{-1}, \qquad (6.8)$$

where $H(c) = \sum_J (\partial_c P_{M,N}(J|c))^2/P_{M,N}(J|c)$ is the Fisher information of the measurement statistics.

We find a formal asymptotic expansion for the Fisher information.

Proposition 6.2.2. *The following asymptotic expansion holds for $H(c)$, if $M+N \to \infty$ with $M-N$ constant, $0 < c < 1$.*

$$H(c) = \frac{M+N}{4c(1-c)} + O(1) \qquad (6.9)$$

For $N = M$, we compute the next to leading term as:

$$H(c) = \frac{N}{2c(1-c)} - \frac{1}{8c^2} + O\left(\frac{1}{N}\right). \tag{6.10}$$

An asymptotically unbiased estimator which saturates at the leading order the Cramér-Rao bound is

$$\tilde{c}_{\rm op}^{\rm loc}(J) := \left(\frac{J}{J_{\max}}\right)^2. \tag{6.11}$$

The proof involves manipulations with asymptotic expansions of Jacobi polynomials, and it is deferred to Appendix A.4. The estimator for the overlap of [BOW19] also saturates the Cramér-Rao bound at leading order, but not at the second order. Since that estimator is the minimum variance unbiased estimator, it means that the Cramér-Rao bound cannot be saturated by unbiased estimators at order $O(1)$. The utility of the second order characterization is to give asymptotic lower bounds for more general estimators, which may not be exactly unbiased. An asymptotic Cramér-Rao bound with the leading order in Eq. 6.11 can be obtained also with the formalism of estimation with nuisance parameters (i.e. unknown parameters which we do not care to estimate) [SYH20], but this approach does not guarantee that the remainder terms do not depend on the dimension.

Moreover, we are able to compute exactly the the optimal global Bayesian cost, with a prior probability given by the probability distribution of overlaps of two Haar random states [AG15]

$$p_d(c) = (d-1)(1-c)^{d-2}. \tag{6.12}$$

In the classical case the solution of the minimum mean square error problem of Theorem 2.4.4 simplifies, and the minimum is attained by the estimator

$$\tilde{c}_{\rm op}^{\rm bay}(J) := \frac{\int dc\, c\, p_d(c) P_{M,N}(J|c)}{\int dc\, p_d(c) P_{M,N}(J|c)}. \tag{6.13}$$

Using graphical calculus techniques for the recoupling theory of Clebsch-Gordan coefficients [VMK88], we obtain the following optimal global Bayesian estimator and corresponding AvMSE (average mean square error):

Proposition 6.2.3. *The global Bayesian optimal estimator and the minimum global*

Bayesian cost are:

$$\tilde{c}_{\text{op}}^{\text{bay}}(J) = \frac{d + J + J^2 + \frac{M+N}{2} - \left(\frac{M+N}{2}\right)^2 + MN}{(d+M)(d+N)}, \tag{6.14}$$

$$v_{\text{op}} = \frac{(d-1)(d+M+N)}{d(d+1)(d+M)(d+N)}. \tag{6.15}$$

We compute these quantities in Sec. 6.7. We pause to highlight the following facts: i) when d is fixed and the number of copies is large, the prior distribution of the states is little informative with respect to the information that can be obtained by the actual measurement; indeed we can see that when $M + N \to \infty$, $M - N$ constant, $\tilde{c}_{\text{op}}^{\text{bay}}(J) \approx \tilde{c}_{\text{op}}^{\text{loc}}(J)$ implying that the local optimal estimator is also a good Bayesian estimator and vice versa; ii) contrarily to the local estimation results, the global MSE of Eq. (6.15) is exact for all M, N and depends on d due to the prior, Eq. (6.12); iii) in particular, v_{op} decays as d^{-2} if one of either M or N is kept finite, whereas it decays only as d^{-1} when $M, N \gg 1$.

6.3 Alternative strategies

In addition to this characterization of the optimal estimators, we consider a family of intermediate strategies that employ 1-LOCC (one way local operations and classical communication) on $|\psi\rangle^{\otimes N}$ and $|\phi\rangle^{\otimes M}$, and compare their performances with the optimal measurement. The estimate-and-project (EP) strategy consists in estimating $|\phi\rangle$ from its M copies, then projecting each copy of $|\psi\rangle$ on this estimate and counting the fraction of successful projections. When $|\phi\rangle$ is known, projecting $|\psi\rangle$ on $|\phi\rangle$ is optimal [Hol11b]. However, EP is not necessarily the optimal 1-LOCC strategy. The corresponding POVM elements can be written as

$$E_{V,k}^{(\text{ep})} = dV E_V^{(M)} \otimes V^{\otimes N} \Pi_k^{(N)} V^{\dagger \otimes N}, \tag{6.16}$$

where $E_V^{(M)} = \binom{n+d-1}{d-1}(V|0\rangle\langle 0|V^\dagger)^{\otimes M}$ is the optimal covariant measurement to estimate $|\phi\rangle$ [Hay97; Hay17b]: we mentioned it as the POVM obtained from coherent states of SU(d) in Sec. 4.2.1. $\Pi_k^{(N)}$ represents k successful projections of the copies of $|\psi\rangle$ on the estimate of $|\phi\rangle$. The estimator is $\tilde{c}_{\text{ep}}^{\text{loc}}(k) = \frac{k}{N}$. The estimate-and-estimate (EE) strategy instead consists in estimating both $|\psi\rangle$ and $|\phi\rangle$ separately, then computing the overlap between the estimated states. The corresponding POVM elements can be written as

$$E_{V,W}^{(\text{ee})} = dV dW E_V^{(M)} \otimes E_W^{(N)}, \tag{6.17}$$

i.e., a product of two covariant measurements to estimate $|\phi\rangle$ and $|\psi\rangle$. We take as local estimator

$$\tilde{c}_{\text{ee}}^{\text{loc}}(V, W) = \left|\langle 0| V^\dagger W |0\rangle\right|^2. \tag{6.18}$$

In the Supplementary Material of [Fan+20a] we obtained exact results for local and Bayesian estimation using EP and EE. We will not repeat the calculation here, and we just recall the asymptotic behaviour of the two estimators.

The mean square error of the local estimator for the EP measurement in the limit $M \to \infty$, N constant is

$$v_{ep}(c) \sim \frac{c(1-c)}{N}, \tag{6.19}$$

which coincides with the one of the optimal strategy, corresponding to a projection on the known direction of $|\phi\rangle$. In the limit $M+N \to \infty$, $M-N$ fixed we have instead

$$v_{ep}(c) \sim \frac{6c(1-c)}{(M+N)}, \tag{6.20}$$

which is 3/2 times larger than the optimal strategy.

The global average Bayesian cost is instead

$$v_{ep} = \int \mathrm{d}c\, p(c)c^2 - \sum_{k=0}^{N} p(k)c(k)^2 = \frac{(d-1)((d+M)^2+(d+2M)N)}{d(1+d)(d+M)^2(d+N)}. \tag{6.21}$$

In the limit $M \to \infty$, N, d constant we have

$$v_{ep} \sim \frac{(d-1)}{d(d+1)(d+N)}, \tag{6.22}$$

which again coincides with the optimal Bayesian strategy in this limit. In the limit $M+N \to \infty$, $M-N, d$ fixed we have instead

$$v_{ep} \sim \frac{6(d-1)}{d(d+1)(M+N)}, \tag{6.23}$$

which again is 3/2 times larger than the optimal Bayesian strategy.

For the EE measurement, in the limit $M \to \infty$, N, d constant we have a mean square error for the local estimator which scales asymptotically as

$$v_{ee}(c) \sim \frac{2c(1-c)}{N}, \tag{6.24}$$

which is twice as large as the optimal strategy in the leading order of N. In the limit $M+N \to \infty$, $M-N, d$ fixed we have instead

$$v_{ee}(c) \sim \frac{8c(1-c)}{(M+N)}, \tag{6.25}$$

which is 2 times larger than the optimal strategy.

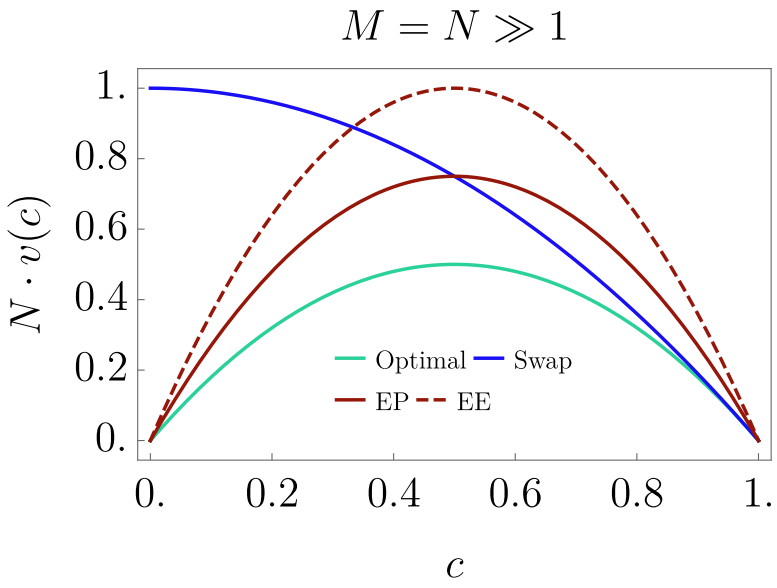


Figure 6.2: Plot of the optimal local MSE scaling coefficient $N \cdot v(c)$ vs. the true value of the overlap c, at leading order in $M = N$, for the strategies analyzed in the chapter.

In the Bayesian case we have instead in the limit $M \to \infty$, N, d constant

$$v_{ee} \sim \frac{(d-1)(d+2N)}{d(d+1)(d+N)^2}, \tag{6.26}$$

which is $(d+2N)/(d+N)$ times larger than the optimal Bayesian strategy. In the limit $M+N \to \infty$, $M-N, d$ fixed we have instead

$$v_{ee} \sim \frac{8(d-1)}{d(d+1)(M+N)}, \tag{6.27}$$

which is 2 times larger than the optimal Bayesian strategy.

Note that these asymptotic behaviours are valid when d is fixed, and it is needed that $N + M >> d$ in order to discard next to leading order terms. This is expected, as covariant estimation is used as a subroutine.

Table (6.1) summarizes the performance of these strategies compared with the optimal one, in the different asymptotic limits.

As we anticipated in the introduction, a very economic strategy to estimate the overlap for multi-qubit states is to use the swap test [Buh+01], which requires only a control qubit, two Hadamard gates and controlled swaps, controlled from the control qubit, which act on pairs of the qubits of the two states. Mathematically, it is a projection on

the completely symmetric subspace of $\mathcal{H}_{2^k}^{\otimes M+N}$ and its orthogonal, hence it coincides with the optimal measurement for $M = N = 1$. As the SWT acts on pairs of states, we restrict to the case $M = N$. The probability of a symmetric projection is $s(c) = \frac{1}{2}(1 + c)$ and the statistics of k successful projections out of N trials is given by the binomial distribution. The optimal local MSE attainable by this test is simply,

$$v_{\mathrm{sw}}(c) = \frac{1 - c^2}{N}, \tag{6.28}$$

while for the optimal global MSE v_{sw} one can derive an exact expression for each value of k, then compute the sum numerically, as detailed in the Supplementary Material of [Fan+20a]. In the asymptotic limit of $M = N \gg d$ a good approximation is provided by averaging the optimal local MSE: $v_{\mathrm{sw}} \simeq \int dc\, p_d(c) v_{\mathrm{sw}}(c) = (d+2)(d-1)/(d(d+1)N)$, which is $O(d)$ times larger than the optimal Bayesian cost v_{op}.

In the same limit, we can compare the local MSE of all the strategies, see Fig. 6.2. First, we observe a gap between the optimal strategy, that attains the QFI, and all the other strategies. This means that, even with a large number of copies, the collective measurement on $|\psi\rangle^{\otimes N} \otimes |\phi\rangle^{\otimes M}$ has a clear advantage over a non-collective one. Second, we observe that the relative error $\frac{\sqrt{v(c)}}{c}$ for small c scales as $\frac{1}{c\sqrt{N}}$ for the SWT and as $\frac{1}{\sqrt{cN}}$ for the other strategies, implying a quadratic improvement in $\frac{1}{\sqrt{c}}$ in the number of copies needed to reach a fixed relative error, while the optimal measurement is still computationally efficient (see Sec.6.6). This is particularly relevant since for large d small overlaps are exponentially more likely, see Eq. (6.12). This phenomenon is also at the source of the so-called "barren plateau" problem [McC+18; Day+19] for quantum variational circuits, and other types of strategies have been proposed to address this issue [GB10; Ben+19; Kha+19].

We notice similar features for the global MSE, plotted in Fig. (6.3) as a function of N for M fixed and increasing d (inset). We observe that the SWT is comparable with EE for $M \sim N$ and $d = 2$, but with a small increase in dimension this feature disappears. Moreover, there is in general a gap between the EP and EE strategies, the former being closer to the optimal one.

Note that all these strategies except the optimal one require labeling of the states.

6.4 Measurement invasiveness

Another relevant figure of merit for applications is the fidelity between the post-measurement state and the initial one, averaged over the measurement outcomes. Both the optimal measurement and the SWT are projective measurements. We assume that the post-measurement states are given by the result of such projections and hence the

Local est.	$\mathbf{v}_{\rm op}(\mathbf{c})$	$\mathbf{v}_{\rm ep}(\mathbf{c})$	$\mathbf{v}_{\rm ee}(\mathbf{c})$
$M=N\to\infty$	$\frac{4c(1-c)}{M+N}$	$\frac{3}{2}v_{\rm op}(c)$	$2v_{\rm op}(c)$
$M\to\infty$	$\frac{c(1-c)}{N}$	$v_{\rm op}(c)$	$2v_{\rm op}(c)$
Bayesian est.	$\mathbf{v}_{\rm op}$	$\mathbf{v}_{\rm ep}$	$\mathbf{v}_{\rm ee}$
$M=N\to\infty$	$\frac{4(d-1)}{d(d+1)(M+N)}$	$\frac{3}{2}v_{\rm op}$	$2v_{\rm op}$
$M\to\infty$	$\frac{(d-1)}{d(d+1)(d+N)}$	$v_{\rm op}$	$\frac{d+2N}{2+N}v_{\rm op}$

Table 6.1: Local MSE and global MSE attainable via the optimal, EP and EE strategies in two asymptotic limits. In all the cases the global MSEs coincide with the corresponding average local MSE values, apart from asymptotically vanishing corrections.

average post-measurement fidelity can be written as

$$F(c)=\int_{U\in{\rm SU}(d)}dU\sum_k|\langle\Psi_U|\,E_k\,|\Psi_U\rangle|^2, \tag{6.29}$$

with $\{E_k=\Pi_J\}$ for the optimal measurement and $\{E_k=\frac{1}{n!}\sum_{\sigma\in S_N}\mathsf{s}_N(\sigma)\otimes\mathsf{s}_N(\sigma)[(\Pi_2^S)^{\otimes k}\otimes(\Pi_2^A)^{\otimes N-k}]\mathsf{s}_N^\dagger(\sigma)\otimes\mathsf{s}_N^\dagger(\sigma)\}$ for the SWT, where $\Pi_2^{S/A}$ are the projectors on the completely symmetric subspace and its orthogonal in $\mathcal{H}_d^{\otimes 2}$. For the optimal measurement $\{\Pi_J\}_J$, Eq. (6.29) evaluates to

$$F_{op}=\sum_{J=J_{\rm min}}^{J_{\rm max}}|\langle\Psi|\,\Pi_J\,|\Psi\rangle|^2=\sum_J P_{M,N}(J|c)^2, \tag{6.30}$$

For the swap test we restrict to $M=N$ and we consider that the measurement is separable and identical on each pair of copies. Moreover, the measurement on a single pair SU(d)-invariant, and succeeds/fails with probability $(1\pm c)/2$. Hence we have

$$F_{sw}=\left[\left(\frac{1+c}{2}\right)^2+\left(\frac{1-c}{2}\right)^2\right]^N=\left(\frac{1+c^2}{2}\right)^N. \tag{6.31}$$

In Fig. 6.3 we plot these two quantities as a function of c, showing that the optimal measurement is less invasive than the SWT, especially for small overlap values.

6.5 Noise-robustness

We also considered how the optimal strategy changes when the states, which are expected to be pure, are affected by depolarizing noise acting independently on each qudit before reaching the measurement stage. Note that if the noisy channel is of a different kind, one

can at least reach the optimal MSE for the depolarizing channel by performing a twirling operation, realizable by pre- and post-processing with random unitaries on each qudit plus classical forward communication. This operation is $\int dU U^\dagger \mathcal{N}(U\rho U^\dagger)U = \Phi_r[\rho]$ for some r, where Φ_r is the depolarizing channel, $\Phi_r[\rho] = r\rho + (1-r)\frac{I}{d}$. After this operation the overall state of the system can now be written as $\Phi_{r_0}[\psi]^{\otimes N} \otimes \Phi_{r_1}[\phi]^{\otimes M}$.

In Appendix A.5 we sketch the computation of optimal MSE in this case, restricting to $d = 2$. Apart from some analytic computation to simplify the expressions, we again use lemma A.3.1 to compute the asymptotics. In the limit $M, N \to \infty$ with $\frac{M}{N}$ finite, the global MSE at leading order is $v_{\text{op,mix}} = \frac{1}{6Mr_0^2} + \frac{1}{6Nr_1^2}$, which agrees with the previously found limit of Eq. (6.15) for zero-noise, $r_i = 1$. Hence the net effect of white noise is to rescale the MSE by a factor r_i^{-2} for each state.

6.6 Gate complexity and noisy implementations

The advantage in the precision of the optimal estimation comes with the tradeoff that the optimal measurement requires entangling operations over the whole system of $N + M$ qudits. As already remarked, the weak Schur transform [Kro19; Har05b] is a way to perform the optimal measurement, and requires $O(\text{poly}(N + M, \log d, \log \frac{1}{\epsilon}))$ qudit gates for precision ϵ. The resulting algorithm is efficient, but still unfeasible without error correction. The SWT instead requires N independent circuits of fixed depth, and may still be convenient for large overlaps or very noisy gates.

A mid-term solution is to divide input data in R groups of S copies of $|\phi\rangle$ and $|\psi\rangle$, such that S is the largest integer for which the given architecture can perform the optimal measurement with high fidelity, repeat the measurement R times and do classical post-processing. The performances of these intermediate protocols are between SWTs and optimal measurement. In this section we sketch an evaluation of the effect of imperfect gates on the accuracy of the estimate of the overlap. First of all we model the error of each iteration of the swap test as white noise for each iteration: $\mathcal{N}_{sw}(s(c)) = (1 - \epsilon_{sw})s(c) + \epsilon_{sw}\frac{1}{2}$, the Fisher information becomes

$$H(\mathcal{N}_{sw}(s(c))) = \frac{(1-\epsilon_{sw})^2}{1 - c^2(1-\epsilon_{sw})^2}, \tag{6.32}$$

For N repetitions, one gets a resulting MSE

$$v_{sw,noisy}(c) = \frac{1 - c^2(1-\epsilon_{sw})^2}{(1-\epsilon_{sw})^2 N}. \tag{6.33}$$

We model the noise on the Schur transform measurement outcomes also as mixing with a probability distribution $q(c)$: $\mathcal{N}_{sw}(P_{M,N}(J|c)) = (1 - \epsilon_{Sch})P_{M,N}(J|c) + \epsilon_{Sch}q(c)$, with a

probability of mixing that scales exponentially in the number of gates, $1-\epsilon_{Sch} \approx (1-\epsilon)^g$, where ϵ is the error per gate, and g is the total number of gates. We recall the joint convexity property of the Fisher Information, coming from its monotonicity:

$$F(\lambda p(c) + (1-\lambda)q(c)) \leq \lambda F(p(c)) + (1-\lambda)F(q(c)), \tag{6.34}$$

If we assume $q(c)$ to be overlap independent, we obtain the bound

$$F(\mathcal{N}_{Sch}(P_{M,N}(J|c))) \leq (1-\epsilon_{Sch})F(P_{M,N}(J|c)), \tag{6.35}$$

so that

$$v_{Sch,noisy}(c) \geq \frac{2c(1-c)}{(1-\epsilon_{Sch})N}. \tag{6.36}$$

This is a very conservative estimate, as we are assuming we are acquiring useful information with exponentially small probability. Hence the Swap test outperforms our optimal strategy, based on the Schur transform, when the respective implementation errors satisfy the following relation

$$\frac{(1-\epsilon_{Sw})^2}{(1-c^2)(1-\epsilon_{Sw}^2)} \geq \frac{1-\epsilon_{Sch}}{2c(1-c)}. \tag{6.37}$$

One can express ϵ_{Sch} and ϵ_{Sw} in terms of the error per gate, ϵ, raised to gate complexity of their respective circuits. An intermediate strategy could be to divide the N copies of both $|\phi\rangle$ and $|\psi\rangle$, into R groups of S copies, and perform the optimal measurement on each group, followed by classical post-processing. If $N = M = RS$ and $F(J|c,S)$ is the optimal Fisher information for the case with $M = N = S$ copies, the Cramér-Rao bound reads

$$v(c) \geq \frac{1}{RF(J|c,S)}. \tag{6.38}$$

The best option would be to choose S as the highest number of copies such that the architecture can perform the optimal measurement in a sufficiently precise way. On the other hand, if one requires to be in the asymptotic regime of the approximation for $c > c_0$, one can just find the minimum S for which the approximation works, and perform the optimal measurement with S copies R times. The classical post processing will have the optimal asymptotic performance for $c > c_0$. In any case the bound (6.38) is asymptotically achieved by a maximum likelihood estimator when $R \to \infty$.

6.7 Optimal global mean squared error

In this section we derive the optimal estimator and corresponding global average mean squared error (AvMSE) for the case where the overlap c is a random variable with a distribution induced by the Haar-uniform measure of SU(d), proving Proposition 6.2.3.

While we could directly work with the classical probability distribution $P_{M,N}(J|c)$ and the the prior probability distribution,

$$p_d(c) = \int_{\mathrm{SU}(d)} \mathrm{d}U\, \delta(c - |\langle\psi| U |\psi\rangle|^2) = (d-1)(1-c)^{d-2}, \tag{6.39}$$

we can simplify the calculations following the more general treatment in [Per71] and recalled as Theorem 2.4.4. We already know that the optimal observable is a post-processing of projectors $\{\Pi_J\}$, and that probability of outcome J is $P_{M,N}(J|c)$. Applying our case to Theorem 2.4.4, we consider the two operators

$$\begin{aligned} \Gamma &:= \int p_d(c)\, \rho(c)\, \mathrm{d}c \\ \eta &:= \int c\, p_d(c)\, \rho(c)\, \mathrm{d}c. \end{aligned} \tag{6.40}$$

We can evaluate Γ as

$$\begin{aligned} \Gamma &= \int_0^1 \mathrm{d}c p_d(c) \left(\int_{\mathrm{SU}(d)} \mathrm{d}U \left(U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \otimes \left(UW(c) |\psi\rangle\langle\psi| W^\dagger(c) U^\dagger \right)^{\otimes M} \right) \\ &= \int_{\mathrm{SU}(d)} \mathrm{d}U \left(U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \\ &\otimes \int_0^1 \mathrm{d}c \int_{\mathrm{SU}(d)} \mathrm{d}V\, \delta(c - |\langle\psi| V |\psi\rangle|^2) \left(UW(c) |\psi\rangle\langle\psi| W^\dagger(c) U^\dagger \right)^{\otimes M} \\ &= \int_{\mathrm{SU}(d)} \mathrm{d}U \left(U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \otimes \int_{\mathrm{SU}(d)} \mathrm{d}V \left(UW_V V |\psi\rangle\langle\psi| V^\dagger W_V^\dagger U^\dagger \right)^{\otimes M} \\ &= \int_{\mathrm{SU}(d)} \mathrm{d}U \left(U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \otimes \int_{\mathrm{SU}(d)} \mathrm{d}V \left(V |\psi\rangle\langle\psi| V^\dagger \right)^{\otimes M} \\ &= \frac{I_{\lambda_{N,N/2}}}{\omega^{(d)}_{\lambda_{N,N/2}}} \otimes I^{(N)}_{\lambda_{N,N/2}} \otimes \frac{I_{\lambda_{M,M/2}}}{\omega^{(d)}_{\lambda_{M,M/2}}} \otimes I^{(M)}_{\lambda_{M,M/2}} \\ &= \frac{1}{\omega^{(d)}_{\lambda_{N,N/2}} \omega^{(d)}_{\lambda_{M,M/2}}} \sum_{J=J_{\min}}^{J_{\max}} I_{\lambda_{N+M,J}} \otimes |J_{N,M}\rangle\langle J_{N,M}|, \end{aligned} \tag{6.41}$$

where in the third equality we have made use of the fact that there always exists a unitary a unitary W_V such that $W_V |\psi\rangle \langle\psi| W_V^\dagger = |\psi\rangle \langle\psi|$ and $W_V V = W(|\langle\psi| V |\psi\rangle|^2)$, which we can insert for free by the invariance of the Haar measure. In the fourth equality we integrate over c and use the invariance of the Haar measure again to decouple the two integrals. In the next equality we used the fact that coherent states of $\mathrm{S}U(d)$ are a resolution

of the projector on the completely symmetric subspace. Here $\lambda_{l,j} := (\frac{h+j}{2}, \frac{h-j}{2}, 0, ..., 0)$, $I_{\lambda_{N,N/2}}$ is the projector on $\mathcal{U}_{\lambda_{N,N/2}}(\mathrm{SU}(d))$ and $I^{(N)}_{\lambda_{N,N/2}}$ is the projector on $\mathcal{V}_{\lambda_{N,N/2}}(S_N)$ (which is one dimensional), and similarly for M. In the last equality we used that the multiplicity is one for each irreducible representation of $\mathrm{SU}(d)$ in the coupling of two irreducible representation of Young diagrams with one row is one, and the multiplicity vector is determined by the form of $\rho(c)$, Eq. (6.6).

To compute η we make use of the following identity, written for multi-qubit states

$$\begin{aligned}
&\int_{\mathrm{SU}(2)} \mathrm{d}g(d-1)\left(1-|D^{\left(\frac{1}{2}\right)}_{\frac{1}{2},\frac{1}{2}}(g)|^2\right)^{d-2} |D^{\left(\frac{1}{2}\right)}_{\frac{1}{2},\frac{1}{2}}(g)|^2 \left(|0\rangle\langle 0|\right)^{\otimes N} \\
&\otimes \left(D^{\left(\frac{1}{2}\right)}(g)^\dagger |0\rangle\langle 0| D^{\left(\frac{1}{2}\right)}(g)\right)^{\otimes M} \\
&= \int_{\mathrm{SU}(2)} \mathrm{d}g D^{\frac{d-1}{2}}_{-\frac{d-3}{2},\frac{d-1}{2}}(g) D^{\frac{d-1}{2}}_{-\frac{d-3}{2},\frac{d-1}{2}}(g)^* D^{\frac{M}{2}}_{k,\frac{M}{2}}(g) D^{\frac{M}{2}}_{k',\frac{M}{2}}(g)^* \\
&\left|\tfrac{N}{2},\tfrac{N}{2}\right\rangle\left\langle\tfrac{N}{2},\tfrac{N}{2}\right| \otimes \left|\tfrac{M}{2},k\right\rangle\left\langle\tfrac{M}{2},k'\right| \qquad (6.42)\\
&= \frac{1}{d+M}\sum_{k=-\frac{M}{2}}^{J-\frac{N}{2}} \left(C^{\frac{d-1+M}{2},-\frac{d-3}{2}+h}_{\frac{d-1}{2},-\frac{d-3}{2},\frac{M}{2},k} C^{J,\frac{N}{2}+k}_{\frac{N}{2},\frac{N}{2},\frac{M}{2},k}\right)^2 \left|J,\frac{N}{2}+k\right\rangle\left\langle J,\frac{N}{2}+k\right|, \qquad (6.43)
\end{aligned}$$

with $D^j_{m,n}(g)$ being Wigner matrices, so that

$$\begin{aligned}
\eta &= \int_0^1 p_d(c) c \left(\int_{\mathrm{SU}(d)} \mathrm{d}U \left(U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \otimes \left(UW(c) |\psi\rangle\langle\psi| W^\dagger(c) U^\dagger \right)^{\otimes M} \right) \mathrm{d}c \\
&= \int_{\mathrm{SU}(d)} \mathrm{d}U \left(U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \\
&\otimes \int_0^1 \mathrm{d}c\, p_d(c)\, c \int_{\mathrm{SU}(2)} \mathrm{d}V \delta(c - |\langle\psi| V |\psi\rangle|^2) \left(UW(c) |\psi\rangle\langle\psi| W^\dagger(c) U^\dagger \right)^{\otimes M} \qquad (6.44) \\
&= \int_{\mathrm{SU}(d)} \mathrm{d}U \left(U |\psi\rangle\langle\psi| U^\dagger \right)^{\otimes N} \int_{\mathrm{SU}(2)} \mathrm{d}V (d-1)(1 - |\langle\psi| V |\psi\rangle|^2)^{d-2} |\langle\psi| V |\psi\rangle|^2 \\
&\otimes \left(UW(|\langle\psi| V |\psi\rangle|^2) |\psi\rangle\langle\psi| W^\dagger(|\langle\psi| V |\psi\rangle|^2) U^\dagger \right)^{\otimes M} \\
&= \int_{SU(d)} U^{\otimes N+M} \otimes \left[\int_{\mathrm{SU}(2)} \mathrm{d}V (d-1)(1 - |\langle\psi| V |\psi\rangle|^2)^{d-2} |\langle\psi| V |\psi\rangle|^2 |\psi\rangle\langle\psi|^{\otimes N} \right. \\
&\otimes \left. \left(V |\psi\rangle\langle\psi| V^\dagger \right)^{\otimes M} \right] U^{\dagger \otimes N+M} \qquad (6.45) \\
&= \frac{1}{d+M} \sum_{J=J_{\min}}^{J_{\max}} \sum_{k=-J-\frac{N}{2}}^{J-\frac{N}{2}} \left(C^{\frac{d-1+M}{2}, -\frac{d-3}{2}+k}_{\frac{d-1}{2}, -\frac{d-3}{2}; \frac{M}{2}, k} C^{J, \frac{N}{2}+k}_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k} \right)^2 \frac{I_{\lambda_{N+M,J}}}{\omega^{(d)}_{\lambda_{N+M,J}}} \otimes |J_{N,M}\rangle\langle J_{N,M}| .
\end{aligned}$$

For an observable of the form $S = \sum_J \tilde{c}(J) \Pi_J$ giving the optimal mean squared error, we have the unique solution

$$\tilde{c}(J) = \frac{\mathrm{tr}[\Pi_J \eta]}{\mathrm{tr}[\Pi_J \Gamma]} = \frac{\frac{1}{d+M} \sum_{k=-J-\frac{N}{2}}^{J-\frac{N}{2}} \left(C^{\frac{d-1+M}{2}, -\frac{d-3}{2}+k}_{\frac{d-1}{2}, -\frac{d-3}{2}; \frac{M}{2}, k} C^{J, \frac{N}{2}+k}_{\frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k} \right)^2}{\frac{\omega^{(d)}_{\lambda_{N+M,J}}}{\omega^{(d)}_{\lambda_{N,N/2}} \omega^{(d)}_{\lambda_{M,M/2}}}}, \qquad (6.46)$$

$\omega^{(d)}_{\lambda_{N+M,J}}$ can be calculated with the Hook formula [Hay17b]. For Young diagrams with two rows one has

$$\omega^{(d)}_{\lambda_{N+M,J}} = (2J+1) \frac{\left(d + J + \frac{N+M}{2} - 1\right)! \left(d - J + \frac{N+M}{2} - 2\right)!}{(d-1)!(d-2)! \left(\frac{N+M}{2} + J + 1\right)! \left(\frac{N+M}{2} - J\right)!}.$$

To simplify the numerator we employ the graphical calculus techniques

in [VMK88]:

$$\frac{1}{d+M}\sum_{k=-J-\frac{N}{2}}^{J-\frac{N}{2}}\left(C^{\frac{d-1+M}{2},-\frac{d-3}{2}+k}_{\frac{d-1}{2},-\frac{d-3}{2};\frac{M}{2},k}C^{J,\frac{N}{2}+k}_{\frac{N}{2},\frac{N}{2};\frac{M}{2},k}\right)^2$$

$$=(2J+1)\sum_{L=\frac{d-3+N}{2}}^{\frac{d-1+N}{2}}\left(C^{L,\frac{d-3+N}{2}}_{\frac{d-1}{2},\frac{d-3}{2};\frac{N}{2},\frac{N}{2}}\right)^2\begin{Bmatrix}\frac{M}{2} & \frac{d-1}{2} & \frac{d+n-1}{2}\\ L & J & \frac{N}{2}\end{Bmatrix}^2 \tag{6.47}$$

$$=(d-1)(2J+1)(4d+4J+4J^2+2N-N^2+2M+2NM-M^2)$$

$$\times\frac{N!M!(d-1+J+\frac{N+M}{2})!(d-2-J+\frac{N+M}{2})!}{4(d+N)!(d+M)!(-J+\frac{N+M}{2})!(1+J+\frac{N+M}{2})!} \tag{6.48}$$

where the term in curly brackets is the Wigner 6-j symbol. Plugging everything together the optimal AvMSE estimator for a given measurement outcome J is given by

$$\tilde{c}^{bay}_{opt}(J)=\frac{d+J+J^2+\frac{M+N}{2}-\left(\frac{M+N}{2}\right)^2+MN}{(d+M)(d+N)}, \tag{6.49}$$

with its corresponding AvMSE

$$v_{\text{op}}=\langle(\tilde{c}^{bay}_{op}-c)^2\rangle=\int_0^1 p(c)c^2-\sum_J p(J)c(J)^2=\int_0^1 p(c)c^2-\sum_J \text{tr}[\Pi_J\Gamma]c(J)^2$$

$$=\frac{(d-1)(d+M+N)}{d(1+d)(d+M)(d+N)}. \tag{6.50}$$

6.8 Remarks

In this chapter we have computed the ultimate precision attainable in estimating the overlap of two arbitrary pure quantum states, as a function of the dimension of their Hilbert space and their number of copies. We showed that the commonly used SWT is highly inefficient for small values of the overlap and also on average over Haar-distributed random states. The optimal strategy is a collective measurement on all the copies and can be implemented efficiently using the Schur transform, although it remains experimentally challenging. A practical alternative is to do Schur sampling on subsets of the dataset, followed by classical post-processing. In addition, we proposed two intuitive strategies that estimate separately one or both states and showed that they also outperform the SWT. Finally, we showed that the optimal measurement is less invasive than the SWT and robust to white noise. It would be important to understand the limits to the estimation of closeness measures of sets of general mixed states. We already pointed out how to estimate a class of such quantities, based only on the spectra of the states and on their convex combinations, in Sec. 4.2.2. The Hilbert-Schmidt distance falls in

one example [BOW19] (see also chapter 7), while the trace distance is an important exception, which does not fall in this subset. As for the programmable discrimination problem, pointing out precise limits on the accuracy of the estimation of closeness measures could be out of reach beyond the simple pure state case. However, it would be already interesting and non-trivial, to understand how the necessary number of copies of the states scales with the dimension, for each closeness measure.

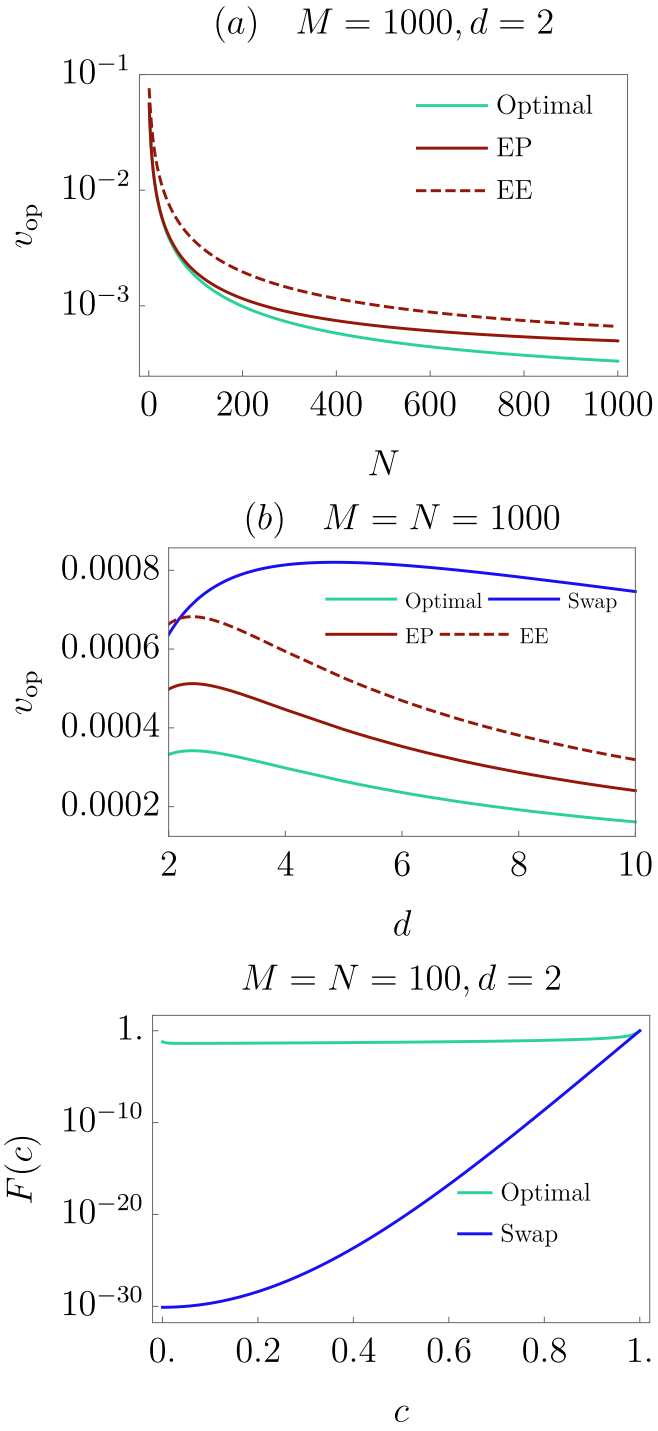


Figure 6.3: (a) Plot of the optimal global MSE v_{opt} vs. the number of copies of one state N, for a fixed number of copies of the other $M = 1000$, in dimension $d = 2$, for the optimal, EP and EE strategies. (b) Plot of the optimal global MSE v_{opt} vs. the dimension d, for a fixed number of copies $M = N = 1000$ for all the strategies studied. (c) Plot of the average post-measurement fidelity with the initial state $F(c)$ vs. the true value of the overlap c with a fixed and equal number of copies $M = N = 100$, for the optimal strategy and SWT.

Chapter 7

Identity testing of collection of quantum states

This chapter is largely based on:

- Marco Fanizza, Raffaele Salvia, and Vittorio Giovannetti. *Testing identity of collections of quantum states: sample complexity analysis*. 2021. arXiv: 2103.14511.

7.1 Introduction

In this chapter we take another look at unitarily invariant quantities, this time from the point of view of property testing [Gol17a; Can20; MW16] and sample complexity. We introduced the concept of finite size effects in hypothesis testing in Sec. 2.4.3. The problem we consider in this chapter is testing if a collection of N d-dimensional states is such that the states are all equal or they are far from being equal in some motivated closeness measure. As in previous chapters, we do not have access to a full classical description of the states but we receive copies of the states. We assume a sample access to these copies: each time the agent request a state, the agent receives a labeled state ρ_i with probability p_i. As mentioned in Sec. 1.1, this model is effective to picture a scenario where a preparation device is modeled as a source of states which are in principle different, each state corresponding to a different measurement outcome of the preparation stage. The provider of the preparation device would like to certify that different outcomes of the preparation stage give equivalent states. At variance with previous chapters, in this case we are not interested in computing exactly the optimal error probability, but rather to understand how the necessary number of copies to answer successfully with

high probability depends on the extensive parameters of the problem, that is N and d. Our analysis of the problem adapts techniques from [BOW19], dealing with the case of certification of two unknown states with the same number of copies, and [Gol17b; DK16], which solved the corresponding classical problem of testing identity between classical distributions. The measurement for the test is again weak Schur sampling, in the nested variant discussed in Sec. 4.2.4.

The chapter is structured as follows: we present the model and state the results in Sec. 7.1.1. We mention related work in Sec. 7.1.2. We discuss the properties of the distance measures we will need in Sec. 7.2. Sec. 7.3 and Sec. 7.4 are devoted to proofs of the main statements. Sec. 7.5 shows how to implement the test with nested weak Schur sampling, while Sec. 7.6 makes final remarks on the relation between the problem that we consider and testing independence.

7.1.1 Results

Given a collection of d-dimensional quantum states $\{\rho_i\}_{i=1,\dots,N}$, and a probability distribution p_i $(0 < p_i < 1)$, we consider a *sampling model* [Gol17b; DK16] where we have access to M copies of the density matrix

$$\rho = \sum_{i=1}^{N} p_i \, |i\rangle\langle i| \otimes \rho_i, \tag{7.1}$$

where $\{|i\rangle\}_{i=1,\dots,N}$ is an orthonormal basis of a N dimensional (classical) register. We are promised that one of the two following properties holds:

- **Case** A: $\rho_1 = \rho_2 = ... = \rho_N$, which can be equivalently stated by saying that there exists a d-dimensional state σ such that $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) = 0$;
- **Case** B: For any d-dimensional state σ it holds $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) > \epsilon$.

Our goal is to find the values of M for which there is a two-outcome test that can discriminate the two cases with high probability of success. Explicitly, indicating with `"accept"` and `"reject"` the outcomes of the test, we require the probability of getting `"accept"` to be larger than 2/3 in case A, and smaller than 1/3 in case B, i.e.

$$\begin{cases} P(\texttt{test} \mapsto \texttt{"accept"} \,|\text{Case A}) > 2/3 \,, \\ \\ P(\texttt{test} \mapsto \texttt{"accept"} \,|\text{Case B}) < 1/3 \,. \end{cases} \tag{7.2}$$

Note that the values 2/3 and 1/3 are entirely conventional, and can be replaced by any constant respectively in $(1/2, 1)$ and $(0, 1/2)$. The main result of the paper is to provide an estimate of necessary and sufficient values of M to fulfill the above conditions. We use

the notations $O(f(d,N,\epsilon))$ and $\Omega(g(d,N,\epsilon))$ to indicate respectively asymptotic upper and lower bounds to sample complexities. If lower and upper bounds which differ by a multiplicative constant can be obtained, the sample complexity is considered to be determined and indicated as $\Theta(f(d,N,\epsilon)) = \Theta(g(d,N,\epsilon))$.

Specifically we prove the following results:

Theorem 7.1.1. *Given access to $O(\frac{\sqrt{N}d}{\epsilon^2})$ samples of the density matrix ρ of Eq. (1), there is an algorithm which can distinguish with high probability whether $\sum_i p_i D_{\mathrm{Tr}}(\rho_i,\sigma) > \epsilon$ for every state σ, or there exists a state σ such that $\sum_i p_i D_{\mathrm{Tr}}(\rho_i,\sigma) = 0$ (that is, all the states ρ_i are equal).*

Theorem 7.1.2. *Any algorithm which can distinguish with high probability whether $\sum_i p_i D_{\mathrm{Tr}}(\rho_i,\sigma) > \epsilon$ for every state σ, or there exists a state σ such that $\sum_i p_i D_{\mathrm{Tr}}(\rho_i,\sigma) = 0$ (that is, all the states ρ_i are equal), given access to M copies of the density matrix ρ of Eq. (1), requires at least $M = \Omega(\frac{\sqrt{N}d}{\epsilon^2})$ copies.*

The proof of Theorem 7.1.2 is presented in Sec. 7.4 and it relies on the fact that a test working with M copies could be used to discriminate between two states which are close in trace distance unless $M = \Omega(\frac{\sqrt{N}d}{\epsilon^2})$. These states are obtained as the average input ρ of the form of Eq. (7.1) for two different sets of collections of states: in the first case the set is made of only one collection consisting in completely mixed states (thus satisfying case A), and in the second the set of collections is such that its elements satisfy case B with high probability. The derivation of the upper bound for M given in Theorem 7.1.1 is instead presented in Sec. 7.3 and it is obtained by constructing an observable $\mathcal{D}_M$ whose expected value is the mean squared Hilbert-Schmidt distance between the states ρ_i, and we bound the variance of the estimator. By relating the mean squared Hilbert-Schmidt distance to $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sum_i p_i \rho_i)$ we obtain the test of the theorem. The analysis exploits a *Poissonization* trick [Gol17b] where the number of copies M is not fixed but a random variable, extracted from a Poisson distribution with average μ, $\mathrm{Poi}_\mu(M) := \frac{e^{-\mu}\mu^M}{M!}$ (summarized later on by the notation $M \sim \mathrm{Poi}_\mu$). We then look for a test which can be performed by a two-outcome POVM $\{E_0^{(M)}, E_1^{(M)}\}$ for each M. This is a standard technique that allows for some useful simplification of the analysis by getting rid of unwanted correlations (more on this in Sec. 7.3.1). The equivalence of the Poisson model with the original one is formalised in Appendix A.6.

Analogously to [BOW19] we can refine the upper bound when the states in the collection have low rank. Given the state ρ of Eq. (7.1), we define its reduced average density matrix

$$\bar{\rho} := \sum_{i=1}^{N} p_i \rho_i \ , \tag{7.3}$$

In particular, when $\bar{\rho}$ is η-close to rank k, that is, the sum of its k largest eigenvalues is at least $1-\eta$, we can refine Theorem 7.1.1:

Theorem 7.1.3. *If the density matrix $\bar{\rho}$ of Eq. (7.3) is η-close to rank k, given access to $O(\frac{\sqrt{N}k}{\epsilon^2})$ samples of ρ there exists an algorithm which can distinguish with high probability whether $\sum_i p_i D_{\mathrm{Tr}}(\rho_i, \sigma) > \epsilon + \eta$ for every state σ, or there exists a state σ such that $\sum_i p_i D_{HS}(\rho_i, \sigma) < 8(2-\sqrt{2})\epsilon$.*

7.1.2 Related work

Classical distribution testing

For an overview of learning properties of a classical distribution in the spirit of property testing, we refer to [Gol17a; Can20]. We report a partial list of results which are of direct interest for this chapter, about testing symmetric properties of distribution in variational distance. We use the notation $[d]$ for the set $\{1, ..., d\}$. Learning a classical distribution over $[d]$ in total variation distance requires $\Theta(d/\epsilon^2)$ samples [Gol17a], therefore the interest in testing properties is to get a sample complexity $o(d)$. The problem of testing uniformity was addressed in [GR11] and established to be $O(\sqrt{d}/\epsilon^2)$ in successive works [Pan08; VV14]. More generally, the sample complexity of identity testing to a known distribution has been established to be $\Theta(\sqrt{d}/\epsilon^2)$ [VV14; DKN15]. Identity testing for two unknown distributions is $O(\max(d^{1/2}/\epsilon^2, d^{2/3}/\epsilon^{4/3}))$ [Cha+14]. The problem of testing identity of collection of N distributions was introduced in the classical case in [Gol17b] and solved in [DK16], obtaining $\Theta(\max(\sqrt{dN}/\epsilon^2, d^{2/3}N^{1/3}/\epsilon^{4/3}))$ for the sampling model, where at each sample the tester receives one of N distributions with probabilty p_i, and $\Theta(\max(\sqrt{d}/\epsilon^2, d^{2/3}/\epsilon^{4/3}))$ for the query model, where the tester can choose the distribution to call at each sample. A problem related to testing identity of collections is testing independence of a distribution on $\times_{i=1}^{l}[n_i]$, which was addressed by [Bat+01; Gol17b; AD15] and solved in [DK16], which showed a tight sample complexity $\Theta(\max_j(\prod_{i=1}^{l} n_i^{1/2}/\epsilon^{1/2}, n_j^{1/3}\prod_{i=1}^{l} n_i^{1/3}/\epsilon^{4/3}))$.

Quantum state testing

We already mentioned several results on the sample complexity of quantum tomography 2.4.3 and spectrum estimation 4.2.4, and recall them here for completeness. It has been shown the reconstruction of the classical description of an unknown state, *quantum tomography*, requires $\Theta(d^2/\epsilon^2)$ copies of the state [Haa+17; OW16; OW17]. These algorithms require spectrum learning as a subroutine [ARS88; KW01; HM02; Chr06; Key06], which has sample complexity $O(d^2/\epsilon^2)$, although a matching lower bound is available only for the empirical Young diagram estimator [OW15]. These results have been refined in the case when the state is known to be close to a state of rank less than k. Quantum

entropy estimation has been studied in [AKG19]. In the review of quantum property testing in [MW16] it is shown that testing identity to a pure state requires $O(1/\epsilon^2)$ copies. Testing identity to the completely mixed state requires $\Theta(d/\epsilon^2)$ copies [OW15], and the same is true for a generic state and for testing identity between unknown states (with refinements if the state can be approximated by a rank k state) [BOW19]. In [BOW19], identity testing between unknown states is done by first estimating their Hilbert-Schmidt distance with a minimum variance unbiased estimator, developing a general framework for efficient estimators of sums of traces of polynomials of states. This improves on the swap test [Buh+01] mentioned in Chapter 6, which can also be used for the same purpose. In all of these cases, the algorithms considered are classical post-processing of the measurement used to learn the spectrum of a state, possibly repeated on nested sets of inputs. This measurement can be efficiently implemented, with gate complexity $O(n, \log d, \log 1/\delta)$ [BCH06; Har05a; Kro19], where n is the number of copies of the state, and δ is the precision of the implementation. Testing identity of collections of quantum states in the query model has been established to be $\Theta(d/\epsilon^2)$ [Yu19], while the sampling model complexity was left open and it is addressed in this chapter. Independence testing (checking if a state is a product state or far from it) is also addressed in [Yu19], obtaining a sample complexity $O(d_1 d_2/\epsilon^2)$, which is tight up to logarithmic factors, using the identity test of [BOW19] for testing independence of a state on $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$; similar results hold for the multipartite case (see also [HT16] for the asymptotic setting). Besides these optimality results being valid if one allows any measurement possible according to quantum mechanics, several results have been obtained in the case in which there are restrictions on the measurements: [BCL20] shows that the sample complexity for testing identity to the completely mixed state with independent but possibly adaptive measurements is $\Omega(d^{4/3}/\epsilon^2)$ and $\Theta(d^{3/2}/\epsilon^2)$ for non-adaptive measurement, while the instance optimal case for the same problem is studied in [CLO21]; [Haa+17] shows that the sample complexity for tomography for non-adaptive measurement is $\Omega(d^3/\epsilon^2)$. Algorithms with Pauli measurements only have been considered [Yu19; Yu20], while a general review of the various approaches with attention to feasibility of the measurement can be found in [KR21].

7.2 Distance measures for collection of distributions

In this chapter we will need several relations between classical and quantum closeness measures. In addition to the trace distance $D_{\mathrm{Tr}}(\rho, \sigma)$ between to states, eq. (2.7) we need the Hilbert-Schmidt distance $D_{\mathrm{HS}}(\rho, \sigma)$

$$D_{\mathrm{HS}}(\rho, \sigma) = ||\rho - \sigma||_2 \, . \tag{7.4}$$

These quantities are connected via the following inequalities

$$\frac{1}{2}D_{\mathrm{HS}}(\rho,\sigma) \leq D_{\mathrm{Tr}}(\rho,\sigma) \leq \frac{\sqrt{d}}{2}D_{\mathrm{HS}}(\rho,\sigma), \tag{7.5}$$

where the first inequality comes from monotonicity of Schatten norms [HJ12] and the second from Cauchy-Schwartz inequality.

For ρ and $\bar{\rho}$ as defined in Eq. (7.1) and (7.3), we introduce the quantity

$$\mathcal{M}_{\mathrm{Tr}}(\rho) := \sum_{i=1}^{N} p_i D_{\mathrm{Tr}}(\rho_i,\bar{\rho}) \leq \frac{1}{2}\sum_{i=1}^{N} p_i \sqrt{dD^2_{HS}(\rho_i,\bar{\rho})}\,. \tag{7.6}$$

We also define the mean squared Hilbert-Schmidt distance of the model as

$$\mathcal{M}_{HS}(\rho) := \left[\sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j D^2_{HS}\left(\rho_i,\rho_j\right)\right]^{1/2}, \tag{7.7}$$

observing that it can be equivalently expressed in terms of $\bar{\rho}$ as

$$\begin{aligned}
\mathcal{M}^2_{HS}(\rho) &:= \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j D^2_{HS}\left(\rho_i,\rho_j\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j \operatorname{Tr}\left[(\rho_i-\rho_j)^2\right] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j \operatorname{Tr}\left[(\rho_i-\bar{\rho}+\bar{\rho}-\rho_j)^2\right] \\
&= 2\sum_{i=1}^{N} p_i \operatorname{Tr}\left[(\rho_i-\bar{\rho})^2\right] - 2\sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j \operatorname{Tr}[(\rho_i-\bar{\rho})(\rho_j-\bar{\rho})] \\
&= 2\sum_{i=1}^{N} p_i D^2_{HS}(\rho_i,\bar{\rho}).
\end{aligned} \tag{7.8}$$

Therefore we can derive the following important inequality

$$\begin{aligned}
\mathcal{M}_{\mathrm{Tr}}(\rho) &= \sum_{i=1}^{N} p_i D_{\mathrm{Tr}}(\rho_i,\bar{\rho}) \leq \frac{1}{2}\sum_{i=1}^{N} p_i \sqrt{dD^2_{HS}(\rho_i,\bar{\rho})} \\
&\leq \frac{1}{2}\sqrt{\sum_{i=1}^{N} p_i}\sqrt{\sum_{i=1}^{N} p_i dD^2_{HS}(\rho_i,\bar{\rho})} = \frac{\sqrt{d}}{2\sqrt{2}}\mathcal{M}_{HS}(\rho)\,,
\end{aligned} \tag{7.9}$$

which will be used in the next section to obtain a test for $\mathcal{M}_{\mathrm{Tr}}(\rho)$ starting from a test for $\mathcal{M}_{HS}(\rho)$.

If the state σ is close to having rank k, in the sense that the sum of its largest k eigenvalues is larger than $1-\eta$, then the following inequality (proven in section 5.4 of [BOW19]) holds

$$D_{\mathrm{Tr}}(\rho,\sigma) \leq \frac{\sqrt{k}}{c} D_{HS}(\rho,\sigma) + \eta \,, \tag{7.10}$$

with $c = 2-\sqrt{2}$. Therefore, in the special case in which the average state $\bar{\rho}$ is η-close to having rank k, the inequality (7.9) can be improved by

$$\begin{aligned}
\mathcal{M}_{\mathrm{Tr}}(\rho) &= \sum_{i=1}^{N} p_i D_{\mathrm{Tr}}(\rho_i,\bar{\rho}) \leq \sum_{i=1}^{N} p_i \left(\sqrt{\frac{k}{c^2} D^2_{HS}(\rho_i,\bar{\rho})} + \eta \right) \\
&= \frac{1}{c} \sum_{i=1}^{N} p_i \sqrt{k D^2_{HS}(\rho_i,\bar{\rho})} + \eta = \frac{1}{c} \sum_{i=1}^{N} \sqrt{p_i} \sqrt{p_i k D^2_{HS}(\rho_i,\bar{\rho})} + \eta \\
&\leq \frac{1}{c} \sqrt{\sum_{i=1}^{N} p_i} \sqrt{\sum_{i=1}^{N} p_i k D^2_{HS}(\rho_i,\bar{\rho})} + \eta = \frac{\sqrt{k}}{c\sqrt{2}} \mathcal{M}_{HS}(\rho) + \eta \,.
\end{aligned} \tag{7.11}$$

In our analysis we will also need the following divergences for classical distributions p,q: the *chi-squared* divergence, defined as $\mathrm{d}_{\chi^2}(p||q) := \sum_i \frac{(p_i-q_i)^2}{p_i}$; the *Kullback-Leibler* divergence, which corresponds the relative entropy in Eq. (2.11) evaluated on states which have are diagonal in the same basis, defined as $\mathrm{d}_{KL}(p||q) := \sum_i p_i \log_2 \frac{p_i}{q_i}$; and the total variational distance, defined as $\mathrm{d}_{TV}(p||q) := \frac{1}{2}\sum_i |p_i - q_i|$, which also corresponds to the trace distance between states which are diagonal in the same basis. The following properties can be found in [CT05; SV16]. From the definition of Kullback-Leibler divergence, it follows that it is additive, i.e.

$$\mathrm{d}_{KL}\left(\prod_{j=1}^{N} p^{(j)} || \prod_{j=1}^{N} q^{(j)} \right) = \sum_{j=1}^{N} \mathrm{d}_{KL}(p^{(j)}||q^{(j)}) \,. \tag{7.12}$$

The total variational distance is related to the Kullback-Leibler divergence by Pinsker's inequality:

$$\mathrm{d}_{TV}(p,q) \leq \sqrt{\frac{1}{2}\mathrm{d}_{KL}(p||q)} \,, \tag{7.13}$$

and the Kullback-Leibler can be bounded in terms of the chi-squared divergence, as:

$$\mathrm{d}_{KL}(p,q) \leq \ln\left[1 + \mathrm{d}_{\chi^2}(p,q)\right] \,. \tag{7.14}$$

7.3 Upper bound on the sample complexity

In order to prove Theorem 7.1.1 here we show a stronger version of such statement, i.e.

Theorem 7.3.1. *Given access to $O(\frac{\sqrt{N}}{\delta})$ samples of the state ρ of Eq. (7.1), for $\delta > 0$ there is an algorithm which can distinguish with high probability whether $\mathcal{M}^2_{HS}(\rho) \leq 0.99\delta$ or $\mathcal{M}^2_{HS}(\rho) > \delta$.*

The connection with Theorem 7.1.1 follows by the relations between the functionals $\mathcal{M}_{HS}(\rho)$ and $\mathcal{M}_{\mathrm{Tr}}(\rho)$ discussed in Sec. 7.2. Specifically we notice that $\mathcal{M}_{\mathrm{Tr}}(\rho) = 0$ (case A) implies $\mathcal{M}_{HS}(\rho) = 0$, while having $\mathcal{M}_{\mathrm{Tr}}(\rho) > \epsilon$ (a constraint that holds in Case B) implies $\mathcal{M}^2_{HS}(\rho) > \frac{8\epsilon^2}{d}$ by Eq. (7.9). Therefore a test satisfying the requests of Theorem 7.1.1 can be obtained by taking the algorithm identified by Theorem 7.3.1 with $\delta = \frac{8\epsilon^2}{d}$. [Incidentally we stress that the test can be performed by a two outcome POVMs $\{E_0^{(M)}, E_1^{(M)}\}$ when the number of copies of ρ is M (for any $M \geq 1$), obtained as projectors on the eigenvectors of the observable $\mathcal{D}_M$ defined in the following with eigenvalues larger or lower than a threshold; therefore, it is of the class of test on which we can apply Proposition A.6.1].

In a completely analogous way, Theorem 7.1.3 follows by calling the algorithm of Theorem 7.3.1 with $\delta = \frac{16(2-\sqrt{2})^2\epsilon^2}{k}$, and using the inequality (7.11).

The reminder of the section is hence devoted to the prove Theorem 7.3.1.

7.3.1 Building the estimator for $\mathcal{M}^2_{HS}$

To prove Theorem 7.3.1 we construct an unbiased estimator for $\mathcal{M}^2_{HS}$, generalizing the estimator of $D^2_{HS}(\rho, \sigma)$ discussed in [BOW19]. We start noticing that via permutations that operate on the quantum registers conditioned on measurements performed on the classical registers, the density matrix $\rho^{\otimes M}$ describing M copies of the state ρ, can be cast in the following equivalent form

$$\rho^{(M)} := \sum_{\vec{m} \in \mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} \, |\vec{m}\rangle\langle\vec{m}| \otimes \rho^{\vec{m}}. \tag{7.15}$$

In this expression the summation runs over all vectors $\vec{m} = (m_1, m_2, \cdots, m_N)$ formed by integers that provide a partition of M (i.e. $m_1 + m_2 + \cdots + m_N = M$); $\mathsf{M}(\vec{m})_{\vec{p},M}$ is the multinomial distribution with M extractions and probabilities $\vec{p} = (p_1, p_2, \cdots, p_N)$, i.e.

$$\mathsf{M}(\vec{m})_{\vec{p},M} := \frac{M!}{m_1!...m_N!} p_1^{m_1} p_2^{m_2} \cdots p_N^{m_N} \; ; \tag{7.16}$$

the vectors $|\vec{m}\rangle = |m_1, m_2, \cdots, m_N\rangle$ form an orthonormal set for the classical registers of the model; while finally

$$\rho^{\vec{m}} := \rho_1^{\otimes m_1} \otimes \rho_2^{\otimes m_2} \otimes ... \otimes \rho_N^{\otimes m_N} \ , \tag{7.17}$$

is a state of the quantum registers with m_i elements initialized into ρ_i, which formally operates on an Hilbert space with tensor product structure $\otimes_{i=1}^{N} \mathcal{H}_i$, with $\mathcal{H}_i = (\mathbb{C}^d)^{\otimes m_i}$, with $m_i = 0, ..., M$. Exploiting the representation of Eq. (7.15) we then introduce the observable

$$\mathcal{D}_M := \sum_{\vec{m} \in \mathcal{P}_M} |\vec{m}\rangle\langle\vec{m}| \otimes \mathcal{D}^{\vec{m},M}, \tag{7.18}$$

with

$$\mathcal{D}^{\vec{m},M} := \sum_{i \neq j} \mathcal{D}_{ij}^{m_i,m_j,M}, \tag{7.19}$$

and

$$\mathcal{D}_{ij}^{m_i,m_j,M} := \frac{m_i(m_i-1)}{\mu^2 p_i} p_j \mathcal{O}_{ii}^{m_i,m_j} + \frac{m_j(m_j-1)}{\mu^2 p_j} p_i \mathcal{O}_{jj}^{m_i,m_j} - 2\frac{m_i m_j}{\mu^2} \mathcal{O}_{ij}^{m_i,m_j}. \tag{7.20}$$

In the above expression $\mu > 0$ is a free parameter that will be fixed later on. The operators $\mathcal{O}_{ij}^{m_i,m_j}$ are defined to be the average of all possible different transpositions S_{m_i,m_j} swapping one of the m_i factors of $\mathcal{H}_i$ with one of the m_j factors of $\mathcal{H}_j$, with i and j possibly equal, i.e.

$$\mathcal{O}_{ij}^{m_i,m_j} := \frac{1}{|S_{m_i,m_j}|} \sum_{S \in S_{m_i,m_j}} S \ . \tag{7.21}$$

Since each transposition is Hermitian, $\mathcal{O}_{ij}^{m_i,m_j}$ is Hermitian too.

The expectation values of $\mathcal{D}_M$ on $\rho^{(M)}$ can be formally computed by exploiting the relation

$$\mathrm{Tr}\Big[\mathcal{O}_{ij}^{m_i,m_j} \rho^{\vec{m}}\Big] = \mathrm{Tr}\Big[\mathcal{O}_{ij}^{m_i,m_j} \rho_i^{\otimes m_i} \otimes \rho_j^{\otimes m_j}\Big] = \mathrm{Tr}[\rho_i \rho_j] \ , \tag{7.22}$$

where the first identity follows from the fact that $\mathcal{O}_{ij}^{m_i,m_j}$ acts not trivially only on registers containing copies of ρ_i and ρ_j. Accordingly for $i \neq j$ we have

$$\mathrm{Tr}\Big[\mathcal{D}_{ij}^{m_i,m_j,M} \rho^{\vec{m}}\Big] = \frac{m_i(m_i-1)}{\mu^2 p_i} p_j \,\mathrm{Tr}\big[\rho_i^2\big] + \frac{m_j(m_j-1)}{\mu^2 p_j} p_i \,\mathrm{Tr}\big[\rho_j^2\big] - 2\frac{m_i m_j}{\mu^2} \,\mathrm{Tr}[\rho_i \rho_j] \ , \tag{7.23}$$

which leads to

$$\begin{aligned} &\mathrm{Tr}\Big[\mathcal{D}_M \rho^{(M)}\Big] \\ &= \sum_{\vec{m} \in \mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} \sum_{i \neq j} \Big(\tfrac{m_i(m_i-1)}{\mu^2 p_i} p_j \,\mathrm{Tr}\big[\rho_i^2\big] + \tfrac{m_j(m_j-1)}{\mu^2 p_j} p_i \,\mathrm{Tr}\big[\rho_j^2\big] - 2\tfrac{m_i m_j}{\mu^2} \,\mathrm{Tr}[\rho_i \rho_j] \Big) \ . \end{aligned} \tag{7.24}$$

To simplify the analysis of the performance of a test based on $\mathcal{D}_M$ we can invoke the equivalence of Proposition A.6.1 between the original model and its Poissonized version where the value of M (and hence the density matrix $\rho^{(M)}$ that is presented to us) is randomly generated with probability $\mathrm{Poi}_\mu(M)$ (notice that the mean value of the distribution is taken equal to parameter μ which enters the definition (7.20) of $D_{ij}^{m_i,m_j,M}$). Defining Γ_M the set of eigenvalues of the observables $\mathcal{D}_M$ (7.18), we then introduce a new estimator $\mathcal{D}$ that produces outputs $X \in \Gamma := \bigcup_M \Gamma_M$ with probabilities

$$P_X := \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \sum_{x\in\Gamma_M} \delta_{x,X} P_x^{(M)} \,, \tag{7.25}$$

where $P_x^{(M)}$ is the probability of getting the outcome x from $\mathcal{D}_M$ when acting on $\rho^{(M)}$.

The following facts can then be proved:

Proposition 7.3.1 (Unbiasedness). *Given $\mathbb{E}[\mathcal{D}] := \sum_{X\in\Gamma} X P_X$ the mean value of the estimator $\mathcal{D}$ we have*

$$\mathbb{E}[\mathcal{D}] = \mathcal{M}_{HS}^2(\rho) \,. \tag{7.26}$$

Proof. From Eq. (7.25) and (7.24) we can write

$$\begin{aligned}
\mathbb{E}[\mathcal{D}] &= \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \sum_{x\in\Gamma_M} x P_x^{(M)} = \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M)\, \mathrm{Tr}\Big[\mathcal{D}_M \rho^{(M)}\Big] \\
&= \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \sum_{\vec{m}\in\mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} \\
&\quad \times \sum_{i\neq j} \Big(\tfrac{m_i(m_i-1)}{\mu^2 p_i} p_j\, \mathrm{Tr}\big[\rho_i^2\big] + \tfrac{m_j(m_j-1)}{\mu^2 p_j} p_i\, \mathrm{Tr}\big[\rho_j^2\big] - 2\tfrac{m_i m_j}{\mu^2}\, \mathrm{Tr}[\rho_i,\rho_j] \Big) \\
&= \sum_{m_1=0}^{\infty} \cdots \sum_{m_N=0}^{\infty} \mathrm{Poi}_{p_1\mu}(m_1) \cdots \mathrm{Poi}_{p_N\mu}(m_N) \\
&\quad \times \sum_{i\neq j} \Big(\tfrac{m_i(m_i-1)}{\mu^2 p_i} p_j\, \mathrm{Tr}\big[\rho_i^2\big] + \tfrac{m_j(m_j-1)}{\mu^2 p_j} p_i\, \mathrm{Tr}\big[\rho_j^2\big] - 2\tfrac{m_i m_j}{\mu^2}\, \mathrm{Tr}[\rho_i,\rho_j] \Big) \,,
\end{aligned} \tag{7.27}$$

where in the second identity we used $\sum_{x\in\Gamma_M} x P_x^{(M)} = \mathrm{Tr}\big[\mathcal{D}_M \rho^{(M)}\big]$, while in the last identity we exploit the fact that under Poissanization the random variables m_i become independent due to the property

$$\sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \mathsf{M}(\vec{m})_{\vec{p},M} = \prod_{i=1}^{N} \mathrm{Poi}_{p_i\mu}(m_i) \,, \tag{7.28}$$

with $\mathrm{Poi}_{p_i\mu}(m_i)$ being a Poisson distribution of mean $p_i\mu$. Equation (7.26) then finally follows from the identities

$$\sum_{m_i=0}^{\infty} m_i \,\mathrm{Poi}_{p_i\mu}(m_i) = \mu p_i \,, \qquad \sum_{m_i=0}^{\infty} \frac{m_i(m_i-1)}{p_i} \,\mathrm{Poi}_{p_i\mu}(m_i) = \mu^2 p_i \,. \tag{7.29}$$

□

Proposition 7.3.2 (Bound on the variance). *The variance of the estimator $\mathcal{D}$, $\mathrm{Var}[\mathcal{D}] := \sum_{X\in\Gamma} P_X (X - \mathbb{E}[\mathcal{D}])^2$, satisfies the inequality*

$$\mathrm{Var}[\mathcal{D}] \leq O\left(\frac{N}{\mu^2}\right) + \frac{16\mathcal{M}_{HS}^2(\rho)}{\mu} \,. \tag{7.30}$$

Proof. See Appendix A.7. □

We can now invoke the modified Chebyshev inequality proved in [BOW19], which we restate with a notation adapted to this work:

Lemma 7.3.1 (Lemma 2.1 of [BOW19]). *Let $\mathbf{X}^{(\mu)}$ be a sequence of unbiased estimators for a number $c > 0$, i.e. $\mathbb{E}[\mathbf{X}^{(\mu)}] = c$ for all n. Assume the variance of $\mathbf{X}^{(\mu)}$ can be bounded as*

$$\mathrm{Var}[\mathbf{X}^{(\mu)}] \leq O\left(\frac{v(c)}{\mu} + \frac{b(c)}{\mu^2}\right), \tag{7.31}$$

and $b(c)$, $v(c)$, $c^2/b(c)$ and $c^2/v(c)$ are non-decreasing functions of c. Then, for any $\theta > 0$, provided that

$$\mu \geq \max\left\{\sqrt{\frac{b(\theta)}{\theta^2}}, \frac{v(\theta)}{\theta^2}\right\} \tag{7.32}$$

one can use $\mathbf{X}^{(\mu)}$ to distinguish with high probability whether $c < 0.99\theta$ or $c > \theta$.

We have now all the ingredients necessary to prove Theorem 7.3.1: in particular the thesis is obtained by applying Lemma 7.3.1 to the sequence of observables $\mathcal{D}$, implicitly depending on μ, estimating $c = \mathcal{M}_{HS}^2(\rho)$. Proposition 7.3.2 tells indeed that the estimators $\mathcal{D}$ satisfy the hypothesis (7.31) of Lemma 7.3.1, with the identifications, $b(c) = N \cdot O(1)$, and $v(c) = 16c$, $\theta = \epsilon$.

7.4 Lower bound on the sample complexity

We now explain the idea for proving the lower bound on M that follows from Theorem 7.1.2. First of all we limit ourselves to even d, since for odd d one can simply use the lower bound for $d-1$. We also choose the probability distribution p to be uniform, $p_i = 1/N$. The case $N = 2$ is a straightforward consequence of the lower bound

in [OW15], which gives a lower bound of $O(d/\epsilon^2)$, noting that with access to M copies of ρ_ϵ one can simulate access to M copies of $\frac{1}{2}\left(\frac{I_d}{d}\otimes|1\rangle\langle 1| + \rho_\epsilon \otimes |2\rangle\langle 2|\right)$:

Lemma 7.4.1 (Corollary 4.3 of [OW15]). *Let ρ_ϵ be a quantum state with $d/2$ eigenvalues equal to $\frac{1+2\epsilon}{d}$ and the other $d/2$ eigenvalues equal to $\frac{1-2\epsilon}{d}$. Then any algorithm that can discern between the states $(I_d/d)^{\otimes M}$ and $\rho_\epsilon^{\otimes M}$ with a probability greater than 2/3 must require $M \geq 0.15d/\epsilon^2$.*

This is a lower bound for any N smaller than a constant, say $N < 10$. Therefore we consider $N \geq 10$ in the following. We define two sets of collections of N quantum states. The first set A contains only one collection, namely a collection where all the states are the completely mixed states. Clearly, the only element of A is a collection satisfying the property of case A. For even d, the second set B contains all the collections of states having $d/2$ eigenvalues equal to $\frac{1+8\epsilon}{d}$ and $d/2$ eigenvalues equal to $\frac{1-8\epsilon}{d}$. This means that all the states in a collection of B can be written as $U_i\rho_0 U_i^\dagger$ for ρ_0 with the prescribed spectrum and U_i arbitrary. If each U_i is drawn independently according to the Haar measure of $\mathrm{SU}(d)$, we show that the elements of B satisfy property B with probability larger than a constant. We also show an upper bound on the trace distance between ρ_A and ρ_B, being respectively M samples for a collection of all completely mixed states and the average input of M samples for collections in B. Explicitly, we have

$$\rho_A = \left(\frac{1}{N}\sum_{i=1}^{N}|i\rangle\langle i|\otimes\frac{I}{d}\right)^{\otimes M}, \tag{7.33}$$

$$\rho_B = \int_{U_1,\ldots,U_N\in\mathrm{SU}(d)} \mathrm{d}U_1....\mathrm{d}U_N \left(\frac{1}{N}\sum_{i=1}^{N}|i\rangle\langle i|\otimes U_i\rho_0 U_i^\dagger\right)^{\otimes M}. \tag{7.34}$$

If a test capable of distinguishing with high probability with case A and B exists, then it can be used to distinguish between ρ_A and ρ_B. Since the probability of success in the latter task has to be lower than what we obtain from the bound on the trace distance, we obtain a lower bound on the sample complexity.

Lemma 7.4.2. *Let $\{\rho_i\}_{i,\ldots,N}$ be a collection of states such that $\frac{1}{N}\sum_{i=1}^{N}||\rho_i - \bar{\rho}||_1 > 4\epsilon$. Then $\frac{1}{N}\sum_{i=1}^{N}||\rho_i - \sigma||_1 > 2\epsilon$ for any σ.*

Proof. Suppose that we have $\frac{1}{N}\sum_{i=1}^{N}||\rho_i - \sigma||_1 \leq 2\epsilon$ for some σ. By convexity of the trace norm, $||\bar{\rho} - \sigma||_1 \leq 2\epsilon$. Then

$$\frac{1}{N}\sum_{i=1}^{N}||\rho_i - \bar{\rho}||_1 = \frac{1}{N}\sum_{i=1}^{N}||\rho_i - \sigma + \sigma - \bar{\rho}||_1 \leq \frac{1}{N}\sum_{i=1}^{N}||\rho_i - \sigma||_1 + ||\sigma - \bar{\rho}||_1 \leq 4\epsilon \tag{7.35}$$

which is a contradiction. □

Lemma 7.4.3. *For $N > 10$, let $\{U_i \rho_0 U_i^\dagger\}_{i,\dots,N}$ be a collection of states in B and ρ as in Eq. (7.1), with $p_i = 1/N$. If each U_i is drawn independently according to the Haar measure of* $\mathrm{SU}(d)$*, the probability of having $\mathcal{M}_{\mathrm{Tr}}(\rho) \geq 4\epsilon$ is at least*

$$\underset{U_1,\dots,U_N \sim \mathbf{U}(d)}{\mathrm{P}} \left(\mathcal{M}_{\mathrm{Tr}}(\rho) > 4\epsilon\right) \geq \frac{11}{15} \,. \tag{7.36}$$

Proof. We denote by $|k\rangle_{k=1,\dots,d}$ a basis of eigenvectors of ρ_0, such that $\langle k| \rho_0 |k\rangle = \frac{1+(-1)^k 8\epsilon}{d}$ and define

$$\Theta := \sum_{k=1}^{d} (-1)^k |k\rangle\langle k| \,. \tag{7.37}$$

We can write

$$\begin{aligned}
\mathcal{M}_{\mathrm{Tr}}(\rho) &= \frac{1}{N}\sum_{i=1}^{N} \|\rho_i - \bar{\rho}\|_1 = \frac{1}{N}\sum_{i=1}^{N} \left\| \rho_i - \frac{1}{N}\sum_{j=1}^{N} U_j \rho_0 U_j^\dagger \right\|_1 \\
&= \frac{1}{N}\sum_{i=1}^{N} \left\| U_i \rho_0 U_i^\dagger - \frac{1}{N}\sum_{j=1}^{N} U_j \rho_0 U_j^\dagger \right\|_1 \\
&= \frac{1}{N}\sum_{i=1}^{N} \left\| \rho_0 - \frac{1}{N}\sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right\|_1 \\
&\geq \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{d} \left| \langle k| \rho_0 - \frac{1}{N}\sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \right| \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{d} (-1)^k \left(\langle k| \rho_0 |k\rangle - \langle k| \frac{1}{N}\sum_{j=1}^{N} U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \right) \\
&= 8\epsilon - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{k=1}^{d} (-1)^k \sum_{j=1}^{N} \langle k| U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \\
&= 8\epsilon - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{d} (-1)^k \langle k| U_i^\dagger U_j \rho_0 U_j^\dagger U_i |k\rangle \\
&= 8\epsilon - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Tr}\left[\hat{\Theta} U_i^\dagger U_j \rho_0 U_j^\dagger U_i \right] ,
\end{aligned} \tag{7.38}$$

where the inequality comes from the monotonicity of the trace norm, applied together with the channel that projects on the orthogonal basis $\{|k\rangle\}$. The expected value of the

latter term of (7.38) is

$$\begin{aligned}
&\mathop{\mathbb{E}}_{U_1,\ldots,U_N\sim\mathbf{U}(d)}\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Tr}\left[\hat{\Theta}U_i^\dagger U_j\rho_0 U_j^\dagger U_i\right]\right]\\
&=\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\mathop{\mathbb{E}}_{U_1,\ldots,U_N\sim\mathbf{U}(d)}\left[\mathrm{Tr}\left[\hat{\Theta}U_i^\dagger U_j\rho_0 U_j^\dagger U_i\right]\right]\\
&=\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}8\epsilon\delta_{ij}=8\frac{\epsilon}{N}\,.
\end{aligned}\tag{7.39}$$

Therefore, using Markov inequality, we can write

$$\mathop{\mathrm{P}}_{U_1,\ldots,U_N\sim\mathbf{U}(d)}\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Tr}\left[\hat{\Theta}U_i^\dagger U_j\rho_0 U_j^\dagger U_i\right]>3\epsilon\right)\leq\frac{8}{3N}\tag{7.40}$$

Combining (7.40) with (7.38), we have

$$\mathop{\mathrm{P}}_{U_1,\ldots,U_N\sim\mathbf{U}(d)}\left(\mathcal{M}_{\mathrm{Tr}}(\rho)>4\epsilon\right)\geq 1-\frac{8}{3N}\geq\frac{11}{15},\qquad N\geq 10\tag{7.41}$$

□

Lemma 7.4.4.

$$D_{\mathrm{Tr}}(\rho_A,\rho_B)\leq 16\frac{\epsilon^2 M}{d\sqrt{N}}\,.\tag{7.42}$$

Proof. We have that

$$D_{\mathrm{Tr}}(\rho_A,\rho_B)=E_{\vec{m}\sim\mathsf{M}_{\vec{p},N,M}}\left[D\left(\left(\frac{I}{d}\right)^{\otimes M},\int_{U_i\in\mathrm{SU}(d)}\mathrm{d}U_1\ldots.\mathrm{d}U_N\bigotimes_{i=1}^{N}\left(U_i\rho_0U_i^\dagger\right)^{\otimes m_i}\right)\right]\tag{7.43}$$

Using Schur-Weyl duality, we can write ρ_A and ρ_B as

$$\left(\frac{I}{d}\right)^{\otimes M}=\bigotimes_{i=1}^{N}\left(\sum_{\lambda\in Y_{m_i,d}}\mathrm{SW}_{I/d}^{m_i}(\lambda)\frac{I_{d(\lambda,m_i)\times d(\lambda,m_i)}}{d(\lambda,m_i)}\right)\tag{7.44}$$

$$\int_{U_i\in\mathrm{SU}(d)}\mathrm{d}U_1\ldots.\mathrm{d}U_N\bigotimes_{i=1}^{N}\left(U_i\rho_0U_i^\dagger\right)^{\otimes m_i}=\bigotimes_{i=1}^{N}\left(\sum_{\lambda\in Y_{m_i,d}}\mathrm{SW}_{\rho_0}^{m_i}(\lambda)\frac{I_{d(\lambda,m_i)\times d(\lambda,m_i)}}{d(\lambda,m_i)}\right),\tag{7.45}$$

where $Y_{m_i,d}$ is a set of Young diagrams and $\mathrm{SW}_\rho^M(\lambda)$ is a probability distribution over Young diagrams which depends only on the spectrum of ρ. Defining

$$\mathfrak{D}_0^{\vec{m}} = \mathrm{SW}_d^{m_1} \times \cdots \mathrm{SW}_d^{m_i} \ , \qquad \mathfrak{D}_\epsilon^{\vec{m}} = \mathrm{SW}_{\rho_0}^{m_1} \times \cdots \mathrm{SW}_{\rho_0}^{m_i}, \tag{7.46}$$

we have

$$D_{\mathrm{Tr}}(\rho_A, \rho_B) = E_{\vec{m}\sim \mathsf{M}_{\vec{p},N,M}} d_{TV}(\mathfrak{D}_0^{\vec{m}}, \mathfrak{D}_\epsilon^{\vec{m}}) \tag{7.47}$$

First of all we invoke the bound from [OW15]:

$$\mathrm{d}_{\chi^2}(\mathrm{SW}_\rho^n || \mathrm{SW}_{I/d}^n) \leq \exp\bigl(256 n^2 \epsilon^4 / d^2\bigr) - 1 \tag{7.48}$$

Our first observation is that, when $m_i = 1$, (7.48) can be improved noticing that $d_{KL}(\mathrm{SW}_{\rho_i}^1, \mathrm{SW}_d^1) = 0$ for every possible state ρ_i (since there is only one possible partition of $n = 1$ - in other words, we gain no information on whether the state is mixed by measuring a single copy). This observation, together with (7.48) and (7.14), imply that

$$\mathrm{d}_{KL}(\mathrm{SW}_\rho^{m_i} || \mathrm{SW}_{I/d}^{m_i}) \leq 256 \frac{1_{m_i>1} \cdot m_i^2 \epsilon^4}{d^2} \ . \tag{7.49}$$

Using (7.12) and (7.49) we can write

$$\begin{aligned}
D_{\mathrm{Tr}}(\rho_A, \rho_B) &= E_{\vec{m}\sim \mathsf{M}_{\vec{p},N,M}} d_{TV}(\mathfrak{D}_0^{\vec{m}}, \mathfrak{D}_\epsilon^{\vec{m}}) \leq E_{\vec{m}\sim \mathsf{M}_{\vec{p},N,M}} \sqrt{\frac{1}{2} d_{KL}(\mathfrak{D}_0^{\vec{m}}, \mathfrak{D}_\epsilon^{\vec{m}})} \\
&= E_{\vec{m}\sim \mathsf{M}_{\vec{p},N,M}} \sqrt{\frac{1}{2} \sum_{i=1}^N \mathrm{d}_{KL}(\mathrm{SW}_\rho^{m_i} || \mathrm{SW}_{I/d}^{m_i})} \\
&= E_{\vec{m}\sim \mathsf{M}_{\vec{p},N,M}} \sqrt{\frac{1}{2} \sum_{i=1}^N 256 \frac{1_{m_i>1} \cdot m_i^2 \epsilon^4}{d^2}} \\
&\leq \sqrt{E_{\vec{m}\sim \mathsf{M}_{\vec{p},N,M}} \frac{1}{2} \sum_{i=1}^N 256 \frac{1_{m_i>1} \cdot m_i^2 \epsilon^4}{d^2}} \\
&\leq \sqrt{E_{\vec{m}\sim \mathsf{M}_{\vec{p},N,M}} \sum_{i=1}^N 256 m_i(m_i-1) \frac{\epsilon^4}{d^2}} \\
&\leq 16 \frac{\epsilon^2 M}{d\sqrt{N}},
\end{aligned} \tag{7.50}$$

where the first inequality is from Pinsker's inequality, the second equality is the additivity of the Kullback-Leibler divergence, the second inequality is from concavity of the square root. □

It is now immediate to prove Theorem 7.1.2

Proof of Theorem 7.1.2. If an algorithm as in Theorem 7.1.2 exists, one can use it to try to discriminate between ρ_A and ρ_B. By also invoking the Holevo-Helstrom bound Eq. (2.6), the probability of success has to satisfy

$$\frac{1}{2}\left(1+16\frac{\epsilon^2 M}{d\sqrt{N}}\right) \geq p_{succ} \geq \frac{1}{2}\left(\frac{11}{15}+1\right)\frac{2}{3}\,. \tag{7.51}$$

Therefore

$$M \geq 4\cdot 10^{-3}\frac{\sqrt{N}d}{\epsilon^2}. \tag{7.52}$$

□

7.5 Implementation of the optimal measurement

The measurement of the test defined in Sec. 7.3 to prove Theorem 7.1.1 can be implemented on a quantum computer with gate complexity $O(M, \log d, \log 1/\delta)$, where δ is the precision of the implementation, because it can be realized with a sequence of weak Schur sampling measurements. This was already shown for the observable of [BOW19] for $N = 2$ and it can be easily be shown to be true in the general case too. Indeed, in [BOW19] it is shown that $\mathcal{O}_{ii}^{m_i,m_i}$ can be written as

$$\mathcal{O}_{ii}^{m_i,m_j} = \sum_{\lambda\in Y_{m_i,d}} \mathsf{TN}(\lambda)\Pi_\lambda^{(i)}, \tag{7.53}$$

where $Y_{m_i,d}$ are Young diagrams, Π_λ a complete set of orthogonal projectors and $\mathsf{TN}(\lambda) = \frac{1}{n(n-1)}\sum_{i=1}^{d}((\lambda_i - i + 1/2)^2 - (-i+1/2)^2)$. We now define $\mathcal{O}$ to be the average of all transposition on $\mathcal{H}_d^{\otimes M}$, for which we have:

$$\mathcal{O} = \sum_{\lambda\in Y_{M,d}} \mathsf{TN}(\lambda)\Pi_\lambda. \tag{7.54}$$

Using that

$$\frac{M(M-1)}{2}\mathcal{O} = \frac{1}{2}\sum_{i\neq j} m_i m_j \mathcal{O}_{ij}^{m_i,m_j} + \sum_{i=1}^{N}\frac{m_i(m_i-1)}{2}\mathcal{O}_{ii}^{m_i,m_j}, \tag{7.55}$$

we have

$$\mathcal{D}^{\vec{m},M} := \sum_{i\neq j} \mathcal{D}_{ij}^{m_i,m_j,M} = \sum_{i=1}^{N} \frac{2m_i(m_i-1)}{\mu^2 p_i}\mathcal{O}_{ii}^{m_i,m_j} - \frac{2M(M-1)}{\mu^2}\mathcal{O}$$
$$\sum_{i=1}^{N} \frac{2m_i(m_i-1)}{\mu^2 p_i}\mathcal{O}_{ii}^{m_i,m_j} - \frac{2M(M-1)}{\mu^2}\mathcal{O}. \tag{7.56}$$

Since $[\Pi_\lambda, \otimes_{i=1}^{N}\Pi_{\lambda_i}^{(i)}] = 0$, the measurement can be implemented efficiently by nested weak Schur sampling.

7.6 Remarks

We have established the sample complexity of testing identity of collections of quantum states in the sampling model, with a test that can be also implemented efficiently in terms of gate complexity. Note that for this problem one could have used the independence tester of [Yu19], based on the identity test of [BOW19], since if the states in the collection are equal the input of our problem in Eq. (7.1) is a product state, and far from it otherwise. However, the guaranteed sample complexity in this case would have been $O(Nd/\epsilon^2)$, and to get $\sqrt{N}d/\epsilon^2$ we need to make use of the fact that the state in Eq. (7.1) is a classical-quantum state and that we know the classical marginal. This is a state of zero discord [HV01; OZ01; ABC16], and one could ask how the sample complexity differ if the discord is not zero, for example if the states $|i\rangle$ are not orthogonal. This could be seen as an example of quantum inference problem with quantum flags, proved useful in other contexts, e.g. the evaluation of quantum capacities [SSW08; LDS18; Fan+20b; KFG20; WW19b; FKG21]. More generally, an interesting problem would be to study the sample complexity of independence testing with constraints on the structure of the state, with a rich variety of scenarios possible.

Chapter 8

Designing degradable extensions

This chapter is largely based on:

- Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. "Quantum Flags and New Bounds on the Quantum Capacity of the Depolarizing Channel". In: *Physical Review Letters* 125.2 (2020), p. 020503. DOI: 10.1103/PhysRevLett.125.020503. arXiv: 1911.01977.
- Farzad Kianvash, Marco Fanizza, and Vittorio Giovannetti. *Bounding the quantum capacity with flagged extensions*. 2020. arXiv: 2008.02461.
- Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. *Estimating Quantum and Private capacities of Gaussian channels via degradable extensions*. 2021. arXiv: 2103.09569.

8.1 Introduction

In this chapter we present a series of upper bounds on the quantum capacity of several physically motivated quantum channels. We obtain these bounds finding degradable extensions of the channels, according to the method explained in Sec. 3.3. We also refer to Sec. 3.3 for an overview of the various approaches to upper bounds, with or without degradable extensions, and for the definition of flagged extension. The core idea behind these results is to develop a method to easily design degradable extensions, which is flexible enough to apply to different models and improves bounds beyond the low noise regime where approximate degradability [Sut+17; LLS18b] gives satisfactory answers. The first attempt to go beyond flagged extensions with non-orthogonal flags is

in our paper [FKG20], where we found out improved bounds on the quantum capacity of the depolarizing channel using sufficient conditions for degradability of a flagged extension with non-orthogonal pure flags and mixed commuting non-orthogonal flags. Later [Wan21] used approximate degradability to improve this bound, finding numerical evidence that the degradability region of the flagged extension with pure states was larger. The method was also extended to the BB84 channel and to the generalized amplitude damping channel. In [KFG20] we find new sufficient conditions for degradability that are able to analytically recover the results in [Wan21], and are flexible enough to give even better bounds. In [FKG21] we extended this approach to phase-insensitive Gaussian channels, presented in Sec. 4.4.4. In this chapter we will present the content of [KFG20] (which supersedes [FKG20]), and [FKG21], with the following structure: in Sec. 8.2 we prove sufficient conditions for degradability of flagged extensions, and comment on their general applicability. In Sec. 8.3 we apply this method to Pauli channels, evaluating their capacities. In particular, we present upper bounds on the quantum and private capacities of the depolarizing and BB84 channels. In Sec. 8.4 we discuss degradable extensions of the generalized amplitude damping channel, presenting new bounds. In Sec. 8.5 we present a degradable flagged extension of the additive noise channel, and degradable extensions for the thermal attenuator. These results improve the bounds on the quantum and private capacity of these channels and of the thermal amplifier. In Sec. 8.6 we comment on possible improvements.

8.2 Sufficient conditions for degradability of flagged extensions

The following proposition shows a method to design degradable flagged extensions (Def. 3.3.5) for convex combination of channels.

Proposition 8.2.1 (Sufficient conditions for degradability of flagged extensions). *Let $\mathcal{N} = \sum_{i=1}^{l} p_i \mathcal{N}_i$ be a convex combination of channels acting on the quantum system A, and its flagged extension*

$$\widehat{\mathcal{N}} = \sum_{i=1}^{l} p_i \mathcal{N}_i \otimes |\phi_i\rangle \langle \phi_i| \ , \tag{8.1}$$

with $|\phi_i\rangle$ are pure states of an auxiliary flag system F. The channel $\widehat{\mathcal{N}}$ is degradable if there exists an orthonormal basis $\{|i\rangle\}_i$ for the space of F and a choice of Kraus operators $\{K_j^{(i)}\}_{j=1,\dots,r_i}$ for each channel $\mathcal{N}_i$, such that

$$\langle i' | \phi_i \rangle \sqrt{p_i} K_{j'}^{(i')} K_j^{(i)} = \langle i | \phi_{i'} \rangle \sqrt{p_{i'}} K_j^{(i)} K_{j'}^{(i')} \qquad \forall i, j, i', j' \, . \tag{8.2}$$

Proof. The proof follows by explicitly constructing a degrading map. Each channel $\mathcal{N}_i$ admits the following Stinespring dilation

$$V_i |\psi\rangle_A := \sum_{j=1}^{r_i} K_j^{(i)} |\psi\rangle_A |i\rangle_B |j\rangle_{\bar{B}} , \tag{8.3}$$

for all $|\psi\rangle_A$ states of A, with the systems B and $\bar{B}$ being traced out to obtain $\mathcal{N}_i$.

A Stinespring dilation of the flagged channel in Eq. (8.1) can be constructed from the Stinespring dilations V_i

$$V |\psi\rangle_A := \sum_{i=1}^{l} \sqrt{p_i} V_i |\psi\rangle_A |\phi_i\rangle_F . \tag{8.4}$$

On the other hand, the complementary of the flagged channel is defined by

$$\widehat{\mathcal{N}}^c[|\psi\rangle_A \langle\psi|] = \sum_{i,j} \sqrt{p_i p_j}_F \langle\phi_j|\phi_i\rangle_F \operatorname{Tr}_A[V_i |\psi\rangle_A \langle\psi| V_j^\dagger] . \tag{8.5}$$

We consider a channel W which takes as input systems A and F, with the following Stinespring dilation

$$V' |\psi\rangle_A |i\rangle_F := V_i |\psi\rangle_A . \tag{8.6}$$

The following state is the purification of the state after the action of $W \circ \widehat{\mathcal{N}}$

$$\begin{aligned} V'V |\psi\rangle_A &= \sum_{i=1}^{l} \sqrt{p_i} V' V_i |\psi\rangle_A |\phi_i\rangle_F = \sum_{i=1}^{l} \sum_{i'=1}^{l} \sqrt{p_i} \langle i'|\phi_i\rangle V_{i'} V_i |\psi\rangle_A \\ &= \sum_{i=1}^{l} \sum_{i'=1}^{l} \sum_{j=1}^{r_i} \sum_{j'=1}^{r_{i'}} \langle i'|\phi_i\rangle \sqrt{p_i} K_{j'}^{(i')} K_j^{(i)} |\psi\rangle_A |i\rangle_B |j\rangle_{\bar{B}} |i'\rangle_{B'} |j'\rangle_{\bar{B}'} , \end{aligned} \tag{8.7}$$

where for ease of notation $\langle i'|\phi_i\rangle$ stands for ${}_F\langle i'|\phi_i\rangle_F$. On the other hand the states of subsystem $B\bar{B}$ is equal to $\widehat{\mathcal{N}}^c[|\psi\rangle \langle\psi|]$. Therefore, W is a valid degrading map if $V'V |\psi\rangle_A$ is invariant if we swap subsystem $B\bar{B}$ with $B'\bar{B}'$. We now verify that Eq. (8.2) guarantees

this property. Defining the swap operator as $S_{\leftrightarrow}$, we have

$$\begin{aligned} S_{\leftrightarrow}V'V\left|\psi\right\rangle_A &= \sum_{i=1}^{l}\sum_{i'=1}^{l}\sum_{j=1}^{r_i}\sum_{j'=1}^{r_{i'}} \left\langle i'\middle|\phi_i\right\rangle \sqrt{p_i}K_{j'}^{(i')}K_j^{(i)}\left|\psi\right\rangle_A\left|i\right\rangle_B\left|j\right\rangle_{\bar{B}}\left|i'\right\rangle_{B'}\left|j'\right\rangle_{\bar{B}'} \\ &= \sum_{i=1}^{l}\sum_{i'=1}^{l}\sum_{j=1}^{r_i}\sum_{j'=1}^{r_{i'}} \left\langle i'\middle|\phi_i\right\rangle \sqrt{p_i}K_{j'}^{(i')}K_j^{(i)}\left|\psi\right\rangle_A\left|i'\right\rangle_B\left|j'\right\rangle_{\bar{B}}\left|i\right\rangle_{B'}\left|j\right\rangle_{\bar{B}'} \\ &= \sum_{i=1}^{l}\sum_{i'=1}^{l}\sum_{j=1}^{r_i}\sum_{j'=1}^{r_{i'}} \left\langle i\middle|\phi_{i'}\right\rangle \sqrt{p_{i'}}K_{j}^{(i)}K_{j'}^{(i')}\left|\psi\right\rangle_A\left|i\right\rangle_B\left|j\right\rangle_{\bar{B}}\left|i'\right\rangle_{B'}\left|j'\right\rangle_{\bar{B}'}\,, \\ &= V'V\left|\psi\right\rangle_A\,, \end{aligned} \tag{8.8}$$

where we used Eq. (8.2) in the second equality. □

Let us pause to comment on this result. First of all, we note that a special case where these sufficient conditions are met is the case in which the Kraus operators commute, in which case we can choose the flags to be all equal, $|\phi_i\rangle = |\phi\rangle = \sum_j \sqrt{p_j}\,|j\rangle$. The known fact that channels with commuting Kraus operators are degradable [DS05] is then recovered. However, our sufficient condition does not cover the case of flagged extensions of convex combinations of degradable channels with orthogonal flags, whose degradability was a key result in [SS08]. In the proof, we make a peculiar choice of the degrading map which is tailored to the sufficient conditions we identify, and the symmetry under the swap operator is a fairly restrictive sufficient condition. The result of [SS08] is only recovered by allowing more general degrading maps and checking directly the equality of the partial traces of the purified state.

We can apply Proposition 8.2.1 to a variety of scenarios. For example, we consider a channel with one Kraus operator proportional to a unitary:

$$\mathcal{N}[\rho] = (1-p)U\rho U^\dagger + p\sum_{j=1}^{r} K_j\rho K_j^\dagger\,, \tag{8.9}$$

where $\sqrt{p}K_i$ are the other Kraus operators. As far as we are interested in computing the capacity, by unitary invariance of the capacities we can assume that the unitary operator is an identity and redefine accordingly the other Kraus operators. Now

$$\mathcal{N}[\rho] = (1-p)\rho + p\sum_{j=1}^{r} K_j\rho K_j^\dagger = (1-p)\rho + p\Lambda_1[\rho], \tag{8.10}$$

can be seen as a convex combination of the identity channel and another channel $\mathcal{N}_1$ with Kraus operators $\{K_i\}$. A flagged extension is

$$\widehat{\mathcal{N}}[\rho] = (1-p)\rho\otimes|\phi_0\rangle\langle\phi_0| + p\mathcal{N}_1[\rho]\otimes|\phi_1\rangle\langle\phi_1|\,. \tag{8.11}$$

The degradability conditions in Eq. (8.2) read

$$\langle 1|\phi_0\rangle \sqrt{1-p} = \langle 0|\phi_1\rangle \sqrt{p}\,, \qquad \langle 1|\phi_1\rangle K_j K_{j'} = \langle 1|\phi_1\rangle K_{j'} K_j\,. \tag{8.12}$$

Even if K_j operators do not commute, if $p \leq 1/2$ we find that the conditions are met setting $\langle 1|\phi_1\rangle = 0$ and

$$|\phi_1\rangle = |0\rangle\,, \qquad |\phi_0\rangle = \sqrt{\frac{1-2p}{1-p}}\,|0\rangle + \sqrt{\frac{p}{1-p}}\,|1\rangle\,. \tag{8.13}$$

Therefore, we get the upper bound

$$Q(\mathcal{N}) \leq Q(\widehat{\mathcal{N}}) = Q_1(\widehat{\mathcal{N}})\,. \tag{8.14}$$

In [FKG20] we observed that the following flagged extension of the depolarizing channel

$$\widehat{\Phi}_p[\rho] = (1-p)\rho \otimes \big((1-c^2)\,|\phi_0\rangle\langle\phi_0| + c^2\,|\phi_1\rangle\langle\phi_1|\big) + p\frac{I}{d}\,\mathrm{tr}[\rho] \otimes |\phi_1\rangle\langle\phi_1|\,. \tag{8.15}$$

is degradable for $c^2 = \frac{1-2p}{2(1-p)}$, $\langle\phi_0|\phi_1\rangle = 0$. In fact, this corresponds to a special case of flagged extension of a convex combination of a unitary operator and a channel. More generally a flagged extension

$$\widehat{\mathcal{N}}_{c^2}[\rho] := (1-c^2)(1-p)\rho \otimes |\phi_0\rangle\langle\phi_0| + \big(c^2(1-p)\rho + p\mathcal{N}_1[\rho]\big) \otimes |\phi_1\rangle\langle\phi_1|\,, \tag{8.16}$$

according to Eq. (8.13) is degradable for $|\,\langle\phi_0|\phi_1\rangle\,|^2 = \frac{1-2(p+c^2-pc^2)}{1-(p+c^2-pc^2)}$, for $0 \leq c^2 \leq \frac{1-2p}{2(1-p)}$. We can rewrite the extension of [FKG20] in this form. The optimal degradable extension with pure flags, determined numerically by [Wan21], is found putting $c^2 = 0$. Each of these extensions gives an upper bound, and the best bound is found by minimization.

$$Q(\mathcal{N}) \leq \min_{0\leq c^2\leq \frac{1-2p}{2(1-p)}} Q(\mathcal{N}_{c^2})\,. \tag{8.17}$$

Beyond the case of a convex combination of two channels, the power of Proposition 8.2.1 is to treat the case with multiple non-orthogonal flags. The reason why this is crucial is that the case of two flags is easily addressed with approximate degradability applied to flagged channels [Wan21], which is completely satisfactory in terms of numerical bounds: there are only two parameters to optimize, the overlap between the flags and c^2. However, we will show that better bounds can be found if one allows for more refined decomposition which uses more than two flags. In particular, for a channel which is a convex combination of unitary channels (again we can choose one of them to be the identity)

$$\mathcal{N}[\rho] = (1-p)\rho \otimes |\phi_0\rangle\langle\phi_0| + \sum_{j=1}^{r} p_j U_j \rho U_j^\dagger \otimes |\phi_j\rangle\langle\phi_j| \ , \tag{8.18}$$

and any flag choice such that

$$\langle i|\phi_j\rangle = 0 \ \text{ if } i \neq 0 \,\text{and}\, j \neq 0 \,\text{and}\, i \neq j \ , \tag{8.19}$$

gives a non-trivial degradable extension. A more general structure is allowed if one takes into account the commutation properties of the set of unitaries. Indeed, we explore this possibility for Pauli channels. However, the best upper bounds for the depolarizing channel and BB84 channel using this method have exactly this flag structure.
Finally, we remark that even general extensions with mixed flags can be considered in this framework, by adapting the convex combination considered. For example, a rank two flag can be introduced by splitting a term $K\rho K^\dagger \otimes (q\,|0\rangle\langle 0| + (1-q)\,|1\rangle\langle 1|) = \sqrt{q}K\rho\sqrt{q}K^\dagger \otimes |0\rangle\langle 0| + \sqrt{1-q}K\rho\sqrt{1-q}K^\dagger \otimes |1\rangle\langle 1|$, where we now flag a channel with new Kraus operators $\sqrt{q}K$ and $\sqrt{1-q}K$, each with a pure flag associated.

8.3 Degradable extensions of Pauli channels

Pauli channels, as defined in Def. 4.72, are convex combinations of unitary operators corresponding to elements of W_d^n. As we already pointed out in Sec 3.3, the coherent information of Pauli channels is known to be generically non-additive [SS96; DSS98; SS07; FW08], with the most recent and comprehensive analysis of non-additivity being [BL19]. This makes computing their quantum capacities still prohibitive in most interesting cases. In fact, even for the most symmetric non-unitary channel that can be conceived, the depolarizing channel (Eq. 4.74), the quantum (and private) capacity cannot be computed. The most important progress on the upper bounds has been achieved by exploiting degradable extensions [SS08; Ouy14] and approximate degradability [Sut+17; LLS18b] in the low noise regime, antidegradability [Bru+98; Cer00; Smi08; Ouy14] in the high noise regime, and a method that connects the two regimes [LDS18] which can still be understood as using degradable orthogonal flagged extensions. While it has been shown that in the low noise regime the upper bound given by approximate degradability is tangent to the lower bound given by the coherent information for Pauli channels [LLS18b], the gap between the best lower bounds and the best upper bounds is still large for finite diamond norm distance from the identity channel. This work tries to follow the established road of constructing degradable extensions to find improvements in the upper bounds. The bounds we propose are still much better in the low noise regime than in the high noise one, but the improvement is consistent. Moreover, by setting out a method to design degradable extensions that go beyond our current analysis, we make an argument for continuing to explore this method.

For these channels, our sufficient conditions for degradability allow for a wider set of solutions than in the general case, because of the relations (4.68) occurring for any pair of Pauli unitaries. The flagged version of these channels can be constructed by choosing flags in a Hilbert space $\mathcal{H}_F$ of dimension d^{2n}, with computational basis $\{|x\rangle\}_{x\in\mathbb{Z}_d^{2n}}$. We also consider the space $\mathcal{H}_C \otimes \mathcal{H}_F$, with $\mathcal{H}_C \cong \mathcal{H}_F$, and denote the partial trace with respect to $\mathcal{H}_F$ as $\mathrm{Tr}_F[\cdot]$.
Consider a flagged Pauli channel $\Phi_{\mathbf{\Psi}}$

$$\Phi_\Psi[\rho] = \sum_{x\in\mathbb{Z}_d^{2n}} w_x W_x \rho W_x^\dagger \otimes |\phi_x\rangle\langle\phi_x| \,, \tag{8.20}$$

where the label Ψ determines Φ_Ψ through the definition of the state $|\Psi\rangle \in \mathcal{H}_C \otimes \mathcal{H}_F$:

$$|\Psi\rangle = \sum_{x\in\mathbb{Z}_d^{2n}} \sqrt{w_x}\, |x\rangle_C \otimes |\phi_x\rangle_F = \sum_{x\in\mathbb{Z}_d^{2n}, y\in\mathbb{Z}_d^{2n}} \sqrt{w_x}\, \langle y|\phi_x\rangle\, |x\rangle_C \otimes |y\rangle_F \,. \tag{8.21}$$

We define the projectors Π_j on $\mathcal{H}_C \otimes \mathcal{H}_F$ projecting on $\mathrm{span}\{|x\rangle\,|y\rangle - e^{j\frac{2\pi i}{d}}\,|y\rangle\,|x\rangle : \langle x, y\rangle = j \bmod d\}$. With these definitions, we are equipped to establish the following proposition:

Proposition 8.3.1 (Upper bound on the quantum capacity of Pauli channels). *Given a Pauli channel $\Phi_{\mathbf{w}}$, for any $|\Psi\rangle \in \mathcal{H}_C \otimes \mathcal{H}_F$ satisfying*

$$\mathrm{Tr}[\Pi_j\, |\Psi\rangle\langle\Psi|] = 0 \;\; \forall j \in \{0, ..., d-1\} \qquad \mathrm{Tr}[|x\rangle\langle x| \otimes I\, |\Psi\rangle\langle\Psi|] = w_x \qquad \forall x \in \mathbb{Z}_d^{2n}. \tag{8.22}$$

the quantum and private capacities of $\Phi_{\mathbf{w}}$ satisfy

$$Q(\Phi_{\mathbf{w}}) \leq P(\Phi_{\mathbf{w}}) \leq n \log d - S(\mathbf{w}) + S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|]). \tag{8.23}$$

In particular, the optimal upper bound is obtained by minimizing $S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|])$ with the constraints (8.22).

Proof. Any state $|\Psi\rangle$ on $\mathcal{H}_C \otimes \mathcal{H}_F$ satisfying

$$\mathrm{Tr}[|x\rangle\langle x| \otimes I\, |\Psi\rangle\langle\Psi|] = w_x \qquad \forall x \in \mathbb{Z}_d^{2n}. \tag{8.24}$$

can be written as

$$|\Psi\rangle = \sum_{x\in\mathbb{Z}_d^{2n}} \sqrt{w_x}\, |x\rangle_C \otimes |\phi_x\rangle_F = \sum_{x\in\mathbb{Z}_d^{2n}, y\in\mathbb{Z}_d^{2n}} \sqrt{w_x}\, \langle y|\phi_x\rangle\, |x\rangle_C \otimes |y\rangle_F \,, \tag{8.25}$$

identifying a flagged extension Φ_Ψ of $\Phi_{\mathbf{w}}$. Moreover, the degradability conditions for Φ_Ψ can be rewritten as

$$\mathrm{Tr}[\Pi_j \left|\Psi\right\rangle\!\left\langle\Psi\right|] = 0 \ \ \forall j \in \{0, ..., d-1\}, \tag{8.26}$$

therefore Φ_Ψ is degradable. For flagged degradable Pauli channels, $Q(\Phi_\Psi) = I_c(\Phi_\Psi)$ has a very simple form. By the covariance property of (flagged) Pauli channels, i.e. $\Phi_\Psi[W_x \rho W_x^\dagger] = (W_x \otimes I)\Phi_\Psi[\rho](W_x^\dagger \otimes I)$ for all $W_x \in \mathcal{W}_d^n$. Moreover, we can also write the coherent information as $I_c(\Phi_{\mathbf{\Psi}}, \rho) = S(\Phi_\Psi[\rho]) - S(\Phi_\Psi \otimes \mathcal{I}[|\rho\rangle\rangle\langle\langle\rho|])$, for any purification of ρ, denoted by $|\rho\rangle\rangle$, and $\mathcal{I}$ identity channel, therefore using unitarily invariance of the von Neumann entropy and the covariance property we have $I_c(\Phi_{\mathbf{\Psi}}, \rho) = I_c(\Phi_{\mathbf{\Psi}}, W_x \rho W_x^\dagger)$. By concavity of coherent information for degradable channels [YHD08], we thus get

$$I_c(\Phi_{\mathbf{\Psi}}, \rho) = \frac{1}{d^{2n}} \sum_{x \in \mathbb{Z}_d^{2n}} I_c(\Phi_{\mathbf{\Psi}}, W_x \rho W_x^\dagger) \leq I_c(\Phi_{\mathbf{\Psi}}, \frac{1}{d^{2n}} \sum_{x \in \mathbb{Z}_d^{2n}} W_x \rho W_x^\dagger) = I_c(\Phi_{\mathbf{\Psi}}, \frac{I}{d^n}). \tag{8.27}$$

Therefore, the maximum of coherent information corresponds to the maximally mixed state, which is purified by the maximally entangled state $|\Xi\rangle = \frac{1}{d^{n/2}} \sum_{j=0,...,d-1} |j\rangle \otimes |j\rangle$. It holds that $\langle\Xi| W_x^\dagger W_y \otimes I |\Xi\rangle = \frac{1}{d^n} \mathrm{Tr}\left[W_x^\dagger W_y\right] = \delta_{x,y}$, therefore

$$\begin{aligned} I_c(\Phi_\Psi) &= S\left(\Phi_{\mathbf{\Psi}}\left[\frac{I}{d^n}\right]\right) - S(\Phi_{\mathbf{\Psi}} \otimes \mathcal{I}[|\Xi\rangle\!\langle\Xi|]) = n \log d + S(\sum_{x \in \mathbb{Z}_d^{2n}} w_x \left|\phi_x\right\rangle\!\left\langle\phi_x\right|) - S(\mathbf{w}) \\ &= n \log d - S(\mathbf{w}) + S(\mathrm{Tr}_C[|\Psi\rangle\!\langle\Psi|]). \end{aligned} \tag{8.28}$$

□

Some comments are in order: the minimization problem defined by Proposition 8.3.1 is non-convex, therefore it is hard to treat numerically. Its solution is also not unique in general. In our analysis we restricted our attention to subsets of states $|\Psi\rangle$ satisfying the degradability conditions which can be expressed in terms of a few parameters, so that it is easy to solve the the minimization problem numerically. Moreover, Proposition 8.3.1 does not cover the flagged extension with the structure of Eq. (8.16), used for the depolarizing channel in [FKG20; Wan21]. As explained in the comments after Eq. (8.14), this is easily amended by splitting the Kraus operator proportional to the identity in Eq. (4.74) into two Kraus operators with suitable probabilities, and assigning a different flag to each of them, respecting the sufficient conditions for degradability. This approach gives an easy generalization of Proposition 8.3.1.

Another important point to make is that one can also consider the flagged extension of $\mathcal{N}^{\otimes k}$, which would still give an upper bound on the quantum and private capacity, which can be better than looking just at extensions of $\mathcal{N}$: a degradable flagged extension of

$\mathcal{N}$ gives also a degradable flagged extension of $\mathcal{N}^{\otimes k}$ but the converse is not true. For Pauli channels this is even more relevant since the $\Phi_{\mathbf{w}}^{\otimes k}$ is still a Pauli channel, and the quantum capacity of its flagged extensions has a closed form. We have not found better bounds in this way, but it is possible that that this approach could be fruitful.

Let us now see how these results can be applied to two important Pauli channels.

8.3.1 Depolarizing channel

We remind the definition of the depolarizing channel on one qudit (def. 4.74):

$$\Phi_p^{(d)}[\rho] = (1 - \frac{d^2-1}{d^2}p)\rho + \frac{p}{d^2}\sum_{x\in\mathbb{Z}_d^2\setminus\{0\}} W_x \rho W_x^\dagger \,. \tag{8.29}$$

The symmetries of this channel causes some potential redundancies in the states that achieve the optimal upper bound according to Proposition (8.3.1). Consider the unitary operation U_σ indexed by permutations $\sigma \in S_{d^2-1}$ which act by permuting the orthogonal set $\{|x\rangle\}_{x\in\mathbb{Z}_d^2\setminus\{0\}}$ while leaving $|0\rangle$ invariant. Then, for any state $|\Psi\rangle$ satisfying the constraints, $U_\sigma \otimes U_\sigma |\Psi\rangle$ also satisfies the constraints, and it has the same entanglement entropy $S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|]) = S(\mathrm{Tr}_C[U_\sigma\otimes U_\sigma |\Psi\rangle\langle\Psi| (U_\sigma\otimes U_\sigma)^\dagger])$. We cannot establish if the minimization problem has a unique solution, but if this was the case, then we could restrict the candidate states to those which are invariant under $U_\sigma\otimes U_\sigma$ for every $\sigma \in S_{d^2-1}$. We just take this observation as a suggestion for a guess, and we minimize $S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|])$ on this restricted family of states. This is convenient because $S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|])$ can be determined analytically and we can reduce the problem to a one-parameter minimization.

Proposition 8.3.2. *For $|\Psi\rangle$ satisfying $|\Psi\rangle = U_\sigma\otimes U_\sigma|\Psi\rangle$, we can parametrize $|\Psi\rangle$ with three complex variables $\alpha = \langle 0,0|\Psi\rangle$, $\beta = \langle 0,x|\Psi\rangle$ for $x\neq 0$, $\gamma = \langle x,x|\Psi\rangle$ for $x\neq 0$, and*

$$S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|]) = -(d^2-2)|\gamma|^2\log\big(|\gamma|^2\big) - v_+\log v_+ - v_-\log v_- \tag{8.30}$$

with

$$v_\pm = \frac{1}{2}(|\alpha|^2+|\gamma|^2+2|\beta|^2(d^2-1) \pm \sqrt{(|\alpha|^2-|\gamma|^2)^2 + 4(d^2-1)|\beta|^2|\alpha+\gamma^*|^2}) \tag{8.31}$$

Proof. From the constraints we have that $\beta = \langle 0,x|\Psi\rangle = \langle x,0|\Psi\rangle$, and from the action of a permutation U_{xy} that exchanges $x,y\neq 0$ we have $\langle 0,x|\, U_{xy}\otimes U_{xy}\,|\Psi\rangle = \langle 0,y|\Psi\rangle$. From the constraints we have that $\langle x,y|\Psi\rangle = e^{-\frac{2\pi i}{d}\langle x,y\rangle}\langle y,x|\Psi\rangle$ for $x\neq y\,,\ x,y\neq 0$, then $\langle x,y|\, U_{xy}\otimes U_{xy}\,|\Psi\rangle = \langle y,x|\Psi\rangle = e^{-\frac{2\pi i}{d}\langle x,y\rangle}\langle y,x|\Psi\rangle = 0$. Also, $\langle x,x|\Psi\rangle = \langle x,x|\, U_{xy}\,|\Psi\rangle = \langle y,y|\Psi\rangle = \gamma$ when $x,y\neq 0$. This completes the parametrization. The eigenvalues of

$\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|]$ can be determined from the singular values of the matrix M_{xy} of coefficients of $|\Psi\rangle = \sum_{x\in\mathbb{Z}_d^{2n}, y\in\mathbb{Z}_d^{2n}} M_{xy}\,|x\rangle_C \otimes |y\rangle_F$. We have that the coefficients of $M^\dagger M$ are

$$M^\dagger M_{0,0} = |\alpha|^2 + |\beta|^2(d^2-1) \qquad M^\dagger M_{0,x} = \alpha\beta^* + \beta\gamma^*,\ x \neq 0 \tag{8.32}$$

$$M^\dagger M_{x,y} = |\beta|^2,\ x \neq y,\ x, y \neq 0 \qquad M^\dagger M_{x,x} = |\beta|^2 + |\gamma|^2,\ x \neq 0\,. \tag{8.33}$$

Then $M^\dagger M - |\gamma|^2 I$ has rank 2 and the nonzero eigenvalues can be determined by solving a quadratic equation.

□

Proposition 8.3.3. *For $|\Psi\rangle$ satisfying $|\Psi\rangle = U_\sigma \otimes U_\sigma\,|\Psi\rangle$, the minimization of $S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|])$ is a one-parameter minimization problem.*

Proof. From the expression of $S(\mathrm{Tr}_C[|\Psi\rangle\langle\Psi|])$ in Eq. (8.30) and from Eq. (8.31) it is evident that the result does not depend on the phases of α, β and γ except for the term $|\alpha + \gamma^*|$, which should be maximized. This happens without loss of generality if α and γ^* are real and positive. Then the two constraints $|\alpha|^2 + (d^2-1)|\beta|^2 = (1 - \frac{d^2-1}{d^2}p)$ and $|\beta|^2 + |\gamma|^2 = \frac{p}{d^2}$ eliminate the remaining two parameters. □

The bound Q_{fmin} obtained from this one-parameter minimization can be combined with the no-cloning bound [Bru+98; Cer00; Smi08; Ouy14]

$$Q(\Phi_p^{(d)}) \leq \left(1 - \tfrac{2p(d+1)}{d}\right)\log d\,. \tag{8.34}$$

using the fact that the convex hull of upper bounds from degradable extensions of the depolarizing channel is itself an upper bound [Smi08; Ouy14]. A comparison between the most competitive upper bounds for $d = 2$ is shown in Figure 8.1, where we can see that the bound we obtained outperforms all previous bounds in the whole parameter region. An improvement with respect to previous bounds can be obtained also for generic d, and we show as an example the bound for $d = 4$ in Figure 8.2. In this latter case, the bound from the convex hull is improved considering also the bound from Eq. (8.17).

8.3.2 BB84 channel

In this section we consider the channel that describes the famous quantum key distribution protocol by Bennett and Brassard [BB14]. In its general form the channel is

$$\begin{aligned} B_{p_X,p_Z}[\rho] = {} & (1 - p_X - p_Z + p_X p_Z)\rho + (p_X - p_X p_Z)X\rho X + (p_Z - p_Z p_X)Z\rho Z \\ & + p_X p_Z Y\rho Y, \end{aligned} \tag{8.35}$$

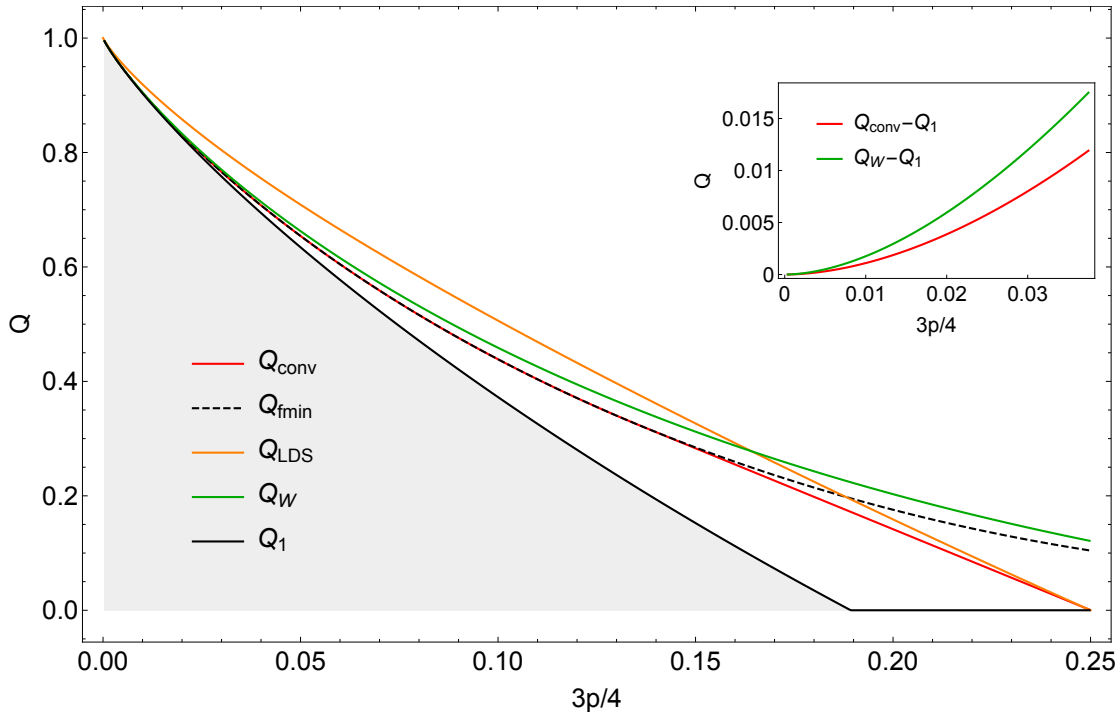


Figure 8.1: Bounds on the quantum capacity of the depolarizing channel for $d = 2$. Here Q_{conv} is the convex hull of the available upper bounds from degradable extensions, Q_{fmin} is the upper bound obtained from Eq. (8.23) by plugging in the expression Eq. (8.30) and minimizing over γ, eliminating the other parameters as explained in the proof of Proposition 8.3.3. Q_1 is the lower bound given by the coherent information of one use of the channel. Q_{LDS} is the bound from [LDS18] and Q_W is the bound from [Wan21].

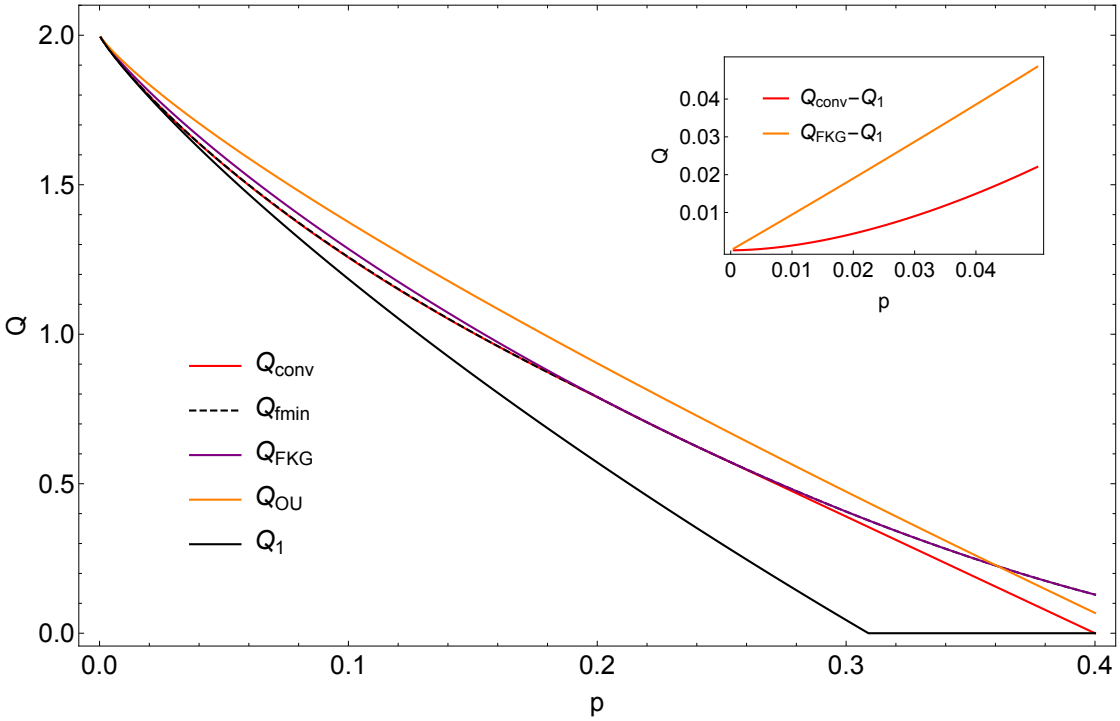


Figure 8.2: Bounds on the quantum capacity of the depolarizing channel for $d = 4$. Here Q_{conv} is the convex hull of the available upper bounds from degradable extensions, Q_{fmin} is the upper bound from non-orthogonal flagged extensions, and Q_1 is the lower bound given by the coherent information of one use of the channel. Q_{FKG} is the bound from [FKG20] and Q_{OU} is the bound from [Ouy14]. Note that in the main plot Q_{fmin} is the bound in Eq. (8.17), since at scale used the bound with more flags is not noticeably better; the situation is different for very small p, in the regime plotted in the inset. In the inset Q_{conv} is effectively the bound obtained from Eq. (8.23) by plugging in the expression Eq. (8.30) and minimizing over γ, eliminating the other parameters as explained in the proof of Proposition 8.3.3.

As in [Sut+17] and [Wan21] we restrict to the case $p_X = p_Z = p$. The flagged extension we consider is

$$B_{p,\Psi}[\rho] = (1-p)^2\rho \otimes |\phi_0\rangle\langle\phi_0| + p(1-p)X\rho X \otimes |\phi_1\rangle\langle\phi_1| + p(1-p)Z\rho Z \otimes |\phi_2\rangle\langle\phi_2| + p^2 Y\rho Y \otimes |\phi_3\rangle\langle\phi_3| \ . \tag{8.36}$$

We choose the following parametrization for the flags

$$\begin{aligned} |\phi_0\rangle &= \sqrt{1-2\alpha^2-\beta^2}\,|0\rangle + \alpha\,|1\rangle + \alpha\,|2\rangle + \beta\,|3\rangle \\ |\phi_1\rangle &= a\,|0\rangle + \sqrt{1-a^2-\gamma^2}\,|1\rangle - \gamma\,|3\rangle \\ |\phi_2\rangle &= a\,|0\rangle + \sqrt{1-a^2-\gamma^2}\,|2\rangle - \gamma\,|3\rangle \\ |\phi_3\rangle &= b\,|0\rangle + c\,|1\rangle + c\,|2\rangle + \sqrt{1-b^2-2c^2}\,|3\rangle \ , \end{aligned} \tag{8.37}$$

where the degradability conditions in Eq. (8.2) imply that $\alpha = a\sqrt{\frac{p(1-p)}{(1-p)^2}}$, $\beta = \frac{bp}{1-p}$ and $\gamma = c\sqrt{\frac{p}{1-p}}$. This is not the most general parametrization for the flags, however, because of the symmetry between the bit flip and phase flip error in Eq. (8.35), we chose this parametrization. Any set of flags in the form of Eq. (8.37) will result in a degradable extension of BB84 channel. Therefore, to get the best upper bound for the quantum capacity or private capacity of BB84 we should minimize the coherent information of its flagged channel with respect to three free parameters a, b, c. We have compared the result of the optimization with the previous bounds in Figure 8.3. The bound in [Wan21] by Wang can be reproduced in our framework just by choosing $a = b = 1$, $c = 0$.

8.4 Degradable extensions of the generalized amplitude damping channel

In this section we consider a bound on the quantum capacity of the generalized amplitude damping channel, which is a model of thermal loss on a qubit, relevant for quantum superconducting processors [CB08]. It can be seen as the analogue of the thermal attenuator in Eq. (4.115) for a discrete variable system.

The generalized amplitude damping channel can be written as

$$\mathcal{A}_{y,N}[\rho] := N\,\mathcal{A}_y[\rho] + (1-N)\,X \circ \mathcal{A}_y \circ X[\rho], \tag{8.38}$$

where $\mathcal{A}_{y,N}$ is the conventional amplitude damping channel, with Kraus operators $K_1 = (|0\rangle\langle 0| + \sqrt{1-y}\,|1\rangle\langle 1|)$ and $K_2 = \sqrt{y}\,|1\rangle\langle 0|$. For $N = 0$, as for the Gaussian thermal attenuator for $N = 0$, $\mathcal{A}_{y,0} = \mathcal{A}_y$ is degradable and its quantum and private capacity

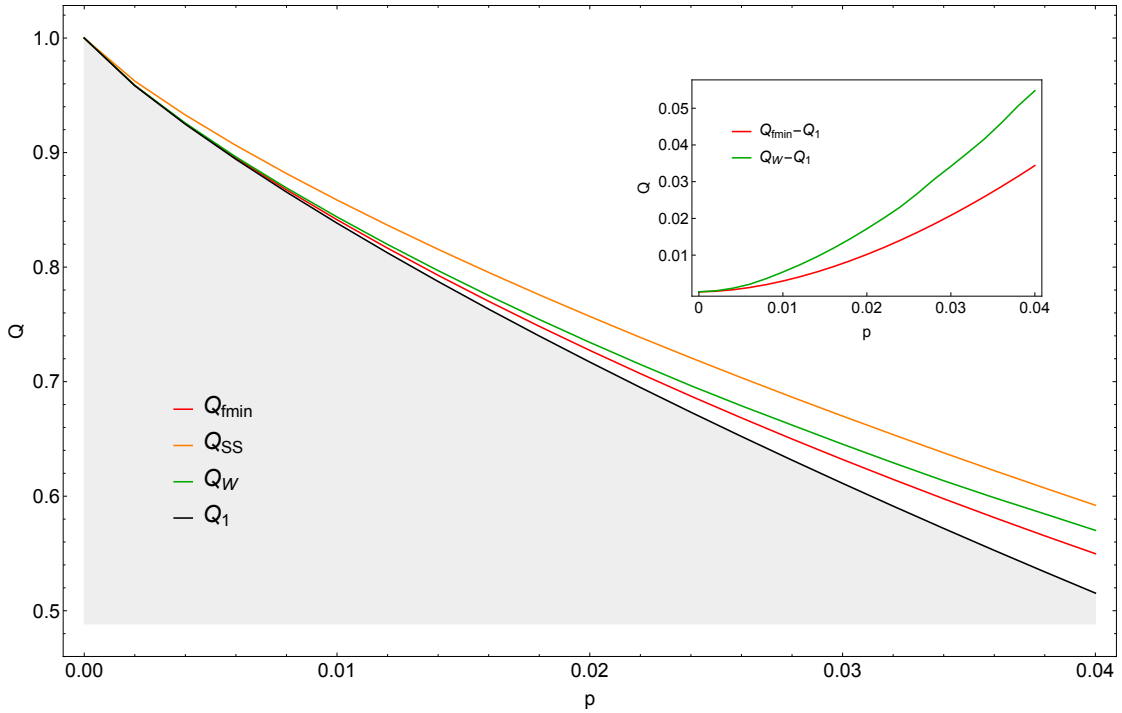


Figure 8.3: Bounds on the quantum and private capacity of BB84 channel. Q_1 is the coherent information of BB84 channel. Q_{fmin} is the new upper bound obtained by the degradable extension, from Eq. (8.23), using the parametrization for the flags in Eq. (8.37), for a suitable choice of the parameters. Q_W is the upper bound obtained in [Wan21]. Q_{SS} is the upper bound derived in [Smi08].

can be computed [GF05]. For $N \neq 0$, while $\mathcal{A}_y$, and $X \circ \mathcal{A}_y \circ X$ are degradable, their convex combination is not, and its quantum and private capacities are not determined. Previous upper bounds have been obtained by [RMG18; KSW20; GP+09; Wan21]. In particular [Wan21] observed that the best bound from approximate degradability of the flagged extension for this decomposition is obtained from the flagged extension with orthogonal flags:

$$\mathcal{A}^F_{y,N}[\rho] = N\,\mathcal{A}_y[\rho] \otimes |0\rangle\langle 0| + (1-N)\,X \circ \mathcal{A}_y \circ X[\rho] \otimes |1\rangle\langle 1|\,, \tag{8.39}$$

In fact this extension is exactly degradable, since it is a flagged convex combination of degradable channels, with orthogonal flags [SS08]. The quantum capacity of this extension can be upper-bounded as $Q(\mathcal{A}^F_{y,N}) \leq (1-N)I_c(A_y,\rho) + NI_c(X \circ A_y \circ X, \rho) = Q(A_y)$. This simple bound seem to not have been pointed out previously, and the actual quantum capacity $Q(\mathcal{A}^F_{y,N})$, which can be evaluated numerically, appears very close to it. An additional family of bounds can be obtained from a different decomposition of the generalized amplitude damping channel. First of all we find a convex hull argument for the generalized amplitude damping channel, following the proof of [SS08] for the depolarizing channel:

Proposition 8.4.1 (Combining bounds of degradable extensions of generalized amplitude damping). *For any collection of degradable extensions $\mathcal{A}^{ext,i}_{y,N}$, $i = 1, ..., l$, for any y_0 the quantum and private capacities of $\mathcal{A}_{y_0,N}$ are upper bounded by the convex hull of $Q(\mathcal{A}^{ext,i}_{y_0,N})$, $i = 1, ..., l$, as functions of the variable N.*

Proof. For any N_1, N_2 such that $N = qN_1 + (1-q)N_2$, $1 - N = 1 - qN_1 + (1-q)N_2 = q(1-N_1) + (1-q)(1-N_2)$, we have

$$\mathcal{A}_{y,N}[\rho] = q(N_1\,\mathcal{A}_y[\rho] + (1-N_1)\,X \circ \mathcal{A}_y \circ X[\rho]) + (1-q)(N_2\,\mathcal{A}_y[\rho] + (1-N_2)\,X \circ \mathcal{A}_y \circ X[\rho]) \tag{8.40}$$

If $\mathcal{A}^{ext,i}_{y,N_1}$ and $\mathcal{A}^{ext,j}_{y,N_2}$ are degradable extensions of $\mathcal{A}_{y,N_1}$ and $\mathcal{A}_{y,N_2}$ respectively, then $q\mathcal{A}^{ext,i}_{y,N_1} \otimes |0\rangle\langle 0| + (1-q)\mathcal{A}^{ext,j}_{y,N_2} \otimes |1\rangle\langle 1|$ is a degradable extension of $\mathcal{A}_{y,N}$ with quantum capacity less than $qQ(\mathcal{A}^{ext,i}_{y,N_1}) + (1-q)Q(\mathcal{A}^{ext,j}_{y,N_2})$. The argument can be iterated to obtain that the convex hull of all degradable extensions is an upper bound.

□

In addition to the extension proposed by [Wan21], we find two other degradable extensions using Proposition 8.2.1. The first is obtained observing that the following set is

also a valid choice of Kraus operators

$$\begin{aligned}
A_1 &= \sqrt{N(1-N)}(\sqrt{1-y}+1)(|0\rangle\langle 0| + |1\rangle\langle 1|) = \sqrt{N(1-N)}(\sqrt{1-y}+1)I, \\
A_2 &= \sqrt{(1-N)y}\,|1\rangle\langle 0|\,, \\
A_3 &= ((1-N) - N\sqrt{1-y})\,|0\rangle\langle 0| + ((1-N)\sqrt{1-y} - N)\,|1\rangle\langle 1|\,, \\
A_4 &= \sqrt{Ny}\,|0\rangle\langle 1|\,.
\end{aligned} \tag{8.41}$$

We notice that A_1 is a rescaled unitary operator, therefore we can directly apply the bound of Eq. (8.14) with $(1-p) = N(1-N)(\sqrt{1-y}+1)^2$. This bound is applicable if $N(1-N)(\sqrt{1-y}+1)^2 > 1/2$. Moreover, at $N = 1/2$, the generalized amplitude damping becomes a Pauli channel:

$$\mathcal{A}_{y,0.5}[\rho] = \frac{1-y/2+\sqrt{1-y}}{2}\rho + \frac{1-y/2-\sqrt{1-y}}{2}Z\rho Z + \frac{y}{4}(Y\rho Y + X\rho X) \tag{8.42}$$

and we get a more refined bound $Q_{\text{fmin}}(y)$, using the techniques of the previous sections, in particular with the same flag structure of BB84 Eq. (8.37). Putting all together, we observe that the bound by [Wan21] remains the best one at high y, but at low y it is beaten by the following bound allowed by the convex hull argument:

$$Q_{\text{conv}}(y,N) = 2NQ_{\text{fmin}}(y) + (1-2N)Q(\mathcal{A}_y) \tag{8.43}$$

and using the full convex hull bound does not give substantial improvements. We plot the results in Figure 8.4.

8.5 Degradable extensions of single mode gauge-covariant Gaussian channels

We begin this section by recalling the state-of-the-art about the quantum and private capacities of single-mode gauge-covariant Gaussian channels, presented in Sec. 4.4.4. We already mentioned these results in Sec. 3.3, and we now make them explicit. The attenuator and amplifier for $N = 0$, that is $\mathcal{E}_{\eta,0}$ in Eq. (4.115) and $\Phi_{g,0}$ in Eq. (4.121) are either degradable or antidegradable [WPG07], and their quantum capacities can be computed as the infinite energy limit of the coherent information of a thermal state:

$$Q(\mathcal{E}_{\eta,0}) = \max\{\log_2\left(\frac{\eta}{1-\eta}\right), 0\} \qquad Q(\Phi_{g,0}) = \log_2\left(\frac{g}{g-1}\right). \tag{8.44}$$

For $N \neq 0$ and for the additive noise channel, we only know upper and lower bounds. A lower bound is again given by the infinite energy limit of the coherent information evaluated on a thermal state [HW01]. We denote these lower bounds by $Q_L(...)$ and we have

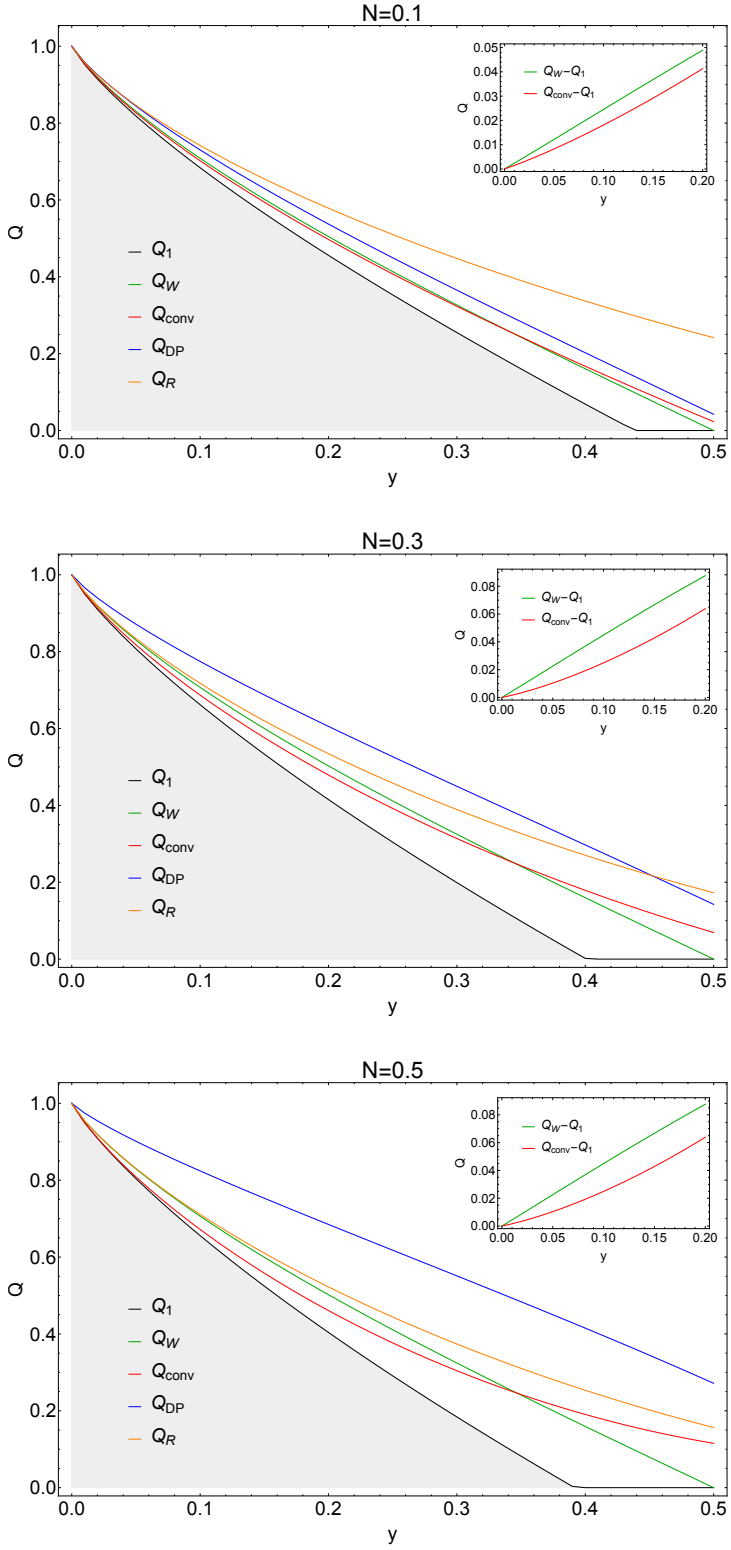


Figure 8.4: Bounds on the quantum capacity of the generalized amplitude damping, for three values of N. Q_1 is the lower bound given by the coherent information of one use of the generalized amplitude damping channel (note that superadditivity has been observed for these channel in [BL20], but the improvement of the lower bound is not quite noticeable at this scale). Q_{conv} is the new upper bound from Eq. (8.43), Q_W is the upper bound obtained by Wang [Wan21], Q_{DP} and Q_R are obtained in [KSW20] respectively from data-processing and Rains information. Previous upper bounds [RMG18; GP+09] are worse and not plotted.

$$Q(\Lambda_\beta) \geq Q_L(\Lambda_\beta) := \max\{\log_2 \beta - 1/\ln 2, 0\}\ , \tag{8.45}$$

$$Q(\mathcal{E}_{\eta,N}) \geq Q_L(\mathcal{E}_{\eta,N}) := \max\{\log_2(\tfrac{\eta}{1-\eta}) - h(2N+1), 0\}. \tag{8.46}$$

$$Q(\Phi_{g,N}) \geq Q_L(\Phi_{g,N}) := \max\{\log_2(\tfrac{g}{g-1}) - h(2N+1), 0\}\ . \tag{8.47}$$

As for the upper bounds, the quantum capacity is zero when these channels become entanglement-breaking, and thus anti-degradable. The entanglement-breaking regions give the following characterizations of the quantum and private capacities [Hol08]:

$$Q(\Lambda_\beta) = P(\Lambda_\beta) = 0 \text{ if } \beta \geq 1, \tag{8.48}$$

$$Q(\mathcal{E}_{\eta,N}) = P(\mathcal{E}_{\eta,N}) = 0 \text{ if } \frac{\eta}{1-\eta} \geq N, \tag{8.49}$$

$$Q(\Phi_{g,N}) = P(\Phi_{g,N}) = 0 \text{ if } \frac{g}{g-1} \geq N. \tag{8.50}$$

$$\tag{8.51}$$

A larger area of zero quantum and private capacity can be obtained using a data-processing argument, and the fact that attenuators with $\eta \leq 1/2$ are anti-degradable. Indeed, when the aforementioned channels are not entanglement breaking, they can be decomposed as

$$\mathcal{E}_{\eta,N} = \mathcal{E}_{\eta',0} \circ \Phi_{\eta/\eta',0}, \qquad \eta' = \eta - (1-\eta)N, \tag{8.52}$$

$$\Phi_{g,N} = \mathcal{E}_{\eta',0} \circ \Phi_{g/\eta',0}, \qquad \eta' = 1 - (g-1)N, \tag{8.53}$$

$$\Lambda_\beta = \mathcal{E}_{\eta',0} \circ \Phi_{1/\eta',0}, \qquad \eta' = 1 - 1/\beta. \tag{8.54}$$

$$\tag{8.55}$$

Therefore, the following upper bounds, valid outside the entanglement-breaking regions, extend the zero capacity regions:

$$Q(\Lambda_\beta) \leq P(\Lambda_\beta) \leq I_c(\mathcal{E}_{1-1/\beta,0}) \tag{8.56}$$

$$Q(\mathcal{E}_{\eta,N}) \leq P(\mathcal{E}_{\eta,N}) \leq I_c(\mathcal{E}_{\eta-(1-\eta)N,0}) \tag{8.57}$$

$$Q(\Phi_{g,N}) \leq P(\Phi_{g,N}) \leq I_c(\mathcal{E}_{1-(g-1)N,0}) \tag{8.58}$$

These bounds can be obtained from the techniques of [Sha+18; RMG18; NAJ19]. Other bounds can be obtained by exchanging the order of amplifier and attenuator in the

decomposition, which where first investigated, and gave worse results [Sha+18] (although this is only true for the unconstrained energy regime, see comment later). In particular, the decomposition with amplifier followed by attenuator was introduced in [RMG18], and Eq. (8.56) was pointed out in [NAJ19].

These upper bounds all use degradability, antidegradability and data-processing techniques, and are state-of-the-art in the high noise regime. In the low noise regime, other upper bounds which are generically valid for the two way quantum and private capacities give better results [Pir+17]. We have the following bounds

$$Q(\Lambda_\beta) \leq Q_{\mathrm{PLOB}}(\beta) = \log_2 \beta - 1/\ln 2 + 1/(\beta \ln 2) \,, \tag{8.59}$$

$$Q(\mathcal{E}_{\eta,N}) \leq Q_{\mathrm{PLOB}}(\eta, N) = -\log_2((1-\eta)\eta^N) - h(2N+1) \,, \tag{8.60}$$

$$Q(\Phi_{g,N}) \leq Q_{\mathrm{AmPLOB}}(g, N) := \log_2(\tfrac{g^{N+1}}{g-1}) - h(2N+1) \,. \tag{8.61}$$

with $h(x) = \frac{x+1}{2}\log_2\left(\frac{x+1}{2}\right) - \frac{x-1}{2}\log_2\left(\frac{x-1}{2}\right)$, as in Eq. (4.104).

Note that, at variance with the classical capacity of these channels [GHGP15], which requires an energy constrained to be finite, the quantum and private capacity are finite whenever the channel do not coincide with the identity [HW01]. Nonetheless, there is interest in estimating the optimal quantum and private rates when there is an energy constraint on the output of the encoding map [WQ16; Sha+18; NAJ19; NPJ20], which is a setting closer to a realistic scenario. The bounds we present in this chapter can also be evaluated in the finite energy setting, but we will perform this analysis in future work.

8.5.1 Flagged extension of the additive gaussian noise

Generalizing the procedure introduced in [KFG20] to infinite dimensional channels we consider the following flagged extension of Λ_β of Eq. 4.123, i.e.

$$\Lambda^e_\beta[\hat{\rho}] := \frac{\beta}{2\pi}\int_{\mathbb{R}^2} \mathrm{d}\mathbf{r}\, e^{-\frac{\beta}{2}\mathbf{r}^\intercal\mathbf{r}} \hat{D}(\mathbf{r})\hat{\rho}\hat{D}(\mathbf{r})^\dagger \otimes |\phi_{\mathbf{r}}\rangle\langle\phi_{\mathbf{r}}| \,, \tag{8.62}$$

where setting $\mathbf{r} := (x, p)$, the states $|\phi_{\mathbf{r}}\rangle$ are product of displaced squeezed states defined by the identity

$$|\phi_{\mathbf{r}}\rangle := \hat{D}(0, -p/2)\,|\beta/2\rangle \otimes \hat{D}(0, x/2)\,|\beta/2\rangle \,, \tag{8.63}$$

with $|\beta/2\rangle$ being a single-mode squeezed vacuum with mean values $\mathbf{m} = 0$ and covariance matrix $V = \begin{pmatrix} 2/\beta & 0 \\ 0 & \beta/2 \end{pmatrix}$.

This channel is Gaussian, as it is an example of Gaussian mixing channels, (see Eq. 4.126), acting on first and second moments as

$$\mathbf{m} \xrightarrow{\Lambda_Y} \mathbf{m}' = \mathbf{m} \,, \qquad V \xrightarrow{\Lambda_Y} V' = V + Y \,. \tag{8.64}$$

By direct comparison with Eq. 4.126 it follows that the flagged additive Gaussian noise of Λ^e_β is a classical mixing channel applied to the state $\hat{\rho} \otimes |\beta/2\rangle\langle\beta/2| \otimes |\beta/2\rangle\langle\beta/2|$ with the matrix Y equal to

$$Y = \begin{pmatrix} \frac{2}{\beta} & 0 & 0 & 0 & 0 & -\frac{1}{\beta} \\ 0 & \frac{2}{\beta} & 0 & \frac{1}{\beta} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 & \frac{1}{2\beta} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{\beta} & 0 & 0 & 0 & 0 & \frac{1}{2\beta} \end{pmatrix} , \tag{8.65}$$

and thus Gaussian.

As explicitly shown in Appendix B.1, Λ^e_β is degradable, with a Gaussian degrading map: here we notice that the intuition for choosing the flags states as in (8.63) comes from Proposition 8.2.1. Indeed, in the special case of Kraus operators proportional to unitary operators, we proved that if $\{\sqrt{p_i}\hat{U}_i\}_{i=1,\dots,n}$ are unitary Kraus operators of a CPTP map $\mathcal{N}$ and $\{|i\rangle\}_{i=1,\dots,n}$ is an orthonormal basis for flags, sufficient conditions for the degradability of $\mathcal{N}^e[\hat{\rho}] := \sum_{i=1}^n p_i \hat{U}_i \hat{\rho} \hat{U}_i^\dagger \otimes |\phi_i\rangle\langle\phi_i|$ are the following

$$\langle i' | \phi_i \rangle \sqrt{p_i} \hat{U}_{i'} \hat{U}_i = \langle i | \phi_{i'} \rangle \sqrt{p_{i'}} \hat{U}_i \hat{U}_{i'} \quad \forall i, i' \,. \tag{8.66}$$

If we have a continuous set of flags, and we pretend that Proposition 8.2.1 still works if we replace the orthonormal basis for the flag space with the (two-mode) pseudo-eigenbasis $\{|\mathrm{x}_1, \mathrm{x}_2\rangle\}_{\mathrm{x}_1,\mathrm{x}_2 \in \mathbb{R}}$ of the position operators of the ancillary modes, the sufficient condition for degradability for Λ^e_β reads

$$\langle \gamma x', \gamma p' | \phi_{\mathbf{r}} \rangle \, e^{-\frac{\beta}{4}\mathbf{r}^\intercal \mathbf{r}} \hat{D}(\mathbf{r}') \hat{D}(\mathbf{r}) = \langle \gamma x, \gamma p | \phi_{\mathbf{r}'} \rangle \, e^{-\frac{\beta}{4}\mathbf{r}'^\intercal \mathbf{r}'} \hat{D}(\mathbf{r}) \hat{D}(\mathbf{r}') \,, \tag{8.67}$$

for all $\mathbf{r}' = (x', p'), \mathbf{r} = (x, p) \in \mathbb{R}^2$ with γ a suitable rescaling factor to determine. With the choice (8.63) of the flags we have explictly

$$\langle \gamma x', \gamma p' | \phi_{\mathbf{r}} \rangle = \sqrt{\frac{\beta}{2\pi}} e^{-\beta \frac{\gamma^2 x'^2 + \gamma^2 p'^2}{4} - \gamma \frac{i p' x - i x' p}{2}} , \tag{8.68}$$

which satisfies the condition in Eq. (8.67) with $\gamma = 1$. Moreover, Λ^e_β is gauge-covariant (in the generalized sense as in Def. 4.4.2). Indeed we have

$$\Lambda^e_\beta[\hat{U}(\theta)\hat{\rho}\hat{U}(\theta)^\dagger] = \hat{U}'(\theta)\Lambda^e_\beta[\hat{\rho}]\hat{U}'(\theta)^\dagger , \tag{8.69}$$

with $\hat{U}'(\theta)$ being a three mode Gaussian unitary acting on $\hat{\mathbf{r}} = (\hat{x}_1, \hat{p}_1, \hat{x}_2, \hat{p}_2, \hat{x}_3, \hat{p}_3)$ as

$$\hat{U}'(\theta)\hat{\mathbf{r}}\hat{U}'(\theta)^\dagger = R'(\theta)\hat{\mathbf{r}}, \tag{8.70}$$

with

$$R'(\theta) := \begin{pmatrix} \cos\theta & \sin\theta & 0 & 0 & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos\theta & 0 & \sin\theta & 0 \\ 0 & 0 & 0 & \cos\theta & 0 & \sin\theta \\ 0 & 0 & -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & -\sin\theta & 0 & \cos\theta \end{pmatrix}. \tag{8.71}$$

According to the remarks after Theorem 4.4.3, we can then evaluate the quantum capacity as the infinite energy limit of the coherent information of thermal states (see Appendix B.2):

$$Q(\Lambda_\beta) \leq P(\Lambda_\beta) \leq Q(\Lambda_\beta^e) = \log_2 \beta - 1/\ln 2 + 2h\left(\sqrt{1+1/\beta^2}\right), \tag{8.72}$$

As shown in Fig. 8.5, Eq. (8.72) is better than (8.59) where $Q(\Lambda_\beta)$ is supposed to be non-zero, i.e. for $1/\beta \leq 0.5$. In the high $1/\beta$ regime, both bounds are surpassed by the bound in [NAJ19], which comes directly from Eq. (8.56):

$$Q(\Lambda_\beta) \leq Q_{\mathrm{NAJ}}(\beta) = \max\{\log_2(\beta - 1), 0\}. \tag{8.73}$$

8.5.2 Extension of the thermal attenuator

In this section we present a degradable extension of $\mathcal{E}_{\eta,N}$.

We first define the passive unitary operator

$$\hat{W}_\eta := \hat{U}_{\eta AE} \otimes \hat{U}_{\eta A'E'}, \tag{8.74}$$

where $\hat{U}_{\eta AE}$ and $\hat{U}_{\eta A'E'}$ are beam splitter transformations acting respectively on the couple of modes A, E and A', E'. We introduce hence the channel

$$\mathcal{F}_{\eta,N}[\rho_{AA'}] := \mathrm{Tr}_{EE'}[\hat{W}_\eta(\rho_{AA'} \otimes |\tau\rangle\langle\tau|_{EE'})\hat{W}_\eta^\dagger], \tag{8.75}$$

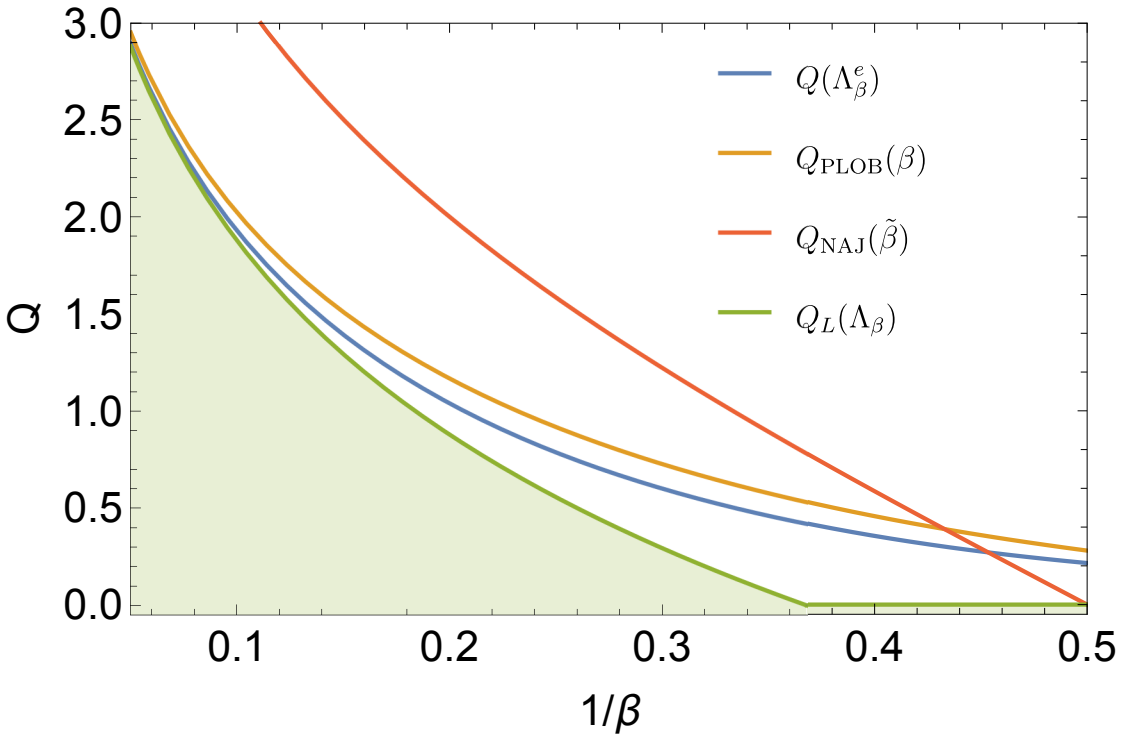


Figure 8.5: Quantum capacity region for the AGNC Λ_β: comparison of the upper bound $Q(\Lambda_\beta^e)$ of Eq. (8.72) with Q_{PLOB} in Eq. (8.59) [Pir+17] and $Q_{\mathrm{NAJ}}(\beta)$ in Eq. (8.73) [NAJ19]. $Q_L(\Lambda_\beta)$ is the lower bound of Eq. (8.45).

and define the extension of $\mathcal{E}_{\eta,N}$ as

$$\mathcal{E}^e_{\eta,N}[\rho_A] := \mathcal{F}_{\eta,N}(\rho_A \otimes |0\rangle\langle 0|_{A'}). \tag{8.76}$$

The map $\mathcal{F}_{\eta,N}$ is manifestly Gaussian, and its action on the first and second moments is

$$\mathbf{m} \xrightarrow{\mathcal{F}_{\eta,N}} \mathbf{m}' = \sqrt{\eta}\mathbf{m}\ , \tag{8.77}$$

$$V \xrightarrow{\mathcal{F}_{\eta,N}} V' = \eta V + (1-\eta)V_{|\tau\rangle}\ , \tag{8.78}$$

With the Stinespring representation in Eq. (8.75) the complementary channel can now computed as $\mathcal{F}^c_{\eta,N} = \mathcal{F}_{1-\eta,N}$. Simple algebra shows that if $\eta > 1/2$ then

$$\mathcal{F}^c_{\eta,N} = \mathcal{F}_{1-\eta,N} = \mathcal{F}_{(1-\eta)/\eta,N} \circ \mathcal{F}_{\eta,N}\ , \tag{8.79}$$

implying that in such regime $\mathcal{F}_{\eta,N}$ (and thus $\mathcal{E}^e_{\eta,N}$) is degradable, with a Gaussian degrading map.

$\mathcal{E}^e_{\eta,N}$ is also gauge-covariant in the generalized sense of Def. 4.4.2:

$$\mathcal{E}^e_{\eta,N}[\hat{U}(\theta)\hat{\rho}\hat{U}(\theta)^\dagger] = \hat{U}''(\theta)\mathcal{E}^e_{\eta,N}[\hat{\rho}]\hat{U}''(\theta)^\dagger, \tag{8.80}$$

with $\hat{U}''(\theta)$ being a two mode Gaussian unitary acting on $\hat{\mathbf{r}} = (\hat{x}_1, \hat{p}_1, \hat{x}_2, \hat{p}_2)$ as

$$\hat{U}''(\theta)\hat{\mathbf{r}}\hat{U}''(\theta)^\dagger = R''(\theta)\hat{\mathbf{r}}, \tag{8.81}$$

with

$$R''(\theta) := \begin{pmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{pmatrix}. \tag{8.82}$$

The quantum capacity of $\mathcal{E}^e_{\eta,N}$ can be thus calculated by evaluating the infinite energy limit of coherent information evaluated on a thermal state, leading to (see App. B.3)

$$\begin{aligned} Q(\mathcal{E}_{\eta,N}) &\leq Q(\mathcal{E}^e_{\eta,N}) \\ &= \log_2(\tfrac{\eta}{1-\eta}) + h((1-\eta)(2N+1)+\eta) - h(\eta(2N+1)+1-\eta)\,. \end{aligned} \tag{8.83}$$

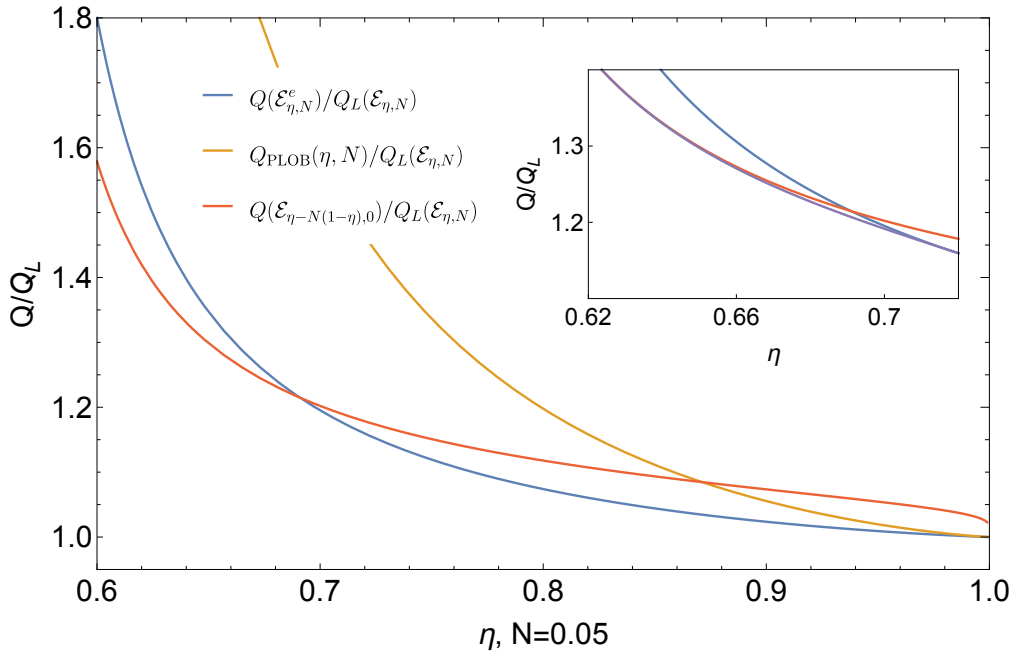


Figure 8.6: Thermal attenuator: ratio between the upper bounds $Q(\mathcal{E}_{\eta-N(1-\eta),0})$ [RMG18], $Q_{\text{PLOB}}(\eta, N)$ [Pir+17] , $Q(\mathcal{E}^e_{\eta,N})$ (this work), and $Q_L(\mathcal{E}_{\eta,N})$ (Eq. 8.46) for $N = 0.05$. In the inset we plot a close-up in the region where $Q(\mathcal{E}_{\eta-N(1-\eta),0})$ and $Q(\mathcal{E}^e_{\eta,N})$ intersect. The purple line is the improved bound using the argument of Eq. (8.87).

A comparison between all these curves is reported in Fig. 8.6, showing that while our inequality (8.83) performs worse than Eq. (8.57) for low η, it gives an improvement with respect to (8.60) for high η.

We finally remark that in our construction the choice of $|0\rangle\langle 0|_{A'}$ in the definition Eq. (8.76) of the extended attenuator is not necessarily optimal. Other Gaussian states could be chosen and the analysis could be done in the same way. In particular, the extension $\mathcal{F}_{\eta,N} \otimes \mathcal{I}[\rho_A \otimes |\tau'\rangle\langle\tau'|_{A'B}]$ gives a slightly better upper bound (optimizing over the single parameter in $|\tau'\rangle$ states), not noticeable on the plot in Fig. 8.6.

8.5.3 Upper bounds for the thermal amplifier

A construction completely analogous to the one in the previous section can be done for the thermal amplifier, but it does not give good upper bounds. It is possible that it is useful in the energy constrained regime, and we will report it in future work. Instead, using the bound on the quantum capacity of the additive noise, and the decomposition

$$\Phi_{g,N} = \Lambda_{\tilde{\beta}} \circ \Phi_{g,0} \tag{8.84}$$

with $\tilde{\beta} = \frac{1}{(g-1)N}$, we get an upper bound which is which is the best known bound to date in the regime $N > 5$ and for intermediate values of g, as seen in Fig. 8.7. As an alternative upper bound we also plot the the upper bound from [NAJ19] on the additive Gaussian noise. We have

$$Q(\Phi_{g,N}) \leq P(\Phi_{g,N}) \leq Q(\Lambda^e_{\tilde{\beta}}) \,, \tag{8.85}$$

$$Q(\Phi_{g,N}) \leq P(\Phi_{g,N}) \leq P(\Lambda_{\tilde{\beta}}) \leq Q_{\mathrm{NAJ}}(\tilde{\beta}) \,, \tag{8.86}$$

which is immediately obtainable also from the bound for the amplifier given by Eq. (8.58).

8.5.4 Combining data-processing and direct bounds

As we observed, the various bounds available are not directly comparable on the whole parameter region. However, taking into account all the possible data-processing decompositions

$$\mathcal{E}_{\eta,N} = \mathcal{E}_{\eta_a,N_{a,1}} \circ \Phi_{g_a,N_{a,2}} = \Phi_{g'_a,N'_{a,1}} \circ \mathcal{E}_{\eta'_a,N'_{a,2}} \,, \tag{8.87}$$

$$\Phi_{g,N} = \mathcal{E}_{\eta_b,N_{b,1}} \circ \Phi_{g_b,N_{b,2}} = \Phi_{g'_b,N'_{b,1}} \circ \mathcal{E}_{\eta'_b,N'_{b,2}} \,, \tag{8.88}$$

using available direct bounds on the channels appearing in the decompositions, and minimizing over the decompositions, one can combine all the direct bounds availabe. An example is seen in the inset of Fig. 8.6, where the purple line corresponds to $\min_{\eta_a,N_{a,1}} Q(\mathcal{E}^e_{\eta_a,N_{a,1}})$ from the decomposition of $\mathcal{E}_{\eta,N}$ in Eq. (8.87).

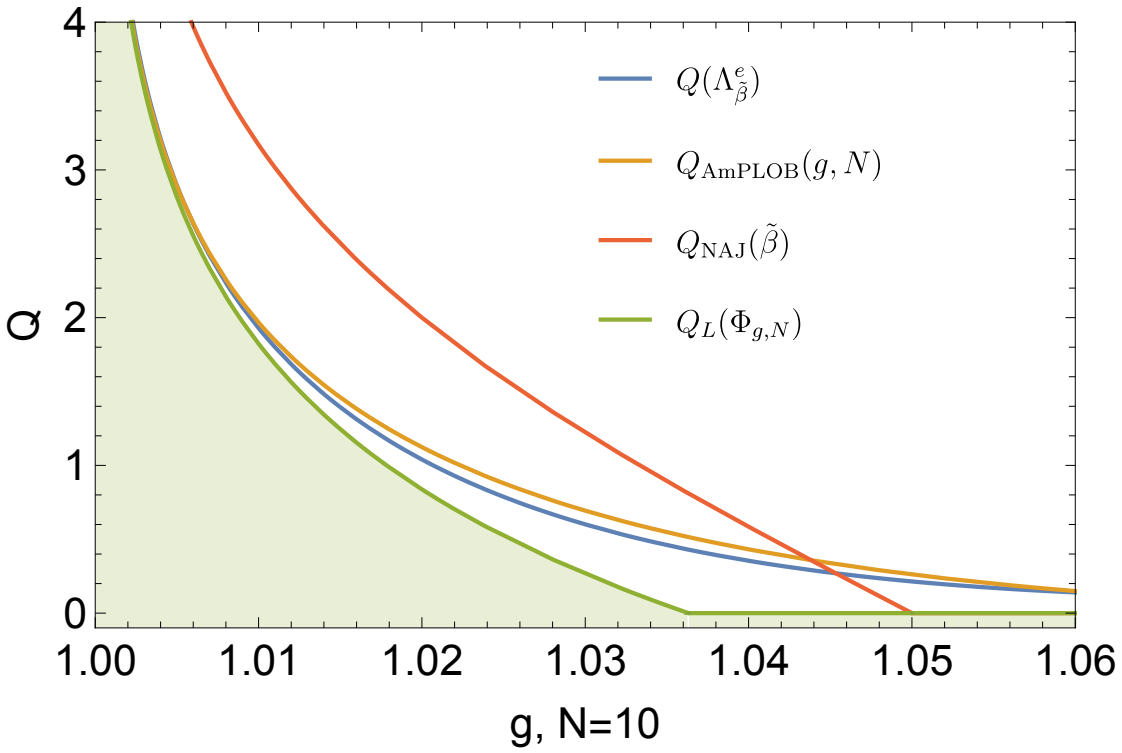


Figure 8.7: Quantum capacity region for the thermal amplifier channel $\Phi_{g,N}$: comparison between the upper bound $Q(\Lambda^e_{\tilde{\beta}})$ of Eq. (8.85) with $Q_{\mathrm{AmPLOB}}(g,N)$ of Eq. (8.61) [Pir+17] and $Q_{\mathrm{NAJ}}(\tilde{\beta})$ in Eq. (8.73) for $N = 10$. $Q_L(\Phi_{g,N})$ is the lower bound in Eq. (8.47).

8.6 Remarks

In this chapter we presented a method for finding bounds on the quantum and private capacities of physically motivated channels. We do not exhaust the possibilities of this method: for the finite dimensional channels, we did not try to numerically optimize in the whole parameter region allowed by the sufficient conditions for degradability. Indeed, the minimization of the upper bound is not a convex optimization problem and would require brute force search, but there are already many parameters for Pauli channels and $d = 2$, $n = 1$. However, we stress the fact that the family of upper bounds for the quantum and private capacity of a channel Λ is even larger in principle, as one can consider the flagged extension of $\Lambda^{\otimes k}$. It is an intriguing possibility that considering more uses could improve the upper bounds towards the high noise regimes, where the gap between the lower bounds and the upper bounds is larger. For infinite-dimensional channels, the construction of degradable extensions is still not systematic, and this is open for future work. Moreover, the results presented here could be easily applied also for the case of energy constrained quantum and private capacity, which will be analyzed elsewhere.

Chapter 9

Classical communication in absence of a shared phase reference

This chapter is largely based on:

- Marco Fanizza, Matteo Rosati, Michalis Skotiniotis, John Calsamiglia, and Vittorio Giovannetti. *Classical capacity of quantum Gaussian codes without a phase reference: when squeezing helps*. 2020. arXiv: 2006.06522.

9.1 Introduction

This chapter is devoted to study how the impossibility of maintaining a shared phase reference frame [BRS07] affects classical communication with continuous variable systems. We have mentioned that the classical capacity of phase-insensitive channels can be exactly computed, since for these channels coherent states are known to minimize the output entropy [Gio+04; MGH14; Gio+14; GHGP15]. However, these models implicitly assume the existence of a shared phase reference, so that the sender and the receiver are phase-locked with a common source. Proposed communication protocols rely on this phase-locking, followed by information encoding into the amplitude and the phase of a coherent state [CGZ11; RMG16; Ban+20], easily generated by a classical source. In this setting, the use of non-classical sources provides no communication advantage. The classical capacity of continuous variable channels requires an energy constraint to be physically meaningful (and finite), therefore it is reasonable to include the energy cost of the phase-locking phase in the energy cost of the communication protocol. This cost is negligible if phase-locking can be done at the start of the protocol and the shared reference can be maintained for the whole duration of the protocol, since in the limit of infinite uses of the channel the energy used for communication goes to infinity. However,

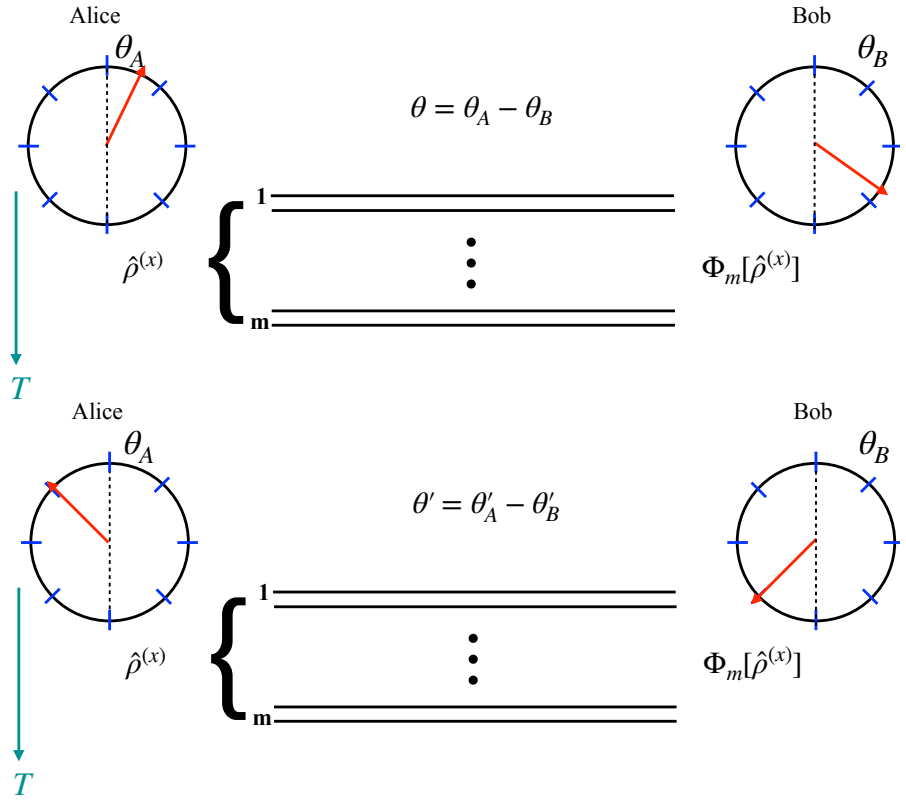


Figure 9.1: Two uses of the phase-noise memory channel Φ_m analyzed in this work, modeling communication in the absence of a phase-reference. The relative phase $\theta \in (0, 2\pi]$ between Alice and Bob is fixed but unknown. Alice exchanges a total of m pulses with Bob, each lasting a time δt, in the form of a quantum state $\hat{\rho}^{(x)} \in \mathcal{B}(L^2(\mathbb{R}^m))$ encoding the classical message $x \in X$. After a time $T = m\delta t$, the relative phase between Alice and Bob's local oscillators changes as a result of each party's local oscillator phase-drifting.

in realistic scenarios the phase reference can be lost or deteriorated. This can happen when the relative phase drifts during transmission due to a physical mechanism in the medium, e.g., Kerr non-linearities and temperature fluctuations in optical fiber [GM90; Wan92; KPB18] or turbulence effects in free space [Sin+14]; but it can also be an effective result of other mechanisms, e.g., the use of a photodetector to measure the signals or the presence of a malicious eavesdropper [DL15; Qi+15]. Indeed, several works analyzed the effect of phase noise on common communication methods based on coherent states encodings [Oli+13; JBDD14; Jar+16; ATP19; DiM+19; COP18]. Since channels modeling phase-noise are not Gaussian channels, computing the capacity is a difficult task. We directly address this problem using a simplified model of phase-noise channel.

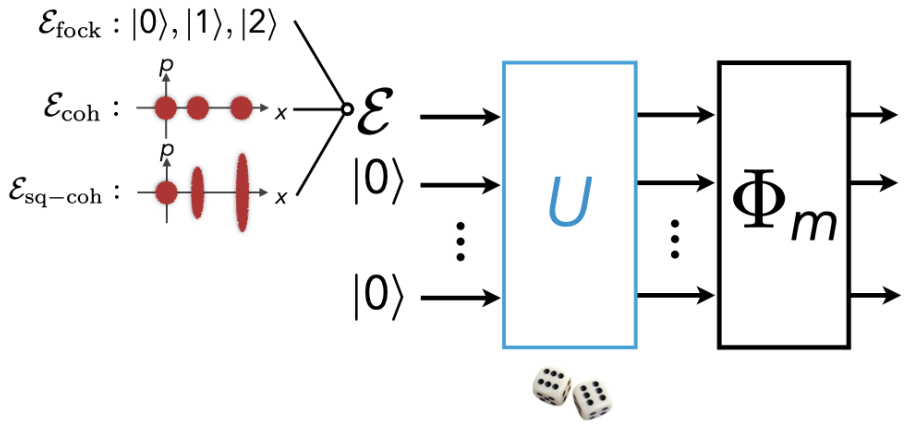


Figure 9.2: Depiction of several communication strategies on phase-noise memory channels studied in the article. Covariant strategies employ a Haar-random passive interferometer $\hat{U}$ to increase the communication rate of any initial ensemble. The input ensembles we consider are discrete constellations comprising the vacuum state and one or more pulse signals, including Fock, coherent and squeezed-coherent states.

We consider a non-Gaussian memory channel [Car+14] that describes the lack of a common frame of reference (see Fig. 9.1). Specifically, we focus on the realistic scenario where decoherence takes place in a finite time T [GM90; Wan92; Sin+14; JBDD14; KPB18] during which the sender can send a total of $m = \lfloor \frac{T}{\delta t} \rfloor$ signals before the onset of decoherence. An alternative is to send a finite number of signals in parallel using different frequency slots. Here δt is the duration of each signal. The channel can be composed with a phase-insensitive attenuator, to have a better representation of both important noise effects on actual communication lines. In absence of loss it is simple to show that the classical capacity for such channels can be achieved using Fock states, while the same is not true in presence of losses. Here we are interested in evaluating the maximum rates achievable via more experimentally friendly encoding schemes. Throughout this chapter we will be interested in transmission rates subject to an average energy constraint and explore strategies that utilize Gaussian state encodings or Fock state with low photon number obeying an average energy constraint (see Fig. 9.2 for a depiction of some analyzed strategies).

In absence of loss, the channel we consider is a special case of the channel studied in [JBDD14], which considers a more general phase noise. There, the authors derived approximations to achievable rates of coherent-state encodings with fixed energy in the low-photon-number sector. Here instead, in the particular case we consider, we analyze

the performance of more general Gaussian encodings, we find exact upper bounds for the optimal coherent state rates, characterize the optimal coherent state rate at low and high energy, and evaluate rates of explicit encodings.

In particular, we show that the addition of squeezing can greatly enhance the communication rate compared to coherent-state encodings, for this family of phase-noise channels. In particular, in the case of complete dephasing ($m = 1$, coinciding with the channel seen by a photodetector), we exhibit an explicit squeezed-coherent-state encoding whose rate surpasses any coherent-state communication strategy for suitable values of the average energy of the signals. Our results provide a clear departure from the case of phase-covariant Gaussian channels and prove the unconditional advantage of using non-classical Gaussian light in a physically motivated communication setting. We note that several works observed an enhancement in discrimination and communication in presence of phase noise via the addition of squeezing to coherent states, though with several restrictions on sources and measurements [YS78; ST87; VR94; Yue04; CCP14; COP18]. To our knowledge, our work is the first to prove an unconditional advantage.

Furthermore, we show that the advantage of squeezed-coherent encodings is robust with respect to the addition of channel losses, contrarily to Fock-state encodings. In this way we identify a regime where the use of a source producing up to two-photon Fock states is sub-optimal with respect to a much simpler squeezed-coherent source with reasonable squeezing values, e.g., ~ 5.8dB at energy $E \sim 2$ vs the 8-15dB attainable at the state of the art [Vah+16; Zha+21].

Finally, we show that the use of part of the signals to establish a common phase reference [SS91; BW00; BRS07] on these channels is in general detrimental for the communication rate, even at large signal energies.

The chapter is structured as follows. In Sec. 9.2 we recap the general communication problem and introduce the channel model. In Sec. 9.3 we first compute the channel capacity in the lossless case, showing it is attained by Fock encodings; we then prove the optimality of covariant encodings, and finally specify the results to Gaussian encodings. In Sec. 9.4 we first determine upper bounds on coherent-state encodings, applying recent advances in bounding the classical Poisson channel capacity [WW14; CR19]; we then we compute attainable lower bounds on coherent and squeezed-coherent encodings via explicit discrete-pulse alphabets. Moreover, in Sec. 9.5 we consider the effect of losses, while in Sec. 9.6 we prove the strict sub-optimality of phase-estimation strategies at large energies. Finally, in Sec. 9.7 we discuss the relevance of our results.

9.2 Phase-noise model

In this section we introduce the mathematical model describing the effective channel associated with the lack of a common phase reference, and establish the notation used throughout the remainder of our work. The sender A wants to communicate classical information to the receiver B. Both A and B are in possession of their own phase reference in the form of local oscillators, i.e., high-intensity laser light with respective phases θ_S and θ_R; in general $\theta_S \neq \theta_R$. The mismatch between the two phase references can be represented as a unitary channel connecting A and B, such that any state $\hat{\rho} \in \mathcal{B}(L^2(\mathbb{R}^m))$, prepared by A is described from B's point of view as $e^{-i\theta\hat{N}}\hat{\rho}e^{i\theta\hat{N}}$, where $\hat{N}$ is the total photon number operator of Eq. 4.93. If the phase mismatch is random, we obtain the channel Φ_m of Eq. 4.108,

$$\Phi_m(\hat{\rho}) = \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi}\, e^{-i\theta\hat{N}}\,\hat{\rho}\, e^{i\theta\hat{N}} = \sum_{n=0}^{\infty} p(n)\hat{\rho}_n \tag{9.1}$$

where now $p(n) := \mathrm{tr}\left[\hat{\Pi}_n\hat{\rho}\right]$, $\hat{\rho}_n := \hat{\Pi}_n\hat{\rho}\hat{\Pi}_n/p(n)$ and $\hat{\Pi}_n$ is the projector on the subspace of $\mathcal{H}$ with total photon number n. The channel Φ_m is a non-Gaussian memory channel and outputs gauge-invariant states. Φ_m models a situation where any phase-variation mechanism will take place after a finite amount of time T, during which the unknown relative phase, $\theta \in (0, 2\pi]$, will remain fixed. Therefore, A and B are capable of exchanging $m = \lfloor \frac{T}{\delta t} \rfloor$ signals, if each of them has duration δt.

In the special case $m = 1$, Φ_1 is equivalent to a photodetector measurement.

$$\Phi_1(\hat{\rho}) = \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi}\, e^{-i\theta\hat{n}_1}\hat{\rho}e^{i\theta\hat{n}_1} = \sum_{n_1=0}^{\infty} |n_1\rangle\langle n_1|\,\hat{\rho}\,|n_1\rangle\langle n_1|\,, \tag{9.2}$$

Observe that whereas A and B cannot use the global-phase degree of freedom of light to encode information—there are no coherences between states with different total photon number—they can still utilise the relative phase between m-mode states with a fixed total photon number since Φ_m commutes with the action of energy-preserving Gaussian unitaries, i.e., m-mode passive interferometers.

An intuitive communication strategy for A and B would be to use part of the energy to perform a standard phase estimation in order to phase-lock their lasers [BW00] and then perform a standard communication protocol. As we will show in Sec. 9.6, this possibility is sub-optimal in practice when A and B's average energy resources are finite, and also asymptotically strictly suboptimal with respect to using a thermal ensemble of coherent states.

A simple model to account for losses is to precede the phase noise channel Φ_m with a lossy bosonic channel $\mathcal{E}_{\eta,N}^{\otimes m}$. Since $\mathcal{E}_{\eta,N}^{\otimes m}$ is phase-insensitive, it commutes with Φ_m and the order in which the channels are placed its irrelevant. We will consider only the case $N = 0$, in Sec 9.5.

9.2.1 Constrained rates

The Holevo quantity of the output of an ensemble gives an achievable rate for the channel (Sec. 3.2.1). In addition to the energy constraint, other constraints on the ensemble can be motivated by practicality. If one is restricted to use a specific type of signals $\hat{\rho}^{(x)} \in \mathcal{S}$, then

$$R_{\mathcal{S}}(\Phi) := \max_{\mathcal{E} \text{ s.t. } \forall x\, \hat{\rho}^{(x)} \in \mathcal{S}} \chi(\Phi, \mathcal{E}) \tag{9.3}$$

determines the maximum rate attainable by sending sequences of signals extracted from $\mathcal{S}$ over multiple uses of the channel and decoding them with an optimal collective quantum measurement. The realization of such measurement is still not trivial in practice, even for coherent-state codes [CGZ11; RG16; RMG16; RMG17; Ban+20], and typical low-energy communication methods commit to specific single-system quantum measurements $\mathcal{M} := \{\hat{M}_y \geq 0, \sum_y \hat{M}_y = I\}$, which, in conjunction with a specific type of signals $\mathcal{S}$, induce a classical channel with maximum information transmission rate given by its Shannon capacity:

$$R_{\mathcal{S}}(\Phi; \mathcal{M}) := \max_{\mathbf{q}} \sum_{x,y} q(x) p^{(x)}(y) \log \frac{p^{(x)}(y)}{\sum_{x'} q(x') p^{(x')}(y)}, \tag{9.4}$$

where the object to maximize is the classical mutual information between the input x, with probability distribution $\mathbf{q}$, and the output y, with conditional probability distribution $\mathbf{p}^{(x)}$ such that $\mathbf{p}^{(x)}(y) := \operatorname{tr}\left[\hat{M}_y \Phi(\hat{\rho}^{(x)})\right]$.

9.3 Phase-noise channel: capacity, covariant and Gaussian rates

In this section we compute the phase-noise memory channel capacity and several maximum information transmission rates with Gaussian encodings. We make use of covariant encodings [Kor+19], which randomize any given encoding, and we show that they attain larger or equal rate for our channel. An important disclaimer: at variance with previous chapters, here the entropies, and therefore the capacities, are computed with natural logarithms instead of base two.

9.3.1 Classical capacity of the phase-noise channel and Fock encodings

In this section we show that the classical capacity of Φ_m is $C(\Phi_m, E) = m\, g(\frac{E}{m})$, where $g(E) := (E+1)\log(E+1) - E\log E$ is the entropy of a single-mode thermal state of average energy E.

It is straightforward to see that for an ensemble $\mathcal{E}_k$ with energy constraint kE

$$\chi(\Phi_m^{\otimes k}, \mathcal{E}_k) \leq S(\hat{\rho}_{\rm th}(kE)) = \sum_{j=1}^{mk} g(E_j) = k\, m\, g\left(\frac{E}{m}\right), \tag{9.5}$$

where the first inequality follows from discarding negative terms in the Holevo quantity and using the fact that the entropy is maximized, under an average-energy constraint, by a thermal state [YO93; CD94]. This follows from $0 \leq D(\hat{\rho}||\hat{\rho}_{\rm th}(E)) = S(\hat{\rho}_{\rm th}(E)) - S(\hat{\rho})$ whenever $\mathrm{tr}\left[\hat{\rho}\hat{N}\right] = E$, with equality if and only if $\hat{\rho} = \hat{\rho}_{\rm th}(E)$. Here $\hat{N}$ is the total photon number of mk modes and $\hat{\rho}_{\rm th}(kE)$ is a thermal state of mk modes with total average energy kE, which can always be written as tensor product of single-mode thermal states with average energies $E_j = E/m$. Finally, $k\, m\, g(\frac{E}{m})$ is monotonic in E, therefore it is not restrictive to constrain the total energy to be exactly kE.

Now note that the upper bound of Eq. (9.5) is achievable using an ensemble of tensor-product Fock states on m modes, i.e., mapping $x \in X \mapsto \bigotimes_{i=1}^m |n_i^{(x)}\rangle$. Indeed, Fock states are pure, giving a zero contribution to the second term of the Holevo quantity, and invariant under the action of Φ_m, so that with a thermal probability distribution of total average energy E, the average output state of the channel is exactly $\hat{\rho}_{\rm th}(E)$. Hence we conclude that $C(\Phi_m, E) = m\, g(\frac{E}{m})$. Moreover, the same arguments above apply to any phase-noise channel with arbitrary phase distribution, provided that the phase-shift is identical on each mode, and their capacity is given by the same expression.

We stress that this is the same rate attainable by m uses of the identity channel with average energy per mode $\frac{E}{m}$, implying that, if the sender and receiver can produce and detect Fock states, then Φ_m is essentially noiseless.

Finally, as Fock states with increasing photon number are increasingly difficult to produce, it makes sense to consider the rate obtainable by sending ensembles of Fock states with fixed maximum photon number. Repeating the proof of this section, the optimal rate of these ensembles as defined in (9.3) is readily characterized as

$$R_t(E) = \max_{s<E} g(s,t), \tag{9.6}$$

where $g(s,t) = \beta(s)s - \log\left(\frac{1-e^{-\beta(s)(t+1)}}{1-e^{-\beta(s)}}\right)$ is the entropy of the thermal state of the

truncated Hilbert space with photon number up to t, with average photon number s, with inverse temperature $\beta(s)$.

Note that for any ensembles of t states the maximum rate is bounded by the maximum value of the Holevo quantity, i.e., $\log t$. For restricted Fock ensembles with photon number up to 1 or 2 states the Holevo quantity saturates to respectively $\log 2$ and $\log 3$ as the energy constraint grows.

9.3.2 Covariant encodings

Since photon-number states are hard to produce, one can be interested in constraining the ensembles to more accessible states. Note that, although the channel is additive, i.e., its capacity is attained with product encodings over different uses, superadditivity may arise due to the constrained input. For simplicity, we will restrict to product encodings in the paper. For coherent states, this restriction is actually free since they are always product states. A drastic simplification in the optimization over any family of states can be obtained by exploiting the symmetry of the channel (see [Kor+19] for a general resource-theoretic treatment of encoding-restricted communication).

We use the fact that the average on Haar-random energy-preserving Gaussian unitaries $\hat{U}$, which act as the group $\mathrm{U}(m)$ in phase space [Ser17] (see Sec. 4.4.3), completely depolarizes the state in blocks of fixed total photon number n, which have dimension $\binom{n+m-1}{m-1}$:

$$\int_{\mathrm{U}(m)} dU\, \hat{U}\hat{\rho}\hat{U}^\dagger = \sum_{n=0}^{\infty} p(n) \frac{\hat{\Pi}_n}{\binom{n+m-1}{m-1}}. \tag{9.7}$$

This is a consequence of Schur's lemma applied to the decomposition into irreducible representations of $\mathrm{U}(m)$ of the Hilbert space of m modes. The decomposition in turn can be understood as a consequence of the connection between coherent states of an infinite-dimensional system with spin-coherent states of finite dimension [Per72; ZFG90], detailed in the Appendix B.4. We also use the term passive interferometers (PI) to refer to these unitary operators, since they can be implemented with simple passive linear optics elements.

Exploiting this property, the rate achievable with an arbitrary ensemble $\mathcal{E} = \{q(x), \hat{\rho}^{(x)}\}$ is then bounded by

$$\begin{aligned} \chi(\Phi_m, \mathcal{E}) \leq \Bigg[& S\left(\int dx\, q(x) \int_{\mathrm{U}(m)} dU \Phi_m(\hat{U}\hat{\rho}^{(x)}\hat{U}^\dagger) \right) \\ & - \int dx\, q(x) \int_{\mathrm{U}(m)} dU S(\Phi_m(\hat{U}\hat{\rho}^{(x)}\hat{U}^\dagger)) \Bigg] = \chi(\Phi_m, \mathcal{E}^{\mathrm{Haar}}) \end{aligned} \tag{9.8}$$

where the inequality follows from the concavity and unitary-invariance of the von Neumann entropy and the fact that $\hat{U}$ and Φ_m commute, while in the last equality we defined $\mathcal{E}^{\text{Haar}}$ as the ensemble obtained by applying a Haar-random $\hat{U}$ to the states extracted from $\mathcal{E}$. The inequality means that one can always restrict the maximization of the Holevo quantity to ensembles of the form $\mathcal{E}^{\text{Haar}}$, which are invariant under total-phase shifts and thus constitute what we refer to as *covariant encodings*.

It follows that for any ensemble of states $\mathcal{E}$ with total photon number distribution $\mathbf{p}^{(x)}$, where $\mathbf{p}^{(x)}(n) = \text{tr}\left[\hat{\Pi}_n \hat{\rho}^{(x)}\right]$, using Eq. (9.7) the Holevo quantity can be computed as

$$\begin{aligned}\chi(\Phi_m, \mathcal{E}^{\text{Haar}}) &= \left[H\left(\int dx\, q(x)\mathbf{p}^{(x)}\right) - \int dx\, q(x) H(\mathbf{p}^{(x)}) \right. \\ &\left. + \sum_{n=0}^{\infty} \int dx\, q(x)\mathbf{p}^{(x)}(n) \left(\log \binom{n+m-1}{m-1} - S(\hat{\rho}_n^{(x)}) \right) \right] \\ &= mg(\frac{E}{m}) - D\left(\int dx\, q(x)\mathbf{p}^{(x)} || \mathbf{p}^{\text{th}}\right) - \int dx\, q(x) \left(H(\mathbf{p}^{(x)}) + \sum_{n=0}^{\infty} \mathbf{p}^{(x)}(n) S(\hat{\rho}_n^{(x)}) \right),\end{aligned}$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence, $H(\cdot)$ the Shannon entropy (in this chapter we use a different notation for $H(...)$ and $S(...)$, the von Neumann entropy, since the first appears prominently) and

$$\mathbf{p}^{\text{th}}(n) = \binom{n+m-1}{m-1} \left(\frac{E}{E+m}\right)^n \left(\frac{m}{E+m}\right)^m \tag{9.9}$$

is the total-photon-number distribution of the thermal state on m modes with average energy per mode $\frac{E}{m}$.

From this expression it is intuitively apparent that states with total-photon-number distribution more concentrated around the mean are preferable as they make $H(\mathbf{p}^{(x)})$ smaller without necessarily increasing $D(\int dx\, q(x)\mathbf{p}^{(x)}||\mathbf{p}^{\text{th}})$. Indeed, as already mentioned above, the capacity of the channel is attained by a thermal ensemble of Fock states. This fact will be crucial in understanding why sub-Poissonian squeezed states offer an enhancement (see Sec. 9.4.2), where sub(super)-Poissonian means that the variance of the photon number distribution is smaller (larger) than the variance of the Poissonian with the same mean.

9.3.3 Gaussian encodings

In the rest of the chapter we will restrict to Gaussian encodings, which are easily realizable in practice. Note that, thanks to the concavity of the entropy and the fact that any Gaussian state can be written as a mixture of pure Gaussian states, it is straightforward to further restrict the optimization to pure Gaussian states (see Appendix B.5).

Any pure Gaussian state $|\psi\rangle$ can be written as

$$|\psi\rangle = \hat{U}\,\hat{S}(\mathbf{r})\,\hat{D}(\mathbf{s})\,|0\rangle^{\otimes m} \tag{9.10}$$

where $\hat{D}(\mathbf{s})$ is a product of single-mode displacements defined in Eq. (4.79), while $\hat{S}(\mathbf{r}) = \hat{S}(r_1)\dots\hat{S}(r_m)$ is a product of single-mode squeezing operators, defined as in Eq. (4.101).

We are then left to consider Gaussian ensembles of the form $\mathcal{E}_{\rm G}^{\rm Haar} := \{q(\mathbf{r},\mathbf{s})\,dU, \hat{U}\,|\mathbf{r},\mathbf{s}\rangle\}$, which can be generated by producing a tensor product of m squeezed-coherent states $|\mathbf{r},\mathbf{s}\rangle := \hat{S}(\mathbf{r})\hat{D}(\mathbf{s})\,|0\rangle^{\otimes m}$ with probability $q(\mathbf{r},\mathbf{s})$ and then acting with a m-mode Haar-random PI $\hat{U}$. Consequently, the optimal Gaussian rate, as defined by Eq. (9.3) for $\mathcal{S}$ being the set of Gaussian states, is obtained by maximizing Eq. (9.9) over $q(\mathbf{r},\mathbf{s})$:

$$\begin{aligned} R_{\rm G}(\Phi_m) = \max_{q(\mathbf{r},\mathbf{s})} \Bigg[& H\left(\int d\mathbf{r}\,d\mathbf{s}\,q(\mathbf{r},\mathbf{s})\mathbf{Q}^{(\mathbf{r},\mathbf{s})}\right) - \int d\mathbf{r}\,d\mathbf{s}\,q(\mathbf{r},\mathbf{s})H(\mathbf{Q}^{(\mathbf{r},\mathbf{s})}) \\ & + \sum_{n=0}^{\infty}\int d\mathbf{r}\,d\mathbf{s}\,q(\mathbf{r},\mathbf{s})\mathbf{Q}^{(\mathbf{r},\mathbf{s})}(n)\log\binom{n+m-1}{m-1}\Bigg], \end{aligned} \tag{9.11}$$

where $\mathbf{Q}^{(\mathbf{r},\mathbf{s})}$ is the total-photon-number distribution of $|\mathbf{r},\mathbf{s}\rangle$ obtainable from that of a single-mode squeezed-coherent state [Yue76; GA90] (see Appendix B.7).

9.4 Bounds on Gaussian communication rates

In this section we determine upper and lower bounds on Gaussian communication rates on Φ_m. We recall that a more general phase-noise model encompassing our Φ_m has been studied in [JBDD14], where the authors derived approximations to the Holevo information of coherent-state ensembles with fixed energy in the low-photon-number sector. By restricting our attention to Φ_m, we perform a wider analysis, extending the attention to general encodings. In particular, in this section we

- evaluate an exact upper bound for the rate of coherent state encodings;
- present explicit strategies with covariant encodings and discrete energy pulses;
- study the high and low energy behaviour of the optimal coherent state rate and show how to achieve them;
- show the advantage of using squeezing with respect to coherent states.

9.4.1 Maximum coherent-state rate and its upper bounds

If we restrict to coherent-state encodings ($\mathbf{r} = 0$ in Eq. 9.10), we can provide a general expression for the optimal rate using the fact that the total-photon-number distribution of

a coherent state $|\mathbf{s}\rangle := \hat{D}(\mathbf{s})\,|0\rangle^{\otimes m}$ is a Poissonian $\mathbf{P}^{(s)}$ with probabilities $\mathbf{P}^{(s)}(n) := e^{-s}\frac{s^n}{n!}$ and depends only on its total energy $s := |\mathbf{s}|^2$.

The optimization in Eq. (9.9) can be restricted to input distributions on the total energy only, $q(s)$, and the optimal coherent-state rate

$$\begin{aligned} R_{\mathrm{c}}(\Phi_m, E) &:= \max_{q(s)} \chi(\Phi_m, \mathcal{E}_{\mathrm{c}}^{\mathrm{Haar}}) \\ &= \max_{q(s)} \left[\sum_{n=0}^{\infty} \int_0^{+\infty} ds\, q(s) \mathbf{P}^{(s)}(n) \log \binom{n+m-1}{m-1} \right. \\ &\left. + H\left(\int_0^{+\infty} ds\, q(s) \mathbf{P}^{(s)} \right) - \int_0^{+\infty} ds\, q(s) H(\mathbf{P}^{(s)}) \right] \end{aligned} \tag{9.12}$$

is attained by producing a single-mode coherent state of energy s with probability $q(s)$ and distributing it with a Haar-random PI. For $m = 1$, only the last two terms of Eq. (9.12) contribute and we recover the well-known discrete-time classical Poisson channel with input distribution $q(s)$. Its capacity is still an open problem in classical information theory for which only upper and lower bounds are known [Mar07; Lap+08; LM09; WW14; CR19]. For $m > 1$, the first term adds a positive contribution depending on the number of modes m and the output photon-number distribution.

Employing the connection with the Poisson channel, we can upper-bound the optimal coherent-state rate by bounding separately the expressions in the second and third rows of Eq. (9.12).

Proposition 9.4.1. *For any upper bound $R_{\mathrm{c}}(\Phi_1, E) < f(E)$ (where $R_{\mathrm{c}}(\Phi_1, E)$ is the capacity of the Poisson channel) we have*

$$R_{\mathrm{c}}(\Phi_m, E) < f(E) + \sum_{n=0}^{\infty} \mathbf{P}^{(E)}(n) \log \binom{n+m-1}{m-1}. \tag{9.13}$$

Proof. In order to prove Proposition 9.4.1 we need to upper bound the first term in the coherent-state rate. We observe that $\sum_{n=0}^{\infty} \mathbf{P}^{(s)}(n) \log \binom{n+m-1}{m-1}$ is a concave function of s. Indeed its second derivative evaluates to

$$
\begin{aligned}
&\frac{d^2}{d^2 s}\{\sum_{n=0}^{\infty}\mathbf{P}^{(s)}(n)\log\binom{n+m-1}{m-1}\}\\
&= s^{-2}\sum_{n=0}^{\infty}\mathbf{P}^{(s)}(n)((s-n)^2-n)\log\binom{n+m-1}{m-1} + s^{-2}\sum_{n=0}^{\infty}\mathbf{P}^{(s)}(n)(s^2\log\binom{n+m-1}{m-1}\\
&+ s(n+1)\log\binom{n+m}{m-1} - s(2s+1)\log\binom{n+m}{m-1})\\
&= \sum_{n=0}^{\infty}\mathbf{P}^{(s)}(n)(\log\binom{n+m-1}{m-1} + \log\binom{n+m+1}{m-1} - 2\log\binom{n+m}{m-1})\\
&\leq \sum_{n=0}^{\infty}\mathbf{P}^{(s)}(n)\sum_{i=1}^{m}\log\frac{(n+i)(n+i+2)}{(n+i+1)^2} < 0
\end{aligned}
\tag{9.14}
$$

where we used $\mathbf{P}^{(s)}(n)n^{\alpha} = s\mathbf{P}^{(s)}(n-1)n^{\alpha-1}$, $\log\binom{x+m-1}{m-1} = \sum_{i=1}^{m}\log\frac{x+i}{i}$, and $\log\frac{x(x+2)}{(x+1)^2} < 0$ by monotonicity of the function $\frac{x}{x+1}$. Therefore, by applying again Jensen's inequality on the integral in s, as well as recalling that $\int ds\, q(s)s = E$, we obtain the following inequality:

$$
\int ds\, q(s)\sum_{n=0}^{\infty}\mathbf{P}^{(s)}(n)\log\binom{n+m-1}{m-1} \leq \sum_{n=0}^{\infty}\mathbf{P}^{(E)}(n)\log\binom{n+m-1}{m-1}, \tag{9.15}
$$

Hence the optimal coherent-state rate can be upper bounded for all E and m by Eq. (9.13). □

Note that the second term in Eq. (9.12) equals the Holevo quantity of an ensemble of coherent states at fixed expected value of the energy, which [JBDD14] evaluated for encodings which are not covariant, therefore suboptimal. In the rest of the article, we employ two upper bounds based on the bounds of Ref. [WW14] and Ref. [CR19], on the capacity of the Poisson channel. Explicitly, the bound $R_{\mathrm{c}}^{\mathrm{up}}(\Phi_m, E)$ is given by Eq. (9.13) with [CR19]:

$$
\begin{aligned}
f(E) &:= E\log\left(\frac{1+\left(1+e^{1+\gamma}\right)E+2E^2}{e^{1+\gamma}E+2E^2}\right)\\
&+ \log\left(1+\frac{1}{\sqrt{2e}}\left(\sqrt{\frac{1+(1+e^{1+\gamma})\,E+2E^2}{1+E}}-1\right)\right)\\
&\geq \max_{q(s)}\left[H\left(\int_0^{+\infty}ds\, q(s)\mathbf{P}^{(s)}\right) - \int_0^{+\infty}ds\, q(s)H(\mathbf{P}^{(s)})\right],
\end{aligned}
\tag{9.16}
$$

where γ is the Euler-Mascheroni constant.

The bound obtained using [WW14] reproduces correctly the expansion of the channel capacity $C(\Phi_m, E)$ at the first two leading orders, but it appears to be larger than $R_{\mathrm{c}}^{\mathrm{up}}(\Phi_m, E)$ everywhere but for extremely low energies. Therefore, we employ the former bound only to derive the low-energy expansion of the coherent-state rate, see Sec. 9.4.3, while the latter bound is employed throughout the rest of the chapter.

9.4.2 Lower bounds on Gaussian rates via discrete-pulse encodings

Randomized on/off modulation (ROOM)

For general m, we can determine an achievable lower bound on both the coherent and squeezed-coherent maximum rates by employing a simple randomized on/off modulation (ROOM) at the encoding: with some probability p we send a Haar-random pulse $\hat{U}\,|\mathbf{r},\mathbf{s}\rangle$ and the vacuum otherwise. The resulting lower bound for general $\mathbf{r}, \mathbf{s}, p$ respecting the mean-energy constraint is

$$\begin{aligned}
R(\Phi_m, E; \mathbf{r}, \mathbf{s}, p) &= p \sum_{n=1}^{\infty} \mathbf{Q}^{(\mathbf{r},\mathbf{s})}(n) \log \binom{n+m-1}{m-1} + \eta\left(1 - p + p\mathbf{Q}^{(\mathbf{r},\mathbf{s})}(0)\right) \\
&+ \sum_{n=1}^{\infty} \eta\left(p\mathbf{Q}^{(\mathbf{r},\mathbf{s})}(n)\right) - p \sum_{n=0}^{\infty} \eta\left(\mathbf{Q}^{(\mathbf{r},\mathbf{s})}(n)\right) \\
&= p \sum_{n=1}^{\infty} \mathbf{Q}^{(\mathbf{r},\mathbf{s})}(n) \log \binom{n+m-1}{m-1} + \eta\left(1 - p + p\mathbf{Q}^{(\mathbf{r},\mathbf{s})}(0)\right) \\
&+ \left(1 - \mathbf{Q}^{(\mathbf{r},\mathbf{s})}(0)\right) \eta(p) - p\,\eta\left(\mathbf{Q}^{(\mathbf{r},\mathbf{s})}(0)\right),
\end{aligned} \tag{9.17}$$

where $\mathbf{Q}^{(\mathbf{r},\mathbf{s})}$ is the total-photon-number distribution of a tensor product of squeezed-coherent states $|\mathbf{r},\mathbf{s}\rangle$ [Yue76; GA90] and $\eta(p) := -p \log p$. In particular, as explained above, for the coherent encoding it always suffices to start with a single-mode pulse, i.e., $\vec{r} = 0$ and $\mathbf{s} = (s, 0, \cdots, 0)$, so that $\mathbf{Q}^{(\mathbf{r},\mathbf{s})}$ reduces to a Poissonian $\mathbf{P}^{(|\alpha|^2)}$.

Consequently, in accordance with Eq. (9.3), we define the best lower bounds on the maximum rate of coherent and squeezed-coherent encodings as

$$R_{\mathrm{c}}^{\mathrm{room}}(\Phi_m, E) := \max_{\mathbf{s},p} R(\Phi_m, E; \mathbf{s}, 0, p), \tag{9.18}$$

$$R_{\mathrm{sc}}^{\mathrm{room}}(\Phi_m, E) := \max_{\mathbf{r},\mathbf{s},p} R(\Phi_m, E; \mathbf{r}, \mathbf{s}, p), \tag{9.19}$$

where the optimization is over values of the parameters respecting the energy constraint. Clearly, in a ROOM communication strategy, the larger the energy of the Haar-random pulses is, the smaller is their joint probability $1 - p$. This fact is particularly clear for coherent encodings, where, as already noted, the problem depends exclusively on

the absolute value of $\mathbf{s}$. Since the energy constraint $E = (1-p)|\mathbf{s}|^2$ imposes $\mathbf{s} = \mathbf{s}_0 := (\sqrt{E/p}, 0, \cdots, 0)$, the optimization of Eq. (9.18) essentially reduces to a single-parameter one:

$$R_{\mathrm{c}}^{\mathrm{room}}(\Phi_m, E) := \max_{\mathbf{p}\in[0,1]} R(\Phi_m, E; \sqrt{E/p}, 0, p). \tag{9.20}$$

On the contrary, for the squeezed-coherent encoding it is difficult to optimize Eq. (9.19) in general. Numerical evidence suggests to concentrate all the energy in one pulse before $\hat{U}$, i.e., $\mathbf{r} = (r, 0, \cdots, 0)$ and $\mathbf{s} = (s, 0, \cdots, 0)$, with $s \in \mathbb{R}$ and $r > 0$. Intuitively, this choice is aimed at reducing the photon-number variance as pointed out after Eq. (9.9).

On/off modulation plus photodetection (OOP)

As explained in Sec. 9.2.1, attaining the maximum rate $R_{\mathrm{c/sc}}^{\mathrm{room}}(\Phi_m, E)$ of a ROOM encoding still requires the realization of collective quantum measurements across several uses of the channel. Hence we consider further a fully explicit communication scheme, employing a non-randomized on/off modulation with coherent, $\mathcal{A}_{\mathrm{c}} := \{|0\rangle, |\mathbf{s}\rangle\}$, or squeezed-coherent signals, $\mathcal{A}_{\mathrm{sc}} := \{|0\rangle, |\mathbf{r}, \mathbf{s}\rangle\}$, plus on/off photodetection measurements (OOP), $\mathcal{M}_{\mathrm{pd}} := \{|0\rangle\langle 0|, I - |0\rangle\langle 0|\}$, whose rate $R_{\mathrm{c/sc}}^{\mathrm{oop}}(\Phi_m, E; \mathcal{A}_{\mathrm{c/sc}}, \mathcal{M}_{\mathrm{pd}})$ can be computed via Eq. (9.4).

Two or more pulses

It is clear that the ROOM encoding can be generalized by considering more than one covariant pulse, obtaining similar expressions to Eq. (9.17). Providing analytical intuition about the behaviour of these strategies is challenging. However, in our numerics we do consider ternary coherent and squeezed-coherent encodings, composed of the vacuum state and two randomized pulses with distinct parameters (see Fig. 9.2 for a depiction of these strategies). As for ROOM, the maximum rate of these encodings can be obtained by optimizing all the parameters subject to the average-energy constraint.

9.4.3 Limiting behaviour of the maximum coherent-state rate

High-energy regime

Let us discuss the high energy regime for coherent-state strategies. The second term of the upper bound in Eq. (9.13) can be further bounded using Jensen's inequality, obtaining the coarser bound

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbf{P}^{(E)}(n) \log \binom{n+m-1}{m-1} &\leq \log \binom{E+m-1}{m-1} \\ &= (m-1)\log E + O(1),\end{aligned} \tag{9.21}$$

which, together with the well-known fact that at high energies the capacity of the Poisson channel is $\frac{1}{2}\log E + O(1)$ [CR19], establishes the following:

Proposition 9.4.2. *The maximum rate of transmission of classical information through the channel Φ_m using high-energy coherent state encodings is given by*

$$R_{\mathrm{c}}(\Phi_m, E) = (m - \frac{1}{2})\log E + O(1). \tag{9.22}$$

We check that this asymptotic rate is achievable with a thermal ensemble of coherent states. An ensemble for which the average state is the thermal state can be obtained by encoding with probability distribution given by a Gamma distribution $q(s) = \left(\frac{m}{E}\right)^m \frac{e^{-\frac{s}{E/m}} s^{m-1}}{(m-1)!}$, indeed

$$\int_0^{+\infty} ds\, q(s)\mathbf{P}^{(s)}(n) = \binom{n+m-1}{m-1}\left(\frac{E}{E+m}\right)^n \left(\frac{m}{E+m}\right)^m = \mathbf{p}^{\mathrm{th}}(n). \tag{9.23}$$

For this distribution one needs to evaluate only the average-output-entropy term. From the well-known fact that $H(\mathbf{P}^{(s)}) \leq \frac{1}{2}\log 2\pi e(s + \frac{1}{12})$ [LM09; ALY10], and from Jensen inequality, we have

$$\int_0^{+\infty} ds\, q(s) H(\mathbf{P}^{(s)}) \leq \int_0^{+\infty} ds\, q(s)\frac{1}{2}\log 2\pi e(s + \frac{1}{12}) \leq \frac{1}{2}\log 2\pi e(E + \frac{1}{12}). \tag{9.24}$$

We then obtain a rate

$$\begin{aligned} R^{\mathrm{th}} &= mg(E/m) - \int_0^{+\infty} ds\, q(s) H(\mathbf{P}^{(s)}) \geq mg(E/m) - \frac{1}{2}\log 2\pi e(E + \frac{1}{12}) \\ &= (m - \frac{1}{2})\log E + O(1). \end{aligned} \tag{9.25}$$

Moreover, by fixing $E/m = k$ and sending m to infinity, we get a rate per use of the transmission line which approaches the identity-channel capacity with energy constraint k:

$$\begin{aligned} \frac{R^{\mathrm{th}}}{m} &= g(E/m) - \int_0^{+\infty} dsp(s) H(\mathbf{P}^{(s)}) \geq g(k) - \frac{1}{2m}\log 2\pi e(km + \frac{1}{12}) \\ &= C(\Phi, k) + O\left(\frac{\log m}{m}\right). \end{aligned} \tag{9.26}$$

Low-energy regime

Let us now discuss the low-energy regime for coherent-state strategies. The series that appears in the bound Eq. (9.13) can be rearranged as a power series in E, and at the leading order (at fixed m) it reads $\sum_{n=0}^{\infty} \mathbf{P}^{(E)}(n) \log \binom{n+m-1}{m-1} = E \log m + o(E)$. Hence the full low-energy expansion of the upper bound Eq. (9.13) reads

$$R_{\mathrm{c}}^{\mathrm{up}}(\Phi_m, E) = E \log \frac{1}{E} + E(c - \gamma + \log m) + o(E), \tag{9.27}$$

where $c = \frac{e^{\frac{1}{2}+\gamma}}{2\sqrt{2}} - 1 \approx 0.038$, and γ is the Euler-Mascheroni constant.

However, at extremely low energies (10^{-60} to 10^{-40}) and for all m, one can use the alternative bound $\tilde{R}_{\mathrm{c}}^{\mathrm{up}}(\Phi_m, E)$, which provides the following asymptotically-tighter upper bound for the rate of our channel:

$$\begin{aligned}\tilde{R}_{\mathrm{c}}^{\mathrm{up}}(\Phi_m, E) = E \log \frac{1}{E} - E \log \log \frac{1}{E} \\ + E(2 + \log 13 - \gamma + \log m) + o(E).\end{aligned} \tag{9.28}$$

An achievable lower bound is instead provided by a corresponding one for the Poisson channel. In the following we will adapt an on/off modulation that attains the Poisson channel capacity with unconstrained decoding at the leading order in E and in general provides a good lower bound for $E \lesssim 1$ [WW14]. Note that this bound can be surpassed at larger energies by that of [Mar07]. The strategy we consider is a randomized on/off modulation (ROOM) and consists in sending a vacuum pulse $|0\rangle$ with probability $1-p$ and otherwise a Haar-random coherent pulse $\hat{U}|\mathbf{s}\rangle \otimes |0\rangle^{\otimes m-1}$ of energy $|\mathbf{s}|^2 = \frac{E}{p}$. Following the same reasoning to obtain the optimal coherent-state rate, it is straightforward to see that the net effect of this encoding is that of inducing an on/off total energy distribution in Eq. 9.12, i.e., $\{q(0) = 1 - p, q(\frac{E}{p}) = p\}$. The corresponding rate is

$$\begin{aligned}R(\Phi_m, E, p) &:= p \sum_{n=1}^{\infty} \mathbf{P}^{(E/p)}(n) \log \binom{n+m-1}{m-1} + \eta\left(1 - p + p\mathbf{P}^{(E/p)}(0)\right) \\ &+ \sum_{n=1}^{\infty} \eta\left(p\mathbf{P}^{(E/p)}(n)\right) - p \sum_{n=0}^{\infty} \eta\left(\mathbf{P}^{(E/p)}(n)\right) \\ &= p \sum_{n=1}^{\infty} \mathbf{P}^{(E/p)}(n) \log \binom{n+m-1}{m-1} \\ &+ \eta\left(1 - p + p\mathbf{P}^{(E/p)}(0)\right) + \left(1 - \mathbf{P}^{(E/p)}(0)\right) \eta(p) - p\eta\left(\mathbf{P}^{(E/p)}(0)\right),\end{aligned}$$

where we used the property $\eta(xy) = x\eta(y) + y\eta(x)$. One can then maximize this function over $p \in [0, 1]$ to obtain the best lower bound $R_{\mathrm{c}}^{\mathrm{room}}(\Phi_m, E) := \max_p R(\Phi_m, E, p)$.

We have already seen that, independently of the value of p, we have $p \sum_{n=1}^{\infty} \mathbf{P}^{(E/p)}(n) \log \binom{n+m-1}{m-1} \leq E \log m + o(E)$. Then we use the fact that for the

remaining terms in Eq. (9.29), at low energies, the maximum is attained for $p = E \log \frac{1}{E}$, see [WW14]. By inserting this value of p we get $p \sum_{n=1}^{\infty} \mathbf{P}^{(E/p)}(n) \log \binom{n+m-1}{m-1} = E \log m + o(E)$ and therefore

$$R_{\mathrm{c}}^{\mathrm{room}}(\Phi_m, E) = E \log \frac{1}{E} - E \log \log \frac{1}{E} + E \log m + o(E). \tag{9.29}$$

Moreover, note that at low energies one can attain this rate at order $O(E)$ by explicit on/off modulation plus photodetection (OOP) that sends in each mode independently a fixed coherent pulse of energy $\frac{E}{pm}$ with probability p and the vacuum otherwise [WW14] (see also [VR94] for a worse-performing generalized PPM strategy). The rate for this strategy is immediately obtained from the on/off rate for the case $m = 1$ [WW14]:

$$R_{\mathrm{c}}^{\mathrm{oop}}(\Phi_m, E) = m R_{\mathrm{c}}^{\mathrm{oop}}(\Phi_1, E/m) = E \log \frac{1}{E} - E \log \log \frac{1}{E} + E \log m + o(E). \tag{9.30}$$

We can summarize the low energy analysis with the following:

Proposition 9.4.3. *The rate of transmission of classical information through the channel Φ_m using low-energy coherent state encodings is bounded by*

$$R_{\mathrm{c}}(\Phi_m, E) \geq E \log \frac{1}{E} - E \log \log \frac{1}{E} + E \log m + o(E) \tag{9.31}$$

$$R_{\mathrm{c}}(\Phi_m, E) \leq E \log \frac{1}{E} - E \log \log \frac{1}{E} + E(2 + \log 13 - \gamma + \log m) + o(E). \tag{9.32}$$

9.4.4 Comparison of all strategies and squeezing advantage

Quantum advantage via squeezing

In Fig. 9.3 we plot the coherent and squeezed-coherent lower bounds $R_{\mathrm{c/sc}}^{\mathrm{room}}$ and the coherent upper bound $R_{\mathrm{c}}^{\mathrm{up}}$ per unit of channel capacity for several values of m, as a function of the mean energy E. First of all, we observe a general increasing trend of the rate with m, implying that one can make use of correlations in the phase noise to increase the Gaussian communication rate.

More importantly, the plot also shows that the use of squeezing provides a large increase of the communication rate for all m and E with respect to its direct coherent counterpart. This is particularly evident for $m = 1$, where there even exists a small range of energies such that $R_{\mathrm{sc}}^{\mathrm{room}}(\Phi_1, E) > R_{\mathrm{c}}^{\mathrm{up}}(\Phi_1, E)$. Since for $m = 1$ it also holds that $R_{\mathrm{sc}}^{\mathrm{room}}(\Phi_1, E) = R_{\mathrm{sc}}^{\mathrm{oop}}(\Phi_1, E)$ (see Sec. 9.4.4), we conclude that the fully explicit OOP squeezed-coherent communication scheme surpasses any coherent-state scheme, answering a question that can be traced back to [ST87]. This proves the existence of an unconditional quantum advantage with respect to classical communication strategies, which can be demonstrated using only experimental-friendly resources such as squeezing and photodetectors.

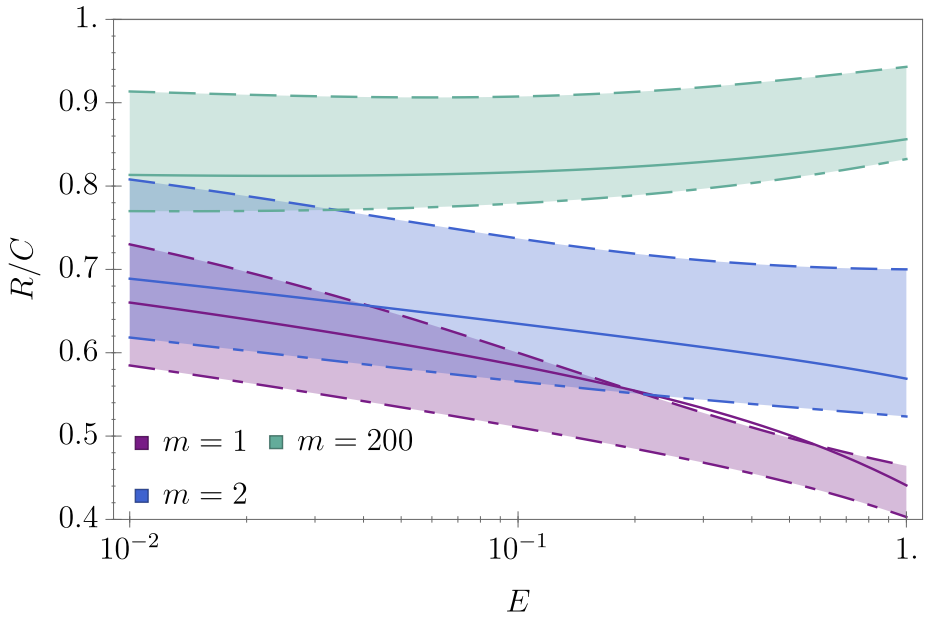


Figure 9.3: Plot (log-linear scale) of several rates per unit of channel capacity $C(\Phi_m, E)$ vs. the average energy E for several values of m: upper (dashed lines) and lower (dot-dashed lines) bounds on the optimal coherent-state rate, $R_{\mathrm{c}}^{\mathrm{up/room}}(\Phi_m, E)$, lower bound (continuous line) on the optimal squeezed-coherent-state rate, $R_{\mathrm{sc}}^{\mathrm{room}}(\Phi_m, E)$. The optimal coherent-state rate lies in the shaded region. Note that as m increases, the coherent-state rate becomes comparable with the capacity. Moreover, squeezing always provides a notable advantage over simple coherent encoding and it can even surpass the coherent-state encoding upper bound for $m = 1$.

Attainability of ROOM rate with fully explicity OOP scheme

We start by observing that, for $m = 1$, the OOP scheme attains the lower bound Eq. (9.17) at all energies both for coherent and squeezed-coherent encodings, i.e., $R_{\mathrm{c/sc}}^{\mathrm{room}}(\Phi_1, E) = R_{\mathrm{c/sc}}^{\mathrm{oop}}(\Phi_1, E)$, implying that both $(m = 1)$-ROOM rates are attainable with an end-to-end explicit protocol.

For $m > 1$ instead we have $R_{\mathrm{c/sc}}^{\mathrm{room}}(\Phi_1, E) > R_{\mathrm{c/sc}}^{\mathrm{oop}}(\Phi_1, E)$ in general (see e.g. Fig. 9.4). Interestingly, in the low-energy regime this gap closes and the same lower bound of (9.31) can be attained by a fully explicit OOP strategy for all m.

Comparison with Fock and ternary encodings

The advantage of a ROOM squeezed-coherent encoding is quite small with respect to that obtained with an on/off encoding using the vacuum and a one-photon Fock state $|1\rangle$,

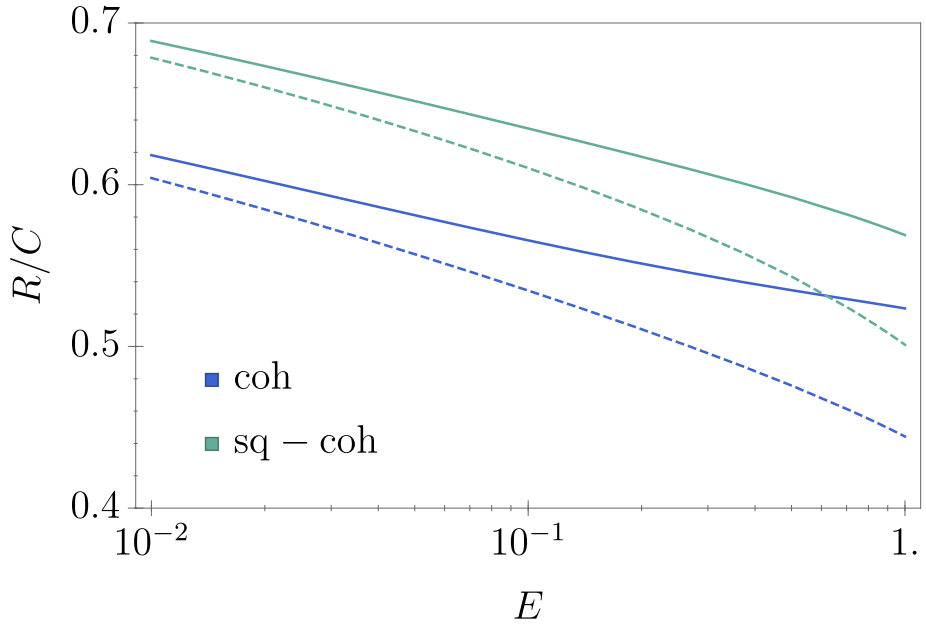


Figure 9.4: Plot (log-linear scale) of several rates per unit of channel capacity: coherent and squeezed-coherent ROOM rate (continuous) and explicit OOP scheme rate (dashed), for $m = 2$.

which is, admittedly, not particularly challenging to produce nowadays. For this reason, we are driven to consider discrete encodings comprising more than two signals.

As we already pointed out, for any ensembles of t states the maximum rate is bounded by $\log t$. The restricted Fock ensembles with photon number up to 1 or 2 states rapidly saturate to respectively $\log 2$ and $\log 3$. Still, as shown in Fig. 9.5, the advantage of squeezed-coherent encodings is enhanced by considering ternary constellations, which can surpass the performance of 0/1 Fock encodings. Importantly, the amount of squeezing required by these optimal encodings is relatively modest, e.g., at $E = 1.1$ the largest $r \simeq 0.62$ corresponds to a squeezing of 5.4dB, while at $E = 2$ the largest $r \simeq 0.68$ corresponds to a squeezing of 5.8dB. Hence our squeezed-coherent encoding is fully within reach of current experimental platforms [Vah+16] (up to 15dB), even for on-chip production [Zha+21](up to 8dB).

Clearly, one can surpass the ternary squeezed-coherent encoding with a ternary Fock encoding, as shown in Fig. 9.5. However, the difficulty of producing Fock states with photon number larger than 1, makes encodings with multiple squeezed-coherent states preferable over those relying on Fock states. Such advantage is further enhanced in the presence of loss, as we will discuss in the next section.

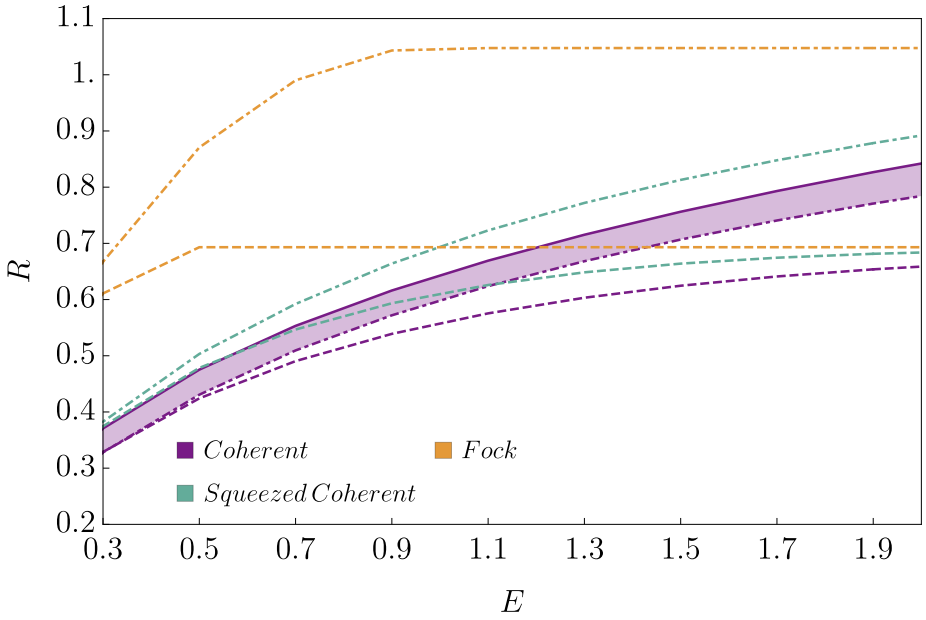


Figure 9.5: $m = 1$, Binary (dashed) and ternary (dot-dashed) rates achievable with Fock encodings with up to two photons (orange), coherent states (violet), and squeezed coherent states (green). The violet solid line is the upper bound on the coherent state rate Eq. (9.33).

9.5 Comparison with Fock encodings in presence of loss

We account for losses by preceding the phase-noise channel Φ_m with a lossy bosonic channel $\mathcal{E}_{\eta,n_{th}}^{\otimes m}$. At the level of the rates, this means replacing ensembles $\mathcal{E} = \{q(x), \hat{\rho}^{(x)}\}$ with $\mathcal{E}_\eta = \{q(x), \mathcal{E}_{\eta,n_{th}}^{\otimes m}(\hat{\rho}^{(x)})\}$ in the rate expression Eq. (9.9).

Since the action of loss on coherent states is $\mathcal{E}_{\eta,0}(|\alpha\rangle\langle\alpha|) = \left|\sqrt{\eta}\alpha\right\rangle\left\langle\sqrt{\eta}\alpha\right|$, we can immediately compute the maximum coherent-state rate in the presence of loss as

$$R_c(\Phi_m \circ \mathcal{E}_{\eta,0}, E) = R_c(\Phi_m, \eta E), \tag{9.33}$$

and employ upper and lower bounds adapted from Sec. 9.4.

To compute the rates of encodings generated by applying a random PI to single-mode squeezed-coherent states, we need instead the photon-number distribution of a generic Gaussian state, which is reported in Appendix B.8, Eq. (B.32). Finally, to compute the rates of encodings generated by Fock states we can use that for $N = 0$ the action of the attenuator on Fock states is

$$\mathcal{E}_{\eta,0}(|n\rangle\langle n|) = \sum_{i=0}^{n} \binom{n}{i} \eta^i (1-\eta)^{n-i} \, |i\rangle\langle i| \, . \tag{9.34}$$

In Figs. 9.6,9.7 we plot these rates in the case $m = 1$ and $m = 2$ and zero-temperature environment, $n_{th} = 0$, showing that squeezed-coherent encodings outperform both coherent and Fock encodings with photon number up to 2 in the presence of a moderate amount of loss, in the regime $E \approx 1$.

9.6 Communication cost of establishing a phase reference

The optimality of covariant encodings makes strategies with phase reference states suboptimal in principle, but it is still worth to compare them with covariant encodings. We consider an encoding employing Ex energy in the first mode to prepare a fixed phase reference state and $E(1-x)$ on the remaining $m-1$ modes with arbitrary encoding, with $0 < x < 1$. By a data-processing argument for the capacity, clearly this rate cannot be better than that of the identity channel on $m-1$ modes with energy constraint $E(1-x)$. Asymptotically at high energy, the leading term of this upper bound is $(m-1)\log E$, independently of x and of the reference state, which is less than the coherent-state rate achieving Eq. (9.22) by $\frac{1}{2}\log E$. We give an expression of the rate with phase synchronization in Appendix B.6, showing that a coherent states encoding achieves this upper bound asymptotically. In Fig. 9.8 we show the comparison in the finite energy regime between a covariant coherent encoding with an average thermal input state and encodings using a truncated phase state $|\psi\rangle = [2xE+1]^{-1/2} \sum_{n=0}^{2xE} |n\rangle$ as reference and a thermal ensemble of coherent states for coding. In fact, while phase estimation procedures [Cav81; Mon06; Yon+12; ŠAF15; SOP15] benefit from super-Poissonian photon-number statistics, the advantage we report in this paper is obtained by trading signals characterized by Poissonian photon-number distribution with sub-Poissionian squeezed-coherent states.

9.7 Remarks

We have analyzed the performance of Gaussian encodings in the presence of phase-noise with a finite decoherence time, such that m successive signals can be sent before losing the phase reference. This is a physically-motivated example of non-Gaussian channel, and we showed that good encodings make an intelligent use of the relative degrees of freedom, rather than trying to synchronize a common phase. Indeed, phase synchronization schemes with quantum-enhanced phase estimation appear to be unfavored with respect to general coherent-state strategies, if the global energy cost is taken into account.

Moreover, we showed that squeezing can greatly enhance the communication rate, as an effect of reducing the entropy of the total-photon-number distribution. In particular for $m = 1$ we proved that, for the first time to our knowledge, an explicit strategy, alternating between the vacuum and a squeezed-coherent state, together with photodetection, outperforms any coherent-state code. This is particularly interesting considering that it can be easily realized with current technology, although the as-yet-unknown optimal coherent-state rate will need in general the use of entangled measurements at the receiver side, which are still challenging.

Finally, we showed that the squeezing advantage over coherent states is robust with respect to additional loss effects in the communication line and that, in this case, squeezed-coherent encodings with multiple pulses can even outperform Fock-state encodings. This fact, in conjunction with the difficulty of realizing photon-number states and the relatively small amount of squeezing required by our strategies, establishes squeezed-coherent states as a robust and efficient coding method for communication without phase-reference.

We leave as open questions: the optimality of strategies employing non-zero squeezing among Gaussian states for any m and E; the sub-optimality of ensembles using states with super-Poissonian statistics, which is good for phase synchronization, with respect to coherent-state strategies. Moreover, we did not consider the possibility of sending entangled squeezed states across the channel uses, which could in principle further enhance the communication rates due to superadditivity.

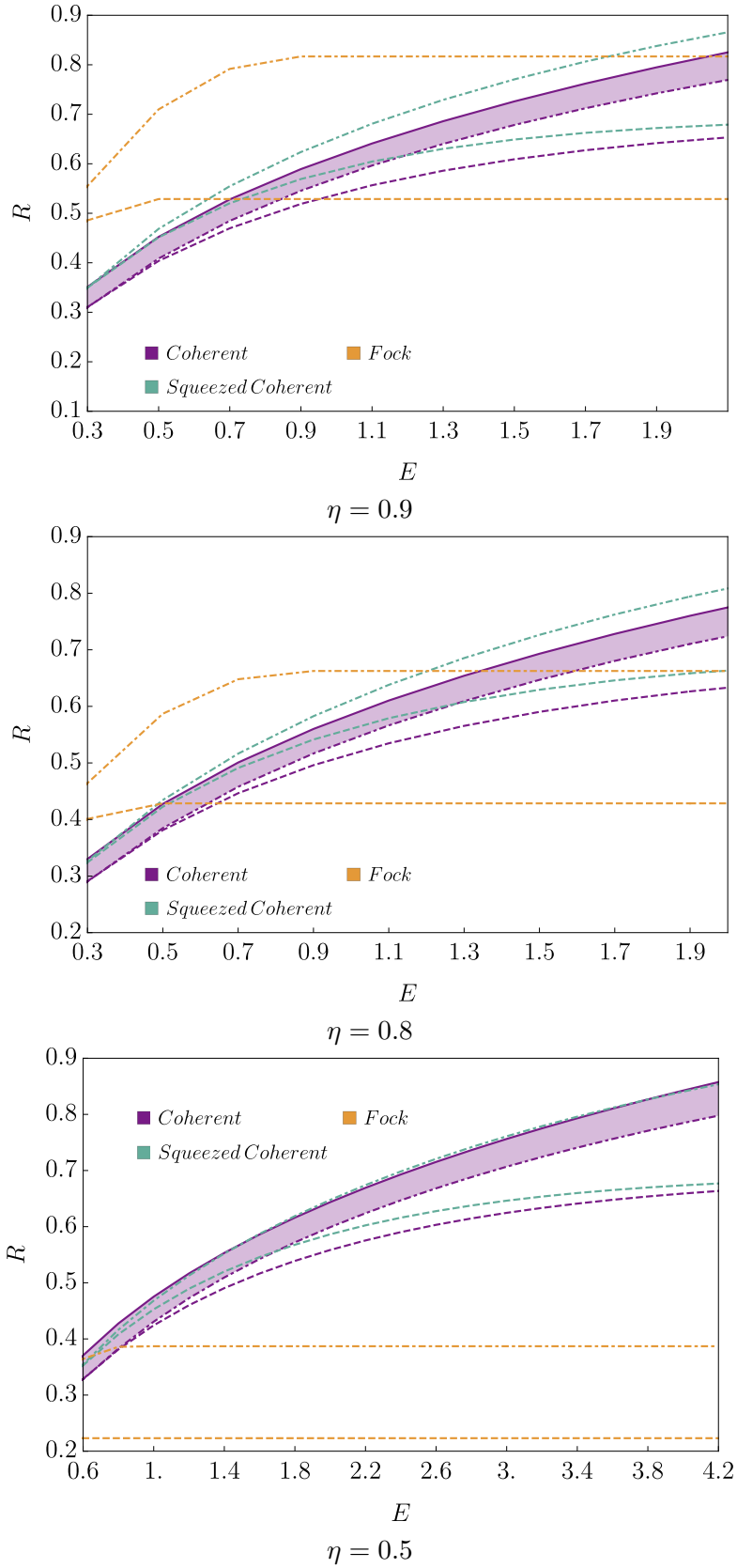


Figure 9.6: $m = 1$, Binary (dashed) and ternary (dot-dashed) rates achievable with Fock encodings with up to three photons (orange), coherent states (violet), and squeezed coherent states (green). The violet solid line is the upper bound on the coherent state rate Eq. 9.33.

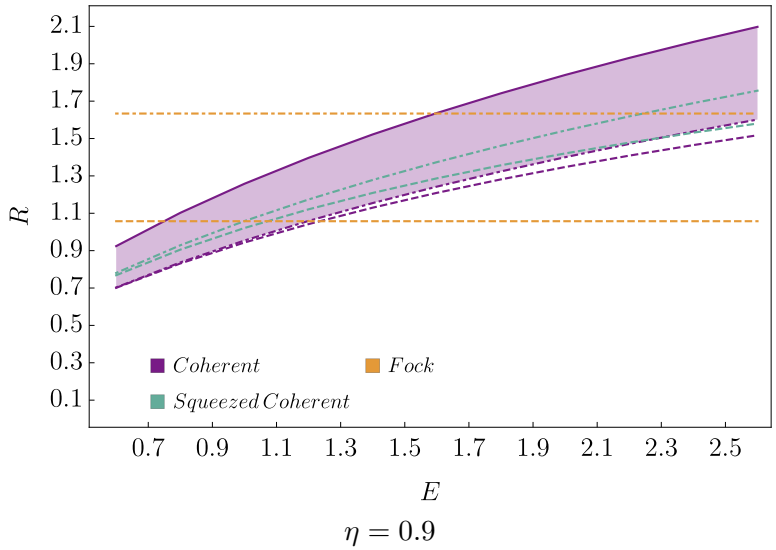


Figure 9.7: $m = 2$, upper bounds on rate with Fock encodings with up to three photons (removing energy constraint), randomized binary (dashed) and randomized ternary (dot-dashed) rates achievable with coherent states (violet), and squeezed coherent states (green). The violet solid line is the upper bound on the coherent state rate Eq. 9.33.

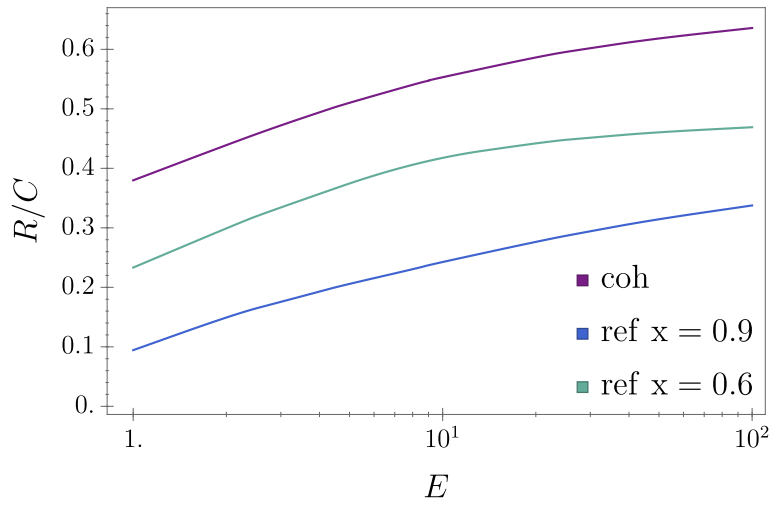


Figure 9.8: Plot (log-linear scale) of several rates per unity of channel capacity: Gaussian coherent-state ensembles on all the $m = 2$ modes or with a fixed reference on one mode.

Chapter 10

Conclusions

The landscape of quantum statistical inference problems is vast and widely uncharted. The fundamental nature of the questions that can be posed, and the potential transformative implications of the answers, make the venture fascinating. We explored some corners in this landscape, where many other interesting problems are in sight. As a common theme across the chapters, we made use of symmetry principles to unravel optimizations and computations.

The first part of our results, Chapters 5, 6, 7 focus on the estimation of unitarily invariant properties of sets of states. In Chapter 5 we computed several asymptotic corrections to the probabilities of error in classifying quantum states with large number of training examples, as an instance of quantum supervised learning. In Chapter 6, we investigated the ultimate limits in the estimation of the overlap between two unknown pure states, comparing the performances of the optimal strategy with several alternatives. This analysis puts on firmer grounds the quantitative understanding of these important primitives. In both Chapters 5, 6, the goal has been to have the most precise answer in the asymptotic regime of large number of copies. This is compelling when the analytic computation of the optimal performance can be actually done. This is the case for Chapter 5, 6, where we concentrate on pure states or qubit states. It would be much more complicated to follow the same approach for general mixed states. However, in realistic situations one could be happy to have a guarantee on the quality of the estimation, rather than the most stringent analysis. Moreover, the asymptotic regime of large number of copies may be not significant if other extensive parameters have to be considered. Indeed, in Chapter 7, we take a different approach, and we evaluate how many copies of labeled states we need to certify if they are all equal or they differ for more than an arbitrary threshold, as a function of the number of states and the dimension. We compute this

dependence up to a multiplicative constant: this is enough to quantify the complexity of the inference problem. In the same way, one could ask how many copies of the template states are needed and are sufficient for realizing a learning machine for state discrimination, such that the difference between the probability of error of the protocol and the optimal probability of error for discriminating between the template states is at most a small quantity. This sample complexity would depend on the number of different template states and the dimension of the template states. If many copies of the test state are also available, one could also determine the sample complexity for identifying the correct classification with high probability. The sample complexity of estimating the trace distance between two states or the Holevo quantity of a set of states is also open. Another important development would be to study the many variations of independence testing [Yu19; HT16]. We also mention other related open sample complexity problems. Spectrum estimation [OW15] and von Neumann entropy estimation [AKG19] are still not completely solved; in [OW15] and [AKG19] it is shown that the sample complexity for both problems is $O(d^2)$, while the sample complexity for quantum tomography is $\Theta(d^2)$ and it is reasonable that spectrum and Von Neumann estimation should be easier. For a comparison, in the classical case the sample complexity of learning the spectrum and the Shannon entropy is $\Theta(d/\log d)$, which is less than the complexity $\Theta(d)$ for learning the distribution in total variation distance. Another open problem is establishing the sample complexity of shadow tomography [Aar18; HKP20], a very interesting approach to learning properties of states. Also in this case, upper and lower bounds exists but do not match. In addition to the model in which the measurement takes as input copies of the states, one could also consider a quantum query model, where the agent can access an unknown circuit producing the states [GL20]. This model could be more appropriate in some situations, and still to be fully explored.

In the second part of this thesis, we improve bounds on optimal classical and quantum communication rates with physically motivated models. In Chapter 8 we developed a method for constructing degradable extensions of convex combination of channels, for which the quantum and private capacity can be easily computed, establishing upper bounds on the quantum and private capacity of the original channel. This method is very flexible, and gives the best upper bounds on a large family of channels, at the time of writing. We do not exhaust the possibilities of this method, since we do not perform a full numerical optimization over the extensions we find. Moreover, there is the possibility that degradable extensions of several uses of the channel can give even better bounds. While this approach is not definitive, since we know that the quantum and private capacity do not coincide in general, it is still interesting to pursue this route. The most pressing theoretical puzzle is to have better bounds on the quantum and private capacity for channels which are far from the identity channel. The current implementations of our method are still more satisfactory in the low noise region, but the

question needs further study. It is also desirable to make the construction of degradable extensions for Gaussian channels more systematic: it is possible that better extensions of the thermal attenuator and amplifier, more akin to the flagged degradable extension of the additive noise channel than to weak-degradability constructions, give tighter bounds in the low noise regime.

In Chapter 9 we started a systematic investigation of the optimal classical communication rates in presence of phase noise and loss. We present conclusive numerical evidence that squeezing helps in these communication settings, at variance with what happens with phase-insensitive noise models. We also find several analytical and asymptotic results on optimal rates with coherent states. The model we consider can be a useful mathematical tool to devise good codes in realistic setting. At the same time, it is very difficult to obtain analytical answers. It would be important to improve the upper bounds on coherent states rates for long coherence times, and to find evidence of advantage of squeezing in higher energy regimes.

Appendix A

Appendix: statistics of quantum invariant measurements

In this appendix we collect computations which we needed in the main text, in Chapters 5, 6, 7.

A.1 Spectrum of the average operator of states at fixed overlap

In this section we compute the spectrum of

$$\rho(c) = \int_{\mathrm{SU}(d)} \mathrm{d}U\, U^{\otimes M+N} \left[(|\psi\rangle\langle\psi|)^{\otimes N} \otimes (|\phi\rangle\langle\phi|)^{\otimes M} \right] U^{\dagger\otimes(M+N)} \tag{A.1}$$

$$= \sum_J P_{M,N}(J|c) \frac{I_{\lambda_J}}{\omega_{\lambda_J}^{(d)}} \otimes |J_{M,N}\rangle\langle J_{M,N}| \,. \tag{A.2}$$

with $|\langle\psi|\phi\rangle|^2 = c$. As explained in Theorem 4.2.6, we can compute the spectrum for $d = 2$, which suffices to determine it for any dimension. Let us define $J_{\max} = \frac{M+N}{2}$ and $J_{\min} = \frac{|M-N|}{2}$. Using the addition rules for angular momentum on $\mathcal{H}_2^{\otimes M+N}$ we can write

$$\begin{aligned} |\psi\rangle^{\otimes N} \otimes |\phi\rangle^{\otimes M} &= \left|\frac{N}{2}, \frac{N}{2}\right\rangle \otimes \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} D^{(\frac{M}{2})}_{k,\frac{M}{2}}(2\arccos\sqrt{c}) \left|\frac{M}{2}, k\right\rangle \\ &= \sum_{J=J_{\min}}^{J_{\max}} \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} C^{J,\frac{N}{2}+k}_{\frac{N}{2},\frac{N}{2};\frac{M}{2},k} D^{(\frac{M}{2})}_{k,\frac{M}{2}}(2\arccos\sqrt{c}) \left|J, \frac{N}{2}+k\right\rangle, \end{aligned} \tag{A.3}$$

where $C^{J,\frac{N}{2}+k}_{\frac{N}{2},\frac{N}{2};\frac{M}{2},k} = \left\langle J, \frac{N}{2}+k \middle| \frac{N}{2}, \frac{N}{2}; \frac{M}{2}, k \right\rangle$ are the Clebsch-Gordan coefficients and $|J,m\rangle$, and $\frac{|M-N|}{2} \leq J \leq \frac{M+N}{2}$ are a basis for $\mathcal{U}_{\lambda_J}(\mathrm{SU}(d)) \otimes \mathcal{V}_{\lambda_J}(S_{M+N})$, $\lambda_J = (\frac{M+N}{2} + J, \frac{M+N}{2} - J, 0, ..., 0)$, and they are eigenvectors of $\rho(c)$ with eigenvalue $\frac{P_{M,N}(J|c)}{\omega^{(d)}_{\lambda_J}}$.

Evaluating these matrix elements we obtain:

$$
\begin{aligned}
P_{M,N}(J|c) &= \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} \left(C^{J,\frac{N}{2}+k}_{\frac{N}{2},\frac{N}{2};\frac{M}{2},k} D^{(\frac{M}{2})}_{k\frac{M}{2}}(2\arccos\sqrt{c}) \right)^2 \\
&= \frac{(2J+1)(J+J_{\min})!N!}{(J-J_{\min})!(J_{\max}-J)!(J_{\max}+1+J)!} \\
&\times \sum_{k=-\frac{M}{2}}^{J-\frac{N}{2}} \frac{(\frac{M}{2}-k)!(J+\frac{N}{2}+k)!}{(J-\frac{N}{2}-k)!(\frac{M}{2}+k)!} \left(D^{(\frac{M}{2})}_{k\frac{M}{2}}(2\arccos\sqrt{c}) \right)^2 \qquad \text{(A.4)} \\
&= \frac{(2J+1)(J+J_{\min})!N!M!}{(J-J_{\min})!(J_{\max}-J)!(J_{\max}+1+J)!} \\
&\times \sum_{k=-\frac{M}{2}}^{J-\frac{N}{2}} \frac{(J+\frac{N}{2}+k)!}{(J-\frac{N}{2}-k)!(\frac{M}{2}+k)!^2} (1-c)^{\frac{M}{2}-k} c^{\frac{M}{2}+k} \qquad \text{(A.5)} \\
&= \frac{(2J+1)N!M!(1-c)^M}{(J_{\max}-J)!(J_{\max}+1+J)!} P^{(0,-2J_{\min})}_{J+J_{\min}}\left(\frac{1+c}{1-c}\right), \qquad \text{(A.6)}
\end{aligned}
$$

where we used the following expression of the Wigner D matrix in going from the second to the third line in Eq. (A.6)

$$
D^{(J)}_{z',z}(\theta) = \sqrt{\frac{(J+z)!(J-z)!}{(J+z')!(J-z')!}} \sin^{(z-z')}\left(\frac{\theta}{2}\right) \cos^{(z+z')}\left(\frac{\theta}{2}\right) P^{(z-z',z+z')}_{(J-z)}(\cos\theta), \qquad \text{(A.7)}
$$

with $P^{(\alpha,\beta)}_n(x)$ the Jacobi polynomials, defined in general as

$$
P^{(\alpha,\beta)}_n(x) = \frac{\Gamma(\alpha+n+1)}{n!\Gamma(\alpha+\beta+n+1)} \sum_{m=0}^{n} \binom{n}{m} \frac{\Gamma(\alpha+\beta+n+m+1)}{\Gamma(\alpha+m+1)} \left(\frac{x-1}{2}\right)^m. \qquad \text{(A.8)}
$$

In the particular case $M = n+1$, $N = n$, $c = \sin^2\frac{\theta}{2}$, we obtain

$$
\begin{aligned}
P_{n+1,n}(J|c) = \sum_h & \frac{n!}{(\frac{n}{2}+h)!\,(\frac{n}{2}-h)!} \left(\cos^2\left(\frac{\pi-\theta}{2}\right)\right)^{\frac{n}{2}+h} \left(\sin^2\left(\frac{\pi-\theta}{2}\right)\right)^{\frac{n}{2}-h} \\
&\times \frac{2\,(\frac{n}{2}-h)!\,(\frac{n}{2}+h+q+\frac{1}{2})!\,(n+1)!}{(\frac{n}{2}-h+q-\frac{1}{2})!\,(\frac{n}{2}+h)!\,(n-q+\frac{1}{2})!\,(n+q+\frac{3}{2})!}. \qquad \text{(A.9)}
\end{aligned}
$$

A.2 Averaged multiqubit states

By Schur-Weyl duality, for ρ with Bloch vector of modulus r one has the identity

$$\int dU \left(U\rho U^\dagger\right)^{\otimes n} = \oplus_j f_j^{(n)}(r) I_j \otimes I_{j,n}, \tag{A.10}$$

where I_j is the identity opertor on $\mathcal{U}_j$ and $I_{j,n}$ is the identity operator on $\mathcal{V}_{j,n}$, as defined in Eq. (5.18)

Let ρ a qubit density matrix characterized by Bloch vector of length r which, without loss of generality we shall assume to be oriented in the positive $\hat{z}$ direction, i.e. $\rho = \left(\frac{1+r}{2}\right)|\uparrow\rangle\langle\uparrow| + \left(\frac{1-r}{2}\right)|\downarrow\rangle\langle\downarrow|$ with $|\uparrow\rangle$, $|\downarrow\rangle$ being the eigenvectors of σ_z. We notice that its n-th tensor power can be expressed as

$$\rho^{\otimes n} = \sum_{l=0}^{n} \left(\frac{1+r}{2}\right)^l \left(\frac{1-r}{2}\right)^{n-l} B_l^{(n)}, \tag{A.11}$$

with

$$B_l^{(n)} \equiv \sum_{\tau\in S_n} \mathbf{s}_n(\tau) \left(|\uparrow\rangle\langle\uparrow|^{\otimes l} \otimes |\downarrow\rangle\langle\downarrow|^{\otimes n-l}\right) \mathbf{s}_n(\tau)^\dagger \,,$$

By construction $B_l^{(n)}$ is the projector on the eigenspace at fixed total angular momentum J_z, therefore it is diagonal in every basis of eigenvectors of J^2, J_z. In particular its support is given by the vectors $\left|j, l-\frac{n}{2}\right\rangle_i$ in each irreducible representation with total angular momentum $J^2 = j(j+1)$ and $l \in \{\frac{n}{2}-j, \cdots, \frac{n}{2}+j\}$, the index i labelling accounting for the multiplicity of the representation, i.e.

$$B_l^{(n)} = \oplus_{j\geq|l-\frac{n}{2}|} \oplus_i \left|j, l-\tfrac{n}{2}\right\rangle_i \left\langle j, l-\tfrac{n}{2}\right| \,. \tag{A.12}$$

We can thus write

$$\rho^{\otimes n} = \sum_{l=0}^{n} \left(\frac{1+r}{2}\right)^l \left(\frac{1-r}{2}\right)^{n-l} \oplus_{j\geq|l-\frac{n}{2}|} \oplus_i \left|j, l-\tfrac{n}{2}\right\rangle_i \left\langle j, l-\tfrac{n}{2}\right|, \tag{A.13}$$

Consider then the operator

$$P_l^{(n)} \;:=\; \int dU U^{\otimes n} B_l^{(n)} {U^\dagger}^{\otimes n} \,. \tag{A.14}$$

Performing the integral, we obtain

$$P_l^{(n)} \;=\; \oplus_{j\geq|l-\frac{n}{2}|} \frac{I_j}{2j+1} \otimes I_{j,n}, \tag{A.15}$$

where I_j is the identity opertor on $\mathcal{U}_j$ and $I_{j,n}$ is the identity operator on $\mathcal{V}_{j,n}$.

Accordingly we have

$$\begin{aligned}\int dU \left(U\rho U^\dagger\right)^{\otimes n} &= \oplus_j \sum_{l=\frac{n}{2}-j}^{\frac{n}{2}+j} \left(\frac{1+r}{2}\right)^l \left(\frac{1-r}{2}\right)^{n-l} \frac{I_j}{2j+1} \otimes I_{j,n} \\ &= \oplus_j f_j^{(n)}(r) I_j \otimes I_{j,n}, \end{aligned} \tag{A.16}$$

with $f_j^{(n)}(r)$ as in (5.35).

A.3 Asymptotic expansion of weighted sums

Lemma A.3.1. *Consider a sequence of probability distributions $\{P_n\}$ for a sequence of random variables Q_n taking integer values from a_n to b_n, with mean μ_n, satisfying a concentration inequality around the mean:*

$$P_n(|Q_n - \mu_n| > nt) < \frac{C}{n^\alpha}, \tag{A.17}$$

for some positive constant C, and a sequence of functions $\{f_n(x)\}$ with the following properties:

- *i) $f_n(x)$ is analytic in the region $|x - \mu_n| \leq nt$,*
- *ii) $|f_n(x)| \leq Cn^\beta$ in the region $|x - \mu_n| > nt$,*
- *iii) For integers $0 \leq k < k_0$, k_0 even, $|\frac{d^k f_n}{dx^k}(\mu_n)|x^k < Cn^\beta$ in the region $|x - \mu_n| > nt$.*
- *iv) $|\frac{d^{k_0} f_n(x)}{dx^{k_0}}|E[(Q_n - \mu_n)^{k_0}] < C\frac{1}{n^{\alpha-\beta}}$ in the region $|x - \mu_n| \leq nt$.*

Then we have the following asymptotic expansion, for any a'_n, b'_n such that $a'_n < \mu_n - nt$, $b'_n > \mu_n + nt$,.

$$\sum_{q=a'_n}^{b'_n} P_n(q) f_n(q) = \sum_{k=0}^{k_0-1} \frac{d^k f_n}{dx^k}(\mu_n) \mathbb{E}[(Q_n - \mu_n)^k] + O\left(\frac{1}{n^{\alpha-\beta}}\right) \tag{A.18}$$

Proof. We have the following chain of equalities

$$\begin{aligned}
&\sum_{q=a'_n}^{b'_n} P_n(q)f_n(q) = \sum_{q=a'_n}^{\mu_n-nt} P_n(q)f_n(q) + \sum_{q=\mu_n-nt}^{\mu_n+nt} P_n(q)f_n(q) + \sum_{q=\mu_n+nt}^{b'_n} P_n(q)f_n(q) \\
&= \sum_{q=\mu_n-nt}^{\mu_n+nt} P_n(q)f_n(q) + O\left(\frac{1}{n^{\alpha-\beta}}\right) \\
&= \sum_{q=\mu_n-nt}^{\mu_n+nt} P_n(q)\left(\sum_{k=0}^{k_0-1}\frac{d^{k_0}f_n}{dx^{k_0}}(\mu_n)(q-\mu_n)^k + \frac{d^{k_0}f_n}{dx^{k_0}}(q^*)(q-\mu_n)^{k_0}\right) \\
&= \sum_{q=\mu_n-nt}^{\mu_n+nt} P_n(q)\left(\sum_{k=0}^{k_0-1}\frac{d^{k_0}f_n}{dx^{k_0}}(\mu_n)(q-\mu_n)^k\right) + O\left(\frac{1}{n^{\alpha-\beta}}\right) \\
&= \sum_{q=a_n}^{b_n} P_n(q)\left(\sum_{k=0}^{k_0-1}\frac{d^{k_0}f_n}{dx^{k_0}}(\mu_n)(q-\mu_n)^k\right) + O\left(\frac{1}{n^{\alpha-\beta}}\right) \\
&= \sum_{k=0}^{k_0-1}\frac{d^k f_n}{dx^k}(\mu_n)\mathbb{E}[(Q_n-\mu_n)^k] + O\left(\frac{1}{n^{\alpha-\beta}}\right),
\end{aligned} \tag{A.19}$$

where in the first equality we just split the summation, in the second we use ii) to bound the tail contribution, in the third we Taylor expand $f_n(x)$ around μ_n, using the Lagrange remainder for some q^*, satysfying $|q^*-\mu_n| < nt$. In the fourth we use iv) to bound the Lagrange remainder contribution, and in the fifth we extend the summation to a_n, b_n by adding tails which can be bounded using iii).

□

We use this lemma repeatedly to evaluate several nested weighted sums in Chapters 5 and 6, albeit we do not show here that all the conditions are met for all the cases. As an example, in scenario iii) (**??**), in Chapter 5, we have that putting $h = \frac{ns}{2}$ and $\frac{1+r}{2} = \cos\left(\frac{\pi-\theta}{2}\right)$ we notice that the distribution

$$P_\theta^{(n)}(s) \equiv D^{\frac{n}{2}}_{h,\frac{n}{2}}(U_0)D^{\frac{n}{2}}_{\frac{n}{2},h}(U_0^\dagger)\,, \tag{A.20}$$

defined in Eq. (5.81) is a binomial distribution in the variable ns, with moments

$$\begin{aligned}
\mu_1 &= \mathbb{E}[s] = r \ , \\
\mu_2 &= \mathbb{E}[(s-\mu_1)^2] = \frac{1-r^2}{n} \ , \\
\mu_3 &= \mathbb{E}[(s-\mu_1)^3] = 2r\frac{1-r^2}{n} \ , \\
\mu_4 &= \mathbb{E}[(s-\mu_1)^4] = \frac{(-1+r^2)(2-6r^2+3n(-1+r^2))}{n^3} \ . \\
\mu_5 &= \mathbb{E}[(s-\mu_1)^5] = \frac{(1-r)r(2r-1)(2(5n-6)(r-1)r-1)}{n^4} \ . \\
\mu_6 &= \mathbb{E}[(s-\mu_1)^6] \\
&= \frac{(1-r)r\left(5(r-1)r\left(3n^2(r-1)r+n(-26(r-1)r-5)+6(1-2r)^2\right)+1\right)}{n^5}
\end{aligned} \tag{A.21}$$

and satisfying a tail bound [CT05]

$$\sum_{|ns-nr|>n\epsilon} P_\theta^{(n)}(s) \leq (n+1)^2 e^{-n\min_{|ns-nr|>n\epsilon} D(s||r)}, \tag{A.22}$$

where $D(s||r)$ is the relative entropy between distributions $(s,1-s)$, $(r,1-r)$, and $D(s||r) > \frac{|r-s|^2}{\ln 2}$ by Pinsker's inequality [CT05].

Instead, setting $h = \frac{ns}{2}$, we notice that the moments of the distribution

$$P_h^{(n)}(q) \equiv \frac{2\left(\frac{n}{2}-h\right)!(n+1)!}{\left(\frac{n}{2}+h\right)!} \times \frac{\left(\frac{n}{2}+h+q+\frac{1}{2}\right)!}{\left(q-\frac{1}{2}-\frac{n}{2}-h\right)!\left(n-q+\frac{1}{2}\right)!\left(n+q+\frac{3}{2}\right)!} \ , \tag{A.23}$$

defined in (5.81), can be expressed in terms of Euler gamma functions as follows

$$\begin{aligned}
\mu_1 &= \mathbb{E}[q] = -\frac{1}{2} + \frac{\Gamma(1/2+h+n/2)\Gamma(2+n)}{\Gamma(1+h+n/2)\Gamma(3/2+n)} \\
&= \frac{n\sqrt{1+s}}{\sqrt{2}} - \frac{1}{2} + \frac{11+5s}{8\sqrt{2}\sqrt{1+s}} + \frac{9+14s-23s^2}{128\sqrt{2}n(1+s)^{3/2}} + O\left(\frac{1}{n^2}\right) \ ,
\end{aligned} \tag{A.24}$$

$$\begin{aligned}
\mu_2 &= \mathbb{E}[(q-\mu_1)^2] = \frac{1}{2}(1+n)(2+2h+n) - \frac{\Gamma(3/2+h+n/2)^2\Gamma(2+n)^2}{\Gamma(1+h+n/2)^2\Gamma(3/2+n)^2} \\
&= \frac{1}{8}n(1-s) + \frac{-1+2s-s^2}{64(1+s)} + O\left(\frac{1}{n}\right)
\end{aligned} \tag{A.25}$$

$$\begin{aligned}\mu_3 = \mathbb{E}[(q-\mu_1)^3] &= \frac{-(8+2h(5+4n)+n(11+4n))\Gamma(3/2+h+n/2)\Gamma(2+n)}{\Gamma(1+h+n/2)\Gamma(3/2+n)} \\ &+ \frac{8\Gamma(3/2+h+n/2)^3\Gamma(2+n)^3}{4\Gamma(1+h+n/2)^3\Gamma(3/2+n)^3} = \frac{(-1+s)^2 n}{32\sqrt{2}\sqrt{1+s}} + O(1) \ ,\end{aligned} \tag{A.26}$$

$$\begin{aligned}\mu_4 = \mathbb{E}[(q-\mu_1)^4] &= \frac{(1+n)(4+10n+4h^2n+6n^2+n^3+4h(1+3n+n^2))}{4} \\ &- \frac{3\Gamma(3/2+h+n/2)^4\Gamma(2+n)^4}{\Gamma(1+h+n/2)^4\Gamma(3/2+n)^4} \\ &+ \frac{(2+2n+n^2+2h(2+n))\pi^2\Gamma(2+2h+n)^2\Gamma(3+2n)^2}{4^{3+2h+3n}\Gamma(1+h+n/2)^4\Gamma(3/2+n)^4} \\ &= \frac{3}{6}4(1-2s+s^2)n^2 + O(n) \ .\end{aligned} \tag{A.27}$$

$$\begin{aligned}\mu_5 &= \mathbb{E}[(q-\mu_1)^5] \\ &= -\{[\left(h^2\left(64n^2+80n+4\right)+4h\left(16n^3+64n^2+71n+19\right)\right) \\ &+ 16n^4+108n^3+241n^2+218n+64] \\ &\times \Gamma\left(n+\frac{3}{2}\right)^4\Gamma(n+2)\Gamma\left(h+\frac{n}{2}+1\right)^4\Gamma\left(h+\frac{n}{2}+\frac{3}{2}\right) \\ &+ 64\Gamma(n+2)^5\Gamma\left(h+\frac{n}{2}+\frac{3}{2}\right)^5 \\ &+ 40(n-2h)\Gamma\left(n+\frac{3}{2}\right)^2\Gamma(n+2)^3\Gamma\left(h+\frac{n}{2}+1\right)^2\Gamma\left(h+\frac{n}{2}+\frac{3}{2}\right)^3\} \\ &\times \frac{1}{16\Gamma\left(n+\frac{3}{2}\right)^5\Gamma\left(h+\frac{n}{2}+1\right)^5} \\ &= \frac{5n^2(1-s)^3}{128\left(\sqrt{2}\sqrt{s+1}\right)} + O(n)\end{aligned} \tag{A.28}$$

$$
\begin{aligned}
\mu_6 &= \mathbb{E}[(q-\mu_1)^6] \\
&= \frac{1}{8}\left(-\frac{20\left(n^2(s+1)-n(s-5)+2\right)\Gamma(n+2)^4\Gamma\left(\frac{1}{2}(sn+n+3)\right)^4}{\Gamma\left(n+\frac{3}{2}\right)^4\Gamma\left(\frac{1}{2}(sn+n+2)\right)^4}\right) \\
&+ \frac{3}{8}\left(6n^4(s+1)^2+2n^3\left(5s^2+24s+19\right)+n^2\left(s^2+62s+81\right)+6n(3s+13)+24\right) \\
&\times \frac{\Gamma(n+2)^2\Gamma\left(\frac{1}{2}(sn+n+3)\right)^2}{\Gamma\left(n+\frac{3}{2}\right)^2\Gamma\left(\frac{1}{2}(sn+n+2)\right)^2} \\
&+ \frac{n+1}{8}\left(n^5(s+1)^3-n^4(s-11)(s+1)^2+n^3\left(-2s^2+40s+42\right)\right) \\
&- \frac{n+1}{8}\left(16n^2(s+4)+4n(s+9)+8\right) \\
&- \frac{1}{8}\left(\frac{40\Gamma(n+2)^6\Gamma\left(\frac{1}{2}(sn+n+3)\right)^6}{\Gamma\left(n+\frac{3}{2}\right)^6\Gamma\left(\frac{1}{2}(sn+n+2)\right)^6}\right) \\
&= \frac{1}{512}\left(15n^3\left(1-s^3+3s^2-3s\right)\right)+O\left(n^2\right)\ . \qquad \text{(A.29)}
\end{aligned}
$$

From the concentration inequality applied to the sixth moment [BLM13]

$$P(|Q_n-\mathbb{E}[Q_n]|>nt)\leq\frac{\mu_6}{n^6t^6}. \qquad \text{(A.30)}$$

We use this concentration bound to apply Lemma A.3.1 and obtain an expansion up to order $O\left(\frac{1}{n^2}\right)$.

A.4 Asymptotics of the Fisher information

We begin by recalling the definition of the Fisher information for the overlap estimation problem

$$H(c)=\sum_{J=J_{\min}}^{J_{\max}}P_{M,N}(J|c)\left(\frac{\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}}{P_{M,N}(J|c)}\right)^2=\sum_{J=J_{\min}}^{J_{\max}}\frac{\left(\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}\right)^2}{P_{M,N}(J|c)}. \qquad \text{(A.31)}$$

Using the identity

$$\frac{\mathrm{d}^m P_n^{(\alpha,\beta)}(x)}{\mathrm{d}x^m}=\frac{(\alpha+\beta+n+m)!}{2^m(\alpha+\beta+n)!}P_{n-m}^{(\alpha+m,\beta+m)}(x), \qquad \text{(A.32)}$$

it follows that

$$\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c} = \frac{(2J+1)M!N!(1-c)^{M-2}}{(J_{\max}-J)!\,(J+J_{\max}+1)!} \times \left((J-J_{\min}+1)P_{J+J_{\min}-1}^{(1,1-2J_{\min})}\left(\frac{1+c}{1-c}\right) - (1-c)MP_{J+J_{\min}}^{(0,-2J_{\min})}\left(\frac{1+c}{1-c}\right)\right), \tag{A.33}$$

and

$$\begin{aligned}\frac{1}{P_{M,N}(J|c)}\left(\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}\right)^2 &= \frac{(2J+1)M!N!(1-c)^{M-4}}{(J_{\max}-J)!\,(J+J_{\max}+1)!} \\ &\times \left((J-J_{\min}+1)^2 P_{J+J_{\min}-1}^{(1,1-2J_{\min})}\left(\frac{1+c}{1-c}\right)\frac{P_{J+J_{\min}-1}^{(1,1-2J_{\min})}\left(\frac{1+c}{1-c}\right)}{P_{J+J_{\min}}^{(0,-2J_{\min})}\left(\frac{1+c}{1-c}\right)}\right. \\ &- 2(J-J_{\min}+1)(1-c)MP_{J+J_{\min}-1}^{(1,1-2J_{\min})}\left(\frac{1+c}{1-c}\right) \\ &\left.+(1-c)^2M^2P_{J+J_{\min}}^{(0,-2J_{\min})}\left(\frac{1+c}{1-c}\right)\right),\end{aligned} \tag{A.34}$$

For $x \geq 1$ the following asymptotic expansion for the Jacobi polynomials holds [Ell71] (note that we use a different asymptotics than the one used in the original paper [Fan+20a], taken from [Sze59], since it lets us to compute the next to leading order). Defining the function

$$Q_n^{(\alpha,\beta)}(x) = \frac{\left(\sqrt{x^2-1}+x\right)^{\frac{1}{2}(\alpha+\beta+1)+n}\Gamma(\alpha+\beta+2n+1)}{2^{\frac{1}{2}(\alpha+\beta+1)+2n}\left((x-1)^{\frac{1}{4}(2\alpha+1)}(x+1)^{\frac{1}{4}(2\beta+1)}\right)(\Gamma(n+1)\Gamma(\alpha+\beta+n+1))}, \tag{A.35}$$

one has

$$\frac{P_n^{(\alpha,\beta)}(x)}{Q_n^{(\alpha,\beta)}(x)} = \sum_{s=0}^{\infty}\frac{f_s(\alpha,\beta,x)}{(2n+\alpha+\beta+1)^s}, \tag{A.36}$$

where we should understand this expression as asymptotic expansion valid on $x \in [1+\delta, +\infty)$, $\delta > 0$. We have

$$f_0\left(\alpha,\beta,\frac{1+c}{1-c}\right)=1 \tag{A.37}$$

$$\begin{aligned} f_1\left(\alpha,\beta,\frac{1+c}{1-c}\right) &= \frac{\left(\sqrt{c}-1\right)\left(4a^2+\left(1-4b^2\right)\sqrt{c}-1\right)}{8\sqrt{c}} \\ f_2\left(\alpha,\beta,\frac{1+c}{1-c}\right) &= \frac{c-1}{128\left(c^{3/2}+c\right)} \\ &\quad\times\left(-16a^4-\left(4b^2-1\right)c\left(8a^2+4b^2+5\right)\right. \\ &\quad+\left(4a^2-1\right)\sqrt{c}\left(4a^2+8b^2+5\right) \\ &\quad\left.+40a^2+\left(16b^4-40b^2+9\right)c^{3/2}-9\right), \end{aligned} \tag{A.38}$$

and also

$$\frac{Q^{(1,1-2J_{\min})}_{J+J_{\min}-1}(\frac{1+c}{1-c})}{Q^{(0,-2J_{\min})}_{J+J_{\min}}(\frac{1+c}{1-c})}=\frac{(1-c)(J+J_{\min})}{\sqrt{c}(J-J_{\min}+1)}, \tag{A.39}$$

Moreover, introducing the following binomial distribution

$$\begin{aligned} q(J_{\max},J,c) &:= \mathrm{Bin}(2J_{\max},\frac{1-\sqrt{c}}{2},J_{\max}-J) \\ &= \binom{2J_{\max}}{J_{\max}-J}\left(\frac{1-\sqrt{c}}{2}\right)^{J_{\max}-J}\left(\frac{1+\sqrt{c}}{2}\right)^{J_{\max}+J}, \end{aligned} \tag{A.40}$$

in the variable $J_{\max}-J$, we have that

$$\frac{Q^{(0,-2J_{\min})}_{J+J_{\min}}\left(\frac{1+c}{1-c}\right)}{q(J_{\max},J,c)}=\frac{(1-c)^{-M}\left(\frac{1+\sqrt{c}}{2}\right)2^{2J_{\max}-2J}(2J)!(J_{\max}-J)!(J+J_{\max})!}{\sqrt[4]{c}(2J_{\max})!(J-J_{\min})!(J+J_{\min})!}. \tag{A.41}$$

we thus find the following asymptotic expansion,

$$\frac{1}{P_{M,N}(J|c)}\left(\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}\right)^2 = \frac{(2J+1)M!N!(1-c)^{M-4}}{(J_{\max}-J)!\,(J+J_{\max}+1)!}$$
$$\times \frac{(1-c)^{-M}\left(\frac{1+\sqrt{c}}{2}\right)2^{2J_{\max}-2J}(2J)!(J_{\max}-J)!(J+J_{\max})!}{\sqrt[4]{c}(2J_{\max})!(J-J_{\min})!(J+J_{\min})!}$$
$$\times \left(\begin{array}{l} (J-J_{\min}+1)^2\left(\frac{(1-c)(J+J_{\min})}{\sqrt{c}(J-J_{\min}+1)}\right)^2 \frac{\left(\sum_{s=0}^{\infty}\frac{f_s(1,1-2J_{\min},\frac{1+c}{1-c})}{(2J+1)^s}\right)^2}{\left(\sum_{s=0}^{\infty}\frac{f_s(0,0-2J_{\min},\frac{1+c}{1-c})}{(2J+1)^s}\right)} \\ -2(J-J_{\min}+1)(1-c)M\left(\frac{(1-c)(J+J_{\min})}{\sqrt{c}(J-J_{\min}+1)}\right)\left(\sum_{s=0}^{\infty}\frac{f_s(1,1-2J_{\min},\frac{1+c}{1-c})}{(2J+1)^s}\right) \\ +(1-c)^2M^2\left(\sum_{s=0}^{\infty}\frac{f_s(0,0-2J_{\min},\frac{1+c}{1-c})}{(2J+1)^s}\right) \end{array}\right), \tag{A.42}$$

which can be recast as

$$\frac{1}{P_{M,N}(J|c)}\left(\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}\right)^2 = q(J_{\max},J,c)$$
$$\times \frac{\left(\frac{1+\sqrt{c}}{2}\right)M!N!(1-c)^{-2}2^{-2J+2J_{\max}}(2J+1)}{\sqrt[4]{c}(2J_{\max})!(J+J_{\max}+1)}\binom{2J}{J}\frac{(J-1)...(J-J_{\min})}{(J+1)....(J+J_{\min})}$$
$$\times\left(g_0(J,J_{\min},c)+\frac{g_1(J,J_{\min},c)}{2J+1}+\frac{g_2(J,J_{\min},c)}{(2J+1)^2}+O\left(\frac{1}{(2J+1)^3}\right)\right), \tag{A.43}$$

where

$$g_0(J,J_{\min},c) = \frac{(J+J_{\min}-\sqrt{c}M)^2}{c} = \frac{(J-\sqrt{c}J_{\max})^2}{c}$$
$$+2\frac{(J-\sqrt{c}J_{\max})(1-\sqrt{c})J_{\min}}{c}+(1-\sqrt{c})^2\frac{J_{\min}^2}{c}, \tag{A.44}$$
$$g_1(J,J_{\min},c) = \frac{(1-\sqrt{c})\,(J_{\min}+J-\sqrt{c}J_{\max})}{8\sqrt{c}}$$
$$\times\left(\sqrt{c}(-8J_{\max}+J_{\min}(16J_{\min}(J_{\min}+J-\sqrt{c}J_{\max}-2)+32s+15)+7(J-\sqrt{c}J_{\max}))\right.$$
$$\left.+8c(4J_{\min}-1)(J_{\min}-J_{\max})-7J_{\min}-7(J-\sqrt{c}J_{\max})\right). \tag{A.45}$$

We use the asymptotic expansion [Ele14] for the central binomial coefficients, for which we need the first three terms

$$\binom{2J}{J} = \frac{4^J}{\sqrt{\pi(J+1/2)}}\left(1+\frac{1}{4(2J+1)}+\frac{1}{32(2J+1)^2}+O\left(\frac{1}{(2J+1)^3}\right)\right) \tag{A.46}$$

Using the techniques of Lemma A.3.1, we obtain that for some $C > 0$

$$\sum_{J_{\min}}^{J_{\max}} q(J_{\max}, J, c)\frac{1}{(2J+1)^{k+1/2}}\frac{(J-1)...(J-J_{\min})}{(J+1)....(J+J_{\min})}\frac{1}{J+J_{\max}+1}(J-\sqrt{c}M)^2$$
$$\leq \frac{C}{J_{\max}^{k+1/2}}, \tag{A.47}$$
$$\sum_{J_{\min}}^{J_{\max}} q(J_{\max}, J, c)\frac{1}{(2J+1)^{k+1/2}}\frac{(J-1)...(J-J_{\min})}{(J+1)....(J+J_{\min})}\frac{1}{J+J_{\max}+1}(J-\sqrt{c}M)$$
$$\leq \frac{C}{J_{\max}^{k+3/2}}, \tag{A.48}$$
$$\sum_{J_{\min}}^{J_{\max}} q(J_{\max}, J, c)\frac{1}{(2J+1)^{k+1/2}}\frac{(J-1)...(J-J_{\min})}{(J+1)....(J+J_{\min})}\frac{1}{J+J_{\max}+1}$$
$$\leq \frac{C}{J_{\max}^{k+3/2}}. \tag{A.49}$$

With the aid of these estimates and the explicit forms of $g_0(J, J_{\min}, c)$ and $g_1(J, J_{\min}, c)$, we can see that there is only a leading contribution from $g_0(J, J_{\min}, c)$, while the remaining terms are subleading by power counting

$$\sum_{J_{\min}}^{J_{\max}} q(J_{\max}, J, c)\binom{2J}{J}2^{2J}\frac{(J-1)...(J-J_{\min})}{(J+1)....(J+J_{\min})}\frac{1}{J+J_{\max}+1}$$
$$\times\left(g_0(J, J_{\min}, c)+\frac{g_1(J, J_{\min}, c)}{2J+1}+\frac{g_2(J, J_{\min}, c)}{(2J+1)^2}+O\left(\frac{1}{(2J+1)^3}\right)\right)$$
$$\sum_{J_{\min}}^{J_{\max}} q(J_{\max}, J, c)\frac{(2J+1)^{1/2}}{\sqrt{\pi/2}}\frac{(J-1)...(J-J_{\min})}{(J+1)....(J+J_{\min})}\frac{1}{J+J_{\max}+1}\frac{(J-\sqrt{c}J_{\max})^2}{c}$$
$$+O\left(\frac{1}{J_{\max}^{1/2}}\right) = \frac{J_{\max}^{1/2}(1-\sqrt{c})}{\sqrt{\pi}c^{3/4}}+O\left(\frac{1}{J_{\max}^{1/2}}\right). \tag{A.50}$$

Putting all together, we find

$$\begin{aligned}
&\frac{1}{P_{M,N}(J|c)}\left(\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}\right)^2 = q(J_{\max},J,c)\\
&\times\frac{\left(\frac{1+\sqrt{c}}{2}\right)M!N!(1-c)^{-2}2^{-2J+2J_{\max}}(2J+1)}{\sqrt[4]{c}(2J_{\max})!(J+J_{\max}+1)}\binom{2J}{J}\frac{(J-1)...(J-J_{\min})}{(J+1)....(J+J_{\min})}\\
&\times\left(g_0(J,J_{\min},c)+\frac{g_1(J,J_{\min},c)}{2J+1}+\frac{g_2(J,J_{\min},c)}{(2J+1)^2}+O\left(\frac{1}{(2J+1)^3}\right)\right)\\
&=\frac{J_{\max}}{2c(1-c)}+O\left(1\right). \qquad \text{(A.51)}
\end{aligned}$$

The next to leading terms can be obtained in the same way. For example, the term of order $O(1)$ can be found with the terms up to order $O\left(\frac{1}{(2J+1)^2}\right)$. We do it for the case $J_{\min}=0$, where the computations are still tractable by Mathematica, obtaining

$$\frac{1}{P_{N,N}(J|c)}\left(\frac{\mathrm{d}P_{N,N}(J|c)}{\mathrm{d}c}\right)^2=\frac{N}{2c(1-c)}-\frac{1}{8c^2}+O\left(\frac{1}{N^2}\right).$$

Finally, we do not make evaluations of the remainder terms. However, since we use asymptotic expansions which are valid for large $2J+1$, by definition we can bound the remainders at order $k-1$ with a term $C_k\left(\frac{1}{(2J+1)^k}\right)$ if J is larger than some J_k. Since the lower extreme of the region where $q(J_{\max},J,c)$ concentrates grow linearly to $J_{\max}$, we can always find $J_{\max}$ such that we can control the sum of the remainder terms with a term which is $O\left(\frac{1}{J_{\max}^k}\right)$. On the other hand, outside the region of concentration we have

$$\begin{aligned}
\frac{P_{M,N}(J|c)}{q(J_{\max},J,c)}&=\frac{(2j+1)2^{2J_{\max}}(1-c)^{J+J_{\min}}(1+\sqrt{c})^{-2J}M!N!P_{J+J_{\min}}^{(0,-2J_{\min})}\left(\frac{1+c}{1-c}\right)}{(J+J_{\max}+1)(M+N)!}\\
&=\frac{(2J+1)2^{2J_{\max}}M!N!(1+\sqrt{c})^{-2J}\sum_{s=0}^{J-J_{\min}}\binom{J+J_{\min}}{s}\binom{J-J_{\min}}{s}c^s}{(J+J_{\max}+1)(M+N)!}, \qquad \text{(A.52)}
\end{aligned}$$

and

$$
\begin{aligned}
&(1+\sqrt{c})^{-2J}\sum_{s=0}^{J-J_{\min}}\binom{J+J_{\min}}{s}\binom{J-J_{\min}}{s}c^s = \\
&\sum_{s=0}^{J-J_{\min}}\binom{J+J_{\min}}{s}\left(\frac{\sqrt{c}}{1+\sqrt{c}}\right)^s\left(\frac{1}{1+\sqrt{c}}\right)^{J+J_{\min}-s} \\
&\times\binom{J-J_{\min}}{s}\left(\frac{\sqrt{c}}{1+\sqrt{c}}\right)^s\left(\frac{1}{1+\sqrt{c}}\right)^{J-J_{\min}-s} \\
&\leq \sum_{s=0}^{J-J_{\min}} 1 = J-J_{\min}. \qquad \text{(A.53)}
\end{aligned}
$$

By bounding also the remaining terms we get that $\frac{P_{M,N}(J|c)}{q(J_{\max},J,c)}$ is bounded by a power law in J, meaning that if $q(J_{\max},J,c)$ concentrates exponentially also $P_{M,N}(J|c)$ concentrates exponentially in the same region. We already knew that this concentration should hold by Theorem 4.2.7. We are left to bound

$$
\begin{aligned}
&\left(\frac{\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}}{P_{M,N}(J|c)}\right) = -\frac{(J_{\max}-J)}{1-c}+\frac{(1-c)^{J_{\max}-J}\sum_{s=0}^{J-J_{\min}} s\binom{J+J_{\min}}{s}\binom{J-J_{\min}}{s}c^s}{cP_{M,N}(J|c)} \\
&\leq -\frac{(J_{\max}-J)}{1-c}+\frac{J-J_{\min}}{c}, \qquad \text{(A.54)}
\end{aligned}
$$

it follows that for $J \leq J_0$,

$$
\left|\left(\frac{\frac{\mathrm{d}P_{M,N}(J|c)}{\mathrm{d}c}}{P_{M,N}(J|c)}\right)\right| \leq \frac{cJ_{\max}+J_0+(1-c)J_{\min}}{c(1-c)}. \qquad \text{(A.55)}
$$

therefore, since the binomial distribution has exponential tails, the region where we cannot apply the asymptotic expansion does not contribute to our asymptotic series for the Fisher information.

A.5 Overlap estimation with depolarizing noise

In this appendix we derive the optimal estimator and corresponding mean squared error for the case where we are given N and M copies of depolarized qubits. We shall restrict our attention to qubit mixed states and for ease of notation we use a purely $SU(2)$ description of the average states, since the multiplicities are not relevant.

The mixed states whose overlap we wish to estimate are

$$\rho_\psi^{\otimes N}(r_0) = \left(r_0 \left|\psi\right\rangle \left\langle\psi\right| + (1-r_0)\frac{I}{2} \right)^{\otimes N},$$
$$\rho_\phi^{\otimes M}(r_1) = \left(r_1 \left|\phi\right\rangle \left\langle\phi\right| + (1-r_1)\frac{I}{2} \right)^{\otimes M}, \tag{A.56}$$

where $r_{0(1)}$ denotes the corresponding purity of the states. From Eq. A.13 these states can be written in the total angular momentum basis, after tracing out multiplicities, as

$$\tilde{\rho}_\psi^{\otimes N} = \sum_{J_0=0}^{\frac{M}{2}} p_{J_0}\, \tau_{J_0}^{(0)}(\vec{n}_0)$$
$$\tilde{\rho}_\phi^{\otimes M} = \sum_{J_1=0}^{\frac{N}{2}} p_{J_1}\, \tau_{J_1}^{(1)}(\vec{n}_1), \tag{A.57}$$

where

$$\tau_{J_0}^{(0)} = \frac{1}{Z_{J_0}^{(0)}} \sum_{k=-J_0}^{J_0} R_0^k \left|J_0, k\right\rangle\!\left\langle J_0, k\right|$$
$$\tau_{J_1}^{(1)} = \frac{1}{Z_{J_1}^{(1)}} \sum_{l=-J_1}^{J_1} R_1^l \sum_{\alpha,\beta=-J_1}^{J_1} D_{\alpha,l}^{(J_1)}(h) D_{l,\beta}^{(J_1)}(h) \left|J_1, \alpha\right\rangle\!\left\langle J_1, \beta\right| \tag{A.58}$$

with $R_i = \frac{1+r_i}{1-r_i}$, $Z_{J_i}^{(i)} = \frac{R_i^{J_i+1} - R_i^{-J_i}}{R_i - 1}$, and just as for the case of pure states, h is such that $h\left|\psi\right\rangle = \left|\phi\right\rangle$. We have chosen $\vec{n}_0 = \vec{z}$ without loss of generality. Moreover,

$$p_{J_0} = \left(\frac{1-r^2}{4}\right)^{\frac{N}{2}} \begin{pmatrix} N \\ \frac{N}{2} - J_0 \end{pmatrix} \frac{2J_0+1}{\frac{N}{2}+J_0+1} Z_{J_0} \tag{A.59}$$

and similarly for p_{J_1}. Integrating over SU(2) we obtain

$$\rho(c, r_0, r_1) := \int U^{\otimes N+M} \rho_\psi^{\otimes N} \otimes \rho_\phi^{\otimes M} U^{\dagger \otimes N+M}$$
$$= \sum_J \sum_{J_0,J_1} p_{J_0} p_{J_1} \sum_{k,l} \frac{R_0^k R_1^l}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} \sum_{\alpha=-J_1}^{J_1} \left(C_{J_0,k;J_1,\alpha}^{J,k+\alpha} D_{\alpha,l}^{(J_1)}(h) \right)^2 \frac{I_{\lambda_J}}{2J+1} \otimes \sigma^{(J_0,J_1)}, \tag{A.60}$$

where $\sigma^{(J_0,J_1)} \in \Sigma(\mathcal{V}_{\lambda_J}(S_{N+M}))$ and they are orthogonal for different pairs (J_0, J_1). To calculate the AvMSE we need to compute the operators Γ, η of Eq. (2.20). A similar

calculation as in Eq. (6.41), (6.43) gives

$$\Gamma = \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \frac{p_{J_0} p_{J_1}}{(2J_0+1)(2J_1+1)} \sum_{J=|J_0-J_1|}^{J_0+J_1} I_{\lambda_J} \otimes \sigma^{(J_0,J_1)},$$

For η one obtains the following expression:

$$\eta = \int_{\mathrm{SU}(2)} U^{\otimes(N+M)} \left(\sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \sum_{k=-J_0}^{J_0} \sum_{l=-J_1}^{J_1} \frac{R_0^k R_1^l}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} \, |J_0,k\rangle\langle J_0,k| \right.$$
$$\left. \otimes \int_{\mathrm{SU}(2)} \mathrm{d}h \, |D^{(\frac{1}{2})}_{\frac{1}{2}\frac{1}{2}}(h)|^2 D^{(J_1)}(h) \, |J_1,l\rangle\langle J_1,l| \, D^{(J_1)}(h)^\dagger \right) U^{\dagger\otimes(N+M)}. \tag{A.61}$$

We finally obtain

$$\eta = \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \frac{p_{J_0} p_{J_1}}{(2J_0+1)(2J_1+1)} \sum_{J=|J_0-J_1|}^{J_0+J_1} \left(1 - \frac{(2J_0+1)(2J_1+1)}{p_{J_0} p_{J_1}} \sum_{k=-J_0}^{J_0} \sum_{l=-J_1}^{J_1} \frac{R_0^k R_1^l}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} \right.$$
$$\left. \times \sum_{L=|J_1-\frac{1}{2}|}^{J_1+\frac{1}{2}} \sum_{h=-J_1}^{J_1} \frac{\left(C^{L,-\frac{1}{2}+h}_{\frac{1}{2},-\frac{1}{2};J_1,h} C^{L,\frac{1}{2}+l}_{\frac{1}{2},\frac{1}{2};J_1,l} C^{J,k+h}_{J_0,k;J_1,h} \right)^2}{(2L+1)} \right) \frac{I_{\lambda_J}}{(2J+1)} \otimes \sigma^{(J_0,J_1)}$$
$$= \sum_{J_0=0}^{\frac{N}{2}} \sum_{J_1=0}^{\frac{M}{2}} \frac{p_{J_0} p_{J_1}}{(2J_0+1)(2J_1+1)} \sum_{J=|J_0-J_1|}^{J_0+J_1} \left(1 - \frac{(2J_0+1)(2J_1+1)}{p_{J_0} p_{J_1}} \sum_{L=|J_1-\frac{1}{2}|}^{J_1+\frac{1}{2}} \sum_{L'=|J_0-\frac{1}{2}|}^{J_0+\frac{1}{2}} \right.$$
$$\left. \times \sum_{k=-J_0}^{J_0} \sum_{l=-J_1}^{J_1} \frac{R_0^k \left(C^{L',\frac{1}{2}+k}_{\frac{1}{2},\frac{1}{2},J_0,k} \right)^2 R_1^l \left(C^{L,\frac{1}{2}+l}_{\frac{1}{2},\frac{1}{2},J_1,l} \right)^2}{Z_{J_0}^{(0)} Z_{J_1}^{(1)}} \begin{Bmatrix} J_1 & \frac{1}{2} & L \\ L' & J & J_0 \end{Bmatrix}^2 \right) I_{\lambda_J} \otimes \sigma^{(J_0,J_1)}. \tag{A.62}$$

For a given J, J_0, J_1 and overlap c the estimator is given by

$$\tilde{c}(J,J_0,J_1) = \frac{\mathrm{tr}[\Pi_J \, (\Pi_{J_0} \otimes \Pi_{J_1} \eta)]}{\mathrm{tr}[\Pi_J \, (\Pi_{J_0} \otimes \Pi_{J_1} \Gamma)]} \tag{A.63}$$

and the AvMSE reads

$$v_{op,mix} = \int_0^1 p(c) c^2 - \sum_{J,J_0,J_1} p(J,J_0,J_1) c(J,J_0,J_1)^2. \tag{A.64}$$

The sums in L, L', m, k can be done exactly, the result being polynomials in J. The sums in J can be done exactly too. The final sum in J_0 and J_1 can be done using the fact that $p_{J_0} \frac{R_{(0)}^{J_0}}{Z_{J_0}}$ can be written as

$$p_{J_0}\frac{R_{(0)}^{J_0}}{Z_{J_0}} = \frac{2J_0+1}{\frac{N}{2}+J_0+1}\mathrm{Bin}(N, \frac{N}{2}-J_0, \frac{1+r}{2}), \tag{A.65}$$

and in the limit $M = \alpha Z$, $N = \beta Z$, $Z \to \infty$ one can approximate the AvMSE expanding in moments around the mean of the binomial distribution, applying Lemma A.3.1. The final result reads

$$v_{op,mix} = \frac{1}{6Mr_0^2} + \frac{1}{6Nr_1^2} + o(Z^{-1}) \tag{A.66}$$

in agreement with the pure state case for $d = 2$, $r_0 = r_1 = 1$.

A.6 Equivalence of sampling model and Poissonized model

The equivalence of the Poisson model with the original one can be formalised in the following proposition.

Proposition A.6.1. *Suppose that given access to M copies of the state ρ of Eq. (7.1), where M is extracted from a Poisson distribution with mean μ, there is a test* `Ptest` *such that*

$$\begin{cases} P(\texttt{Ptest} \mapsto \texttt{"accept"} \,|\mathit{Case}\ \mathsf{A}) > 3/4\,, \\ \\ P(\texttt{Ptest} \mapsto \texttt{"accept"} \,|\mathit{Case}\ \mathsf{B}) < 1/4\,, \end{cases} \tag{A.67}$$

and it can be performed by a two-outcome POVM $\{E_0^{(M)}, E_1^{(M)}\}$ for each M. Then, provided that μ is larger than a fixed constant, there is a test in the sampling model using 2μ copies of ρ satisfying

$$\begin{cases} P(\texttt{test} \mapsto \texttt{"accept"} \,|\mathit{Case}\ \mathsf{A}) > 2/3\,, \\ \\ P(\texttt{test} \mapsto \texttt{"accept"} \,|\mathit{Case}\ \mathsf{B}) < 1/3\,. \end{cases} \tag{A.68}$$

Proof. Given 2μ copies of ρ, we construct the following test. We extract M from a Poisson distribution with mean μ. If $M < 2\mu$, we perform the measurement $\{E_0^{(M)}, E_1^{(M)}\}$, otherwise we declare failure. The difference of the acceptance probabilities of `test` and

`Ptest` is

$$\begin{aligned} & P(\texttt{Ptest} \mapsto \texttt{"accept"}) - P(\texttt{test} \mapsto \texttt{"accept"}) \\ & = \sum_{M=0}^{2\mu} \mathrm{Poi}_\mu(M) \left(\mathrm{Tr}\Big[E_0^{(M)} \rho^{\otimes M} \Big] - \mathrm{Tr}\Big[E_0^{(M)} \rho^{\otimes M} \Big] \right) \\ & + \sum_{M=2\mu+1}^{\infty} \mathrm{Poi}_\mu(M) \left(\mathrm{Tr}\Big[E_0^{(M)} \rho^{\otimes M} \Big] - 0 \right) \\ & = \sum_{M=2\mu+1}^{\infty} \mathrm{Poi}_\mu(M) \, \mathrm{Tr}\Big[E_0^{(M)} \rho^{\otimes M} \Big], \end{aligned} \tag{A.69}$$

which implies

$$\begin{aligned} & 0 \leq P(\texttt{Ptest} \mapsto \texttt{"accept"}) - P(\texttt{test} \mapsto \texttt{"accept"}) \leq \sum_{M=2\mu+1}^{\infty} \mathrm{Poi}_\mu(M) \\ & = P_{M \sim \mathrm{Poi}_\mu}(M > 2\mu). \end{aligned} \tag{A.70}$$

Invoking hence the Cramér-Chernoff tail bound on the Poisson distribution [BLM13], i.e.

$$P_{M \sim \mathrm{Poi}_\mu}(M > t) \leq e^{-t h(t/\mu)} \qquad h(x) = (1+x)\log(1+x) - x \, , \tag{A.71}$$

and setting $\mu > 1$, from Eq. (A.70) we then get

$$0 \leq P(\texttt{Ptest} \mapsto \texttt{"accept"}) - P(\texttt{test} \mapsto \texttt{"accept"}) \leq e^{-\mu h(2)} < 1/10 \, , \tag{A.72}$$

from which the statement of the proposition follows. □

A.7 Proof of Proposition 7.3.2

As in the proof of Proposition 7.3.1 we can invoke Eqs. (7.25), (7.24) and the identity $\sum_{x \in \Gamma_M} x^2 P_x^{(M)} = \mathrm{Tr}\big[\mathcal{D}_M^2 \rho^{(M)} \big]$ to write

$$\begin{aligned} \mathrm{Var}[\mathcal{D}] &= \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \, \mathrm{Tr}\Big[\mathcal{D}_M^2 \rho^{(M)} \Big] - \mathbb{E}[\mathcal{D}]^2 \\ &= \sum_{M=0}^{\infty} \mathrm{Poi}_\mu(M) \sum_{\vec{m} \in \mathcal{P}_M} \mathsf{M}(\vec{m})_{\vec{p},M} \, \mathrm{Tr}\Big[(D^{\vec{m},M})^2 \rho^{\vec{m}} \Big] - \mathbb{E}[\mathcal{D}]^2 \, , \end{aligned} \tag{A.73}$$

where the last passage involves (7.18) and (7.15). Replacing Eqs. (7.15), (7.19), and (7.20) into $\mathrm{Tr}\big[(D^{\vec{m},M})^2 \rho^{\vec{m}} \big]$ reveals that such term can be written as a linear combination of the expectation values of the operators $\mathcal{O}_{ij}^{m_i,m_j} \mathcal{O}_{kl}^{m_k,m_l}$ on $\rho^{\vec{m}}$ which are complicated functions of of the random variable m_i and traces of powers of the ρ_i reported in the next

subsection. Invoking hence (7.28) to decouple the averages over the m_i we can finally write

$$\mathrm{Var}[\mathcal{D}] = V_1 + V_2 \,, \tag{A.74}$$

where setting $\mathrm{Var}_\rho[O] := \mathrm{Tr}\big[(O - \mathrm{Tr}[O\rho])^2\rho\big]$, we defined

$$V_1 = \mathop{\mathbb{E}}_{\substack{m_l\sim\mathrm{Poi}(p_l\mu)\\ l=1,\dots,N}} \mathrm{Var}_{\rho^{\vec{m}}}\left[D^{\vec{m},M}\right], \tag{A.75}$$

$$V_2 = \mathop{\mathbb{E}}_{\substack{m_l\sim\mathrm{Poi}(p_l\mu)\\ l=1,\dots,N}} \left(\mathrm{Tr}\left[\mathcal{D}^{\vec{m},M}\rho^{\vec{m}}\right] - \sum_{i,j} p_i p_j D^2_{HS}(\rho_i,\rho_j)\right)^2 , \tag{A.76}$$

(we remind that the expression $m_l \sim \mathrm{Poi}(p_l\mu)$ indicates that the random variables m_l are extracted from a Poisson distribution of mean $p_l\mu$).

A.7.1 Bound on V_1

The covariance of two observables O, O'on a state ρ is defined as

$$\mathrm{Cov}_\rho[O,O'] := \mathrm{Tr}\big[(O - \mathrm{Tr}[O\rho])(O' - \mathrm{Tr}\big[O'\rho\big])\big]. \tag{A.77}$$

The covariances of the observables $\mathcal{O}_{ij}^{m_i,m_j}$ on $\rho^{\vec{m}}$, read:

$$\mathrm{Var}_{\rho^{\vec{m}}}[\mathcal{O}_{(ii)}^{m_i,m_i}] = \frac{2}{m_i(m_i-1)}(1-(\mathrm{Tr}\big[\rho_i^2\big])^2) + \frac{4(m_i-2)}{m_i(m_i-1)}(\mathrm{Tr}\big[\rho_i^3\big] - (\mathrm{Tr}\big[\rho_i^2\big])^2), \tag{A.78}$$

$$\begin{aligned}\mathrm{Var}_{\rho^{\vec{m}}}[\mathcal{O}_{(ij)}^{m_i,m_j}] = {}& \frac{1}{m_i m_j} + \frac{1-m_i-m_j}{m_i m_j}\,\mathrm{Tr}[\rho_i\rho_j]^2,\\ &+ \frac{1}{m_i}\left(1-\frac{1}{m_j}\right)\mathrm{Tr}\big[\rho_i^2\rho_j\big] + \frac{1}{m_j}\left(1-\frac{1}{m_i}\right)\mathrm{Tr}\big[\rho_i\rho_j^2\big] \qquad i\neq j\end{aligned} \tag{A.79}$$

$$\mathrm{Cov}_{\rho^{\vec{m}}}[\mathcal{O}_{(ii)},\mathcal{O}_{(ij)}] = \frac{2}{m_i}\left(\mathrm{Tr}\big[\rho_i^2\rho_j\big] - \mathrm{Tr}\big[\rho_i^2\big]\,\mathrm{Tr}[\rho_i\rho_j]\right) \qquad i\neq j, \tag{A.80}$$

$$\mathrm{Cov}_{\rho^{\vec{m}}}[\mathcal{O}_{(ij)},\mathcal{O}_{(jk)}] = \frac{\mathrm{Tr}[\rho_i\rho_j\rho_k] - \mathrm{Tr}[\rho_i\rho_j]\,\mathrm{Tr}[\rho_i\rho_k]}{m_i} \qquad i\neq k, \tag{A.81}$$

$$\mathrm{Cov}_{\rho^{\vec{m}}}[\mathcal{O}_{(ij)},\mathcal{O}_{(kl)}] = 0 \qquad i\neq k \wedge j\neq l. \tag{A.82}$$

Replacing the above expressions into Eq. (A.75), we can rewrite it as

$$
\begin{aligned}
V_1 = & \mathop{\mathbb{E}}_{\substack{m_l\sim\mathrm{Poi}(p_l\mu)\\ l=1\ldots N}} \mathrm{Var}_{\rho^{\vec{m}}}\left[\sum_{i\neq j}\frac{m_i(m_i-1)}{\mu^2 p_i}\mathcal{O}_{ii}^{m_i}+\frac{m_j(m_j-1)}{\mu^2 p_j}p_i\mathcal{O}_{jj}^{m_j}-2\frac{m_i m_j}{\mu^2}\mathcal{O}_{ij}^{m_i,m_j}\right] \\
= & \sum_i 4 \mathop{\mathbb{E}}_{m_i\sim\mathrm{Poi}(p_i\mu)} \frac{m_i^2(m_i-1)^2}{\mu^4 p_i^2}(1-p_i)^2\mathrm{Var}[\mathcal{O}_{(ii)}^2] \\
& + 8\sum_{i\neq j} \mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)}} \frac{m_i^2 m_j^2}{\mu^4}\mathrm{Var}[\mathcal{O}_{(ij)}] \\
& - 16\sum_{i\neq j} \mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)}} \frac{m_i^2(m_i-1)m_j}{\mu^4 p_i}(1-p_i)\mathrm{Cov}[\mathcal{O}_{(ii)},\mathcal{O}_{(ij)}] \\
& + 8\sum_{i\neq j\neq k} \mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)\\ m_k\sim\mathrm{Poi}(p_k\mu)}} \frac{m_i^2 m_j m_k}{\mu^4 p_i}\mathrm{Cov}[\mathcal{O}_{(ij)},\mathcal{O}_{(ik)}].
\end{aligned}
\tag{A.83}
$$

Now we proceed to evaluate separately each term of Eq. (A.83).

From (A.78) we get

$$
\begin{aligned}
& \mathop{\mathbb{E}}_{m_i\sim\mathrm{Poi}(p_i\mu)}\left[\frac{m_i^2(m_i-1)^2}{\mu^4 p_i^2}(1-p_i)^2\mathrm{Var}[\mathcal{O}_{(ii)}^2]\right] \\
& = \mathop{\mathbb{E}}_{m_i\sim\mathrm{Poi}(p_i\mu)}\left[\frac{m_i(m_i-1)}{\mu^4 p_i^2}(1-p_i)^2[2(1-(\mathrm{Tr}\left[\rho_i^2\right])^2)+4(m_i-2)(\mathrm{Tr}\left[\rho_i^3\right]-(\mathrm{Tr}\left[\rho_i^2\right])^2]\right] \\
& = \frac{\mu^2 p_i^2}{\mu^4 p_i^2}(1-p_i)^2[2(1-(\mathrm{Tr}\left[\rho_i^2\right])^2)+4\mu_i p_i(\mathrm{Tr}\left[\rho_i^3\right]-(\mathrm{Tr}\left[\rho_i^2\right])^2] \\
& \leq \frac{4p_i(1-p_i)^2}{\mu}(\mathrm{Tr}\left[\rho_i^3\right]-(\mathrm{Tr}\left[\rho_i^2\right])^2)+O(N/\mu^2).
\end{aligned}
\tag{A.84}
$$

where in the third line we used the fact that $\mathbb{E}[m_i(m_i-1)] = \mu_i^2 p_i^2$ and $\mathbb{E}[m_i(m_i-1)(m_i-2)] = \mu_i^3 p_i^3$ for a Poisson distribution with mean $\mu_i p_i$.

Analougously, from (A.79) we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)}} \frac{m_i^2 m_j^2}{\mu^4}\mathrm{Var}[\mathcal{O}_{(ij)}] \\
&= \mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)}} \frac{m_i m_j}{\mu^4}(1+(1-m_i-m_j)\,\mathrm{Tr}[\rho_i\rho_j]^2 \\
&+ (m_i-1)\,\mathrm{Tr}\left[\rho_i^2\rho_j\right] + (m_j-1)\,\mathrm{Tr}\left[\rho_i\rho_j^2\right]) \\
&\leq \frac{p_i p_j^2\,\mathrm{Tr}\left[\rho_i\rho_j^2\right] + p_j p_i^2\,\mathrm{Tr}\left[\rho_j\rho_i^2\right] - p_i p_j (p_i+p_j)\,\mathrm{Tr}[\rho_i\rho_j]^2}{\mu} + O(1/\mu^2),
\end{aligned} \tag{A.85}$$

where in the leading $1/\mu$ term we kept only $\mathbb{E}[m_i^\alpha] = \mu^\alpha p_i^\alpha + O(\mu^{\alpha-1}p^{\alpha-1})$.

The corresponding contribution from Eq. (A.80) is

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)}} \frac{m_i^2(m_i-1)m_j}{\mu^4 p_i}(1-p_i)\mathrm{Cov}[\mathcal{O}_{(ii)},\mathcal{O}_{(ij)}] \\
&= \mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)}} \frac{m_i(m_i-1)m_j}{\mu^4 p_i}(1-p_i)2\left(\mathrm{Tr}\left[\rho_i^2\rho_j\right] - \mathrm{Tr}\left[\rho_i^2\right]\mathrm{Tr}[\rho_i\rho_j]\right) \\
&= \frac{(1-p_i)p_i p_j}{\mu}2\left(\mathrm{Tr}\left[\rho_i^2\rho_j\right] - \mathrm{Tr}\left[\rho_i^2\right]\mathrm{Tr}[\rho_i\rho_j]\right).
\end{aligned} \tag{A.86}$$

Finally, from (A.81) we have

$$\begin{aligned}
&\mathbb{E}\frac{m_i^2 m_j m_k}{M^4 p_i}\mathrm{Cov}[\mathcal{O}_{(ij)},\mathcal{O}_{(ik)}] \\
&= \mathop{\mathbb{E}}_{\substack{m_i\sim\mathrm{Poi}(p_i\mu)\\ m_j\sim\mathrm{Poi}(p_j\mu)\\ m_k\sim\mathrm{Poi}(p_k M)}} \frac{m_i m_j m_k}{\mu^4 p_i}\left(\mathrm{Tr}[\rho_i\rho_j\rho_k] - \mathrm{Tr}[\rho_i\rho_j]\,\mathrm{Tr}[\rho_i\rho_k]\right) \\
&= \frac{p_i p_j p_k}{\mu}(\mathrm{Tr}[\rho_i\rho_j\rho_k] - \mathrm{Tr}[\rho_i\rho_j]\,\mathrm{Tr}[\rho_i\rho_k]).
\end{aligned} \tag{A.87}$$

Inserting (A.84), (A.85) and (A.87) into (A.83) we can finally write

$$\begin{aligned} V_1 = 16 \sum_i \frac{p_i(1-p_i)^2}{\mu} (\mathrm{Tr}\left[\rho_i^3\right] - (\mathrm{Tr}\left[\rho_i^2\right])^2) \\ + 8 \sum_{i \neq j} \frac{p_i p_j^2 \,\mathrm{Tr}\left[\rho_i \rho_j^2\right] + p_j p_i^2 \,\mathrm{Tr}\left[\rho_j \rho_i^2\right] - p_i p_j (p_i + p_j) \,\mathrm{Tr}[\rho_i \rho_j]^2}{\mu} \\ - 32 \sum_{i \neq j} \frac{(1-p_i) p_i p_j}{\mu} \left(\mathrm{Tr}\left[\rho_i^2 \rho_j\right] - \mathrm{Tr}\left[\rho_i^2\right] \mathrm{Tr}[\rho_i \rho_j]\right) \\ + 8 \sum_{i \neq j \neq k} \frac{p_i p_j p_k}{\mu} (\mathrm{Tr}[\rho_i \rho_j \rho_k] - \mathrm{Tr}[\rho_i \rho_j] \,\mathrm{Tr}[\rho_i \rho_k]) + O(N/\mu^2). \end{aligned} \tag{A.88}$$

A.7.2 Bound on V_2

We start defining the quantities

$$o_{ii} = \left(\frac{m_i(m_i-1)}{\mu^2 p_i} - p_i\right) \mathrm{Tr}\left[\rho_i^2\right], \qquad o_{ij} = \left(\frac{m_i m_j}{\mu^2} - p_i p_j\right) \mathrm{Tr}[\rho_i \rho_j], \qquad i \neq j. \tag{A.89}$$

Noticing that

$$\mathrm{Tr}\left[\mathcal{D}^{\vec{m},\mu} \rho^{\vec{m}}\right] - \sum_{ij} p_i p_j D_{HS}^2(\rho_i, \rho_j) = \sum_{i \neq j} p_j o_{ii} + p_i o_{jj} - 2 o_{ij} \,, \tag{A.90}$$

we can rewrite Eq. (A.76) as

$$\begin{aligned} V_2 = \sum_i 4(1-p_i)^2 \mathop{\mathbb{E}}_{m_i \sim \mathrm{Poi}(p_i \mu)} [o_{ii}^2] + 8 \mathop{\mathbb{E}}_{\substack{m_i \sim \mathrm{Poi}(p_i \mu) \\ m_j \sim \mathrm{Poi}(p_j \mu)}} [o_{ij}^2] \\ + 8 \sum_{k \neq i \neq j} \mathop{\mathbb{E}}_{\substack{m_i \sim \mathrm{Poi}(p_i \mu) \\ m_j \sim \mathrm{Poi}(p_j \mu)}} [o_{ij} o_{ik}] - 16 \sum_{i \neq j} (1-p_i)^2 p_i \mathop{\mathbb{E}}_{\substack{m_i \sim \mathrm{Poi}(p_i \mu) \\ m_j \sim \mathrm{Poi}(p_j \mu)}} [o_{ii} o_{ij}]. \end{aligned} \tag{A.91}$$

The expected values which appear in (A.91)can be easily computed:

$$\mathop{\mathbb{E}}_{m_i \sim \mathrm{Poi}(p_i \mu)} [o_{ii}^2] = \frac{2(1 + 2\mu p_i)}{\mu^2} \,\mathrm{Tr}\left[\rho_i^2\right]^2, \tag{A.92}$$

$$\mathop{\mathbb{E}}_{\substack{m_i \sim \mathrm{Poi}(p_i \mu) \\ m_j \sim \mathrm{Poi}(p_j \mu)}} [o_{ij}^2] = \frac{(\mu p_i p_j (p_i + p_j) + p_i p_j)}{\mu^2} \,\mathrm{Tr}[\rho_i \rho_j]^2, \qquad i \neq j, \tag{A.93}$$

$$\mathop{\mathbb{E}}_{\substack{m_i \sim \mathrm{Poi}(p_i \mu) \\ m_j \sim \mathrm{Poi}(p_j \mu) \\ m_k \sim \mathrm{Poi}(p_k \mu)}} [o_{ij} o_{ik}] = \frac{p_i p_j p_k}{\mu} \,\mathrm{Tr}[\rho_i \rho_j] \,\mathrm{Tr}[\rho_i \rho_k], \qquad i \neq j, k, \tag{A.94}$$

$$\mathop{\mathbb{E}}_{\substack{m_i \sim \mathrm{Poi}(p_i \mu) \\ m_j \sim \mathrm{Poi}(p_j \mu)}} [o_{ii} o_{ij}] = \frac{2 p_i p_j}{\mu} \,\mathrm{Tr}[\rho_i \rho_j] \,\mathrm{Tr}\left[\rho_i^2\right], \,. \qquad i \neq j \tag{A.95}$$

Replacing Eqs. (A.92), (A.93), (A.94) and (A.95) into Eq. (A.91), and then isolating the leading order, we can conclude that

$$\begin{aligned}
V_2 &= \sum_i \frac{8(1+2\mu p_i)}{\mu^2}(1-p_i)^2 \operatorname{Tr}\left[\rho_i^2\right]^2 + \sum_{i\neq j} \frac{8(\mu p_i p_j (p_i+p_j) + p_i p_j)}{\mu^2} \operatorname{Tr}[\rho_i\rho_j]^2 \\
&+ \sum_{i\neq j}\sum_{k\neq j} \frac{8(p_i p_j p_k)}{\mu} \operatorname{Tr}[\rho_i\rho_j]\operatorname{Tr}[\rho_i\rho_k] - \sum_{i\neq j} \frac{32 p_i p_j}{\mu}(1-p_i)\operatorname{Tr}[\rho_i\rho_j]\operatorname{Tr}\left[\rho_i^2\right] \\
&\leq \sum_i \frac{16}{\mu}(1-p_i)^2 p_i \operatorname{Tr}\left[\rho_i^2\right]^2 + \sum_{i\neq j} \frac{8 p_i p_j (p_i+p_j)}{\mu} \operatorname{Tr}[\rho_i\rho_j]^2 \\
&+ \sum_{i\neq j}\sum_{k\neq j} \frac{8(p_i p_j p_k)}{\mu} \operatorname{Tr}[\rho_i\rho_j]\operatorname{Tr}[\rho_i\rho_k] \\
&- \sum_{i\neq j} \frac{32 p_i p_j}{\mu}(1-p_i)\operatorname{Tr}[\rho_i\rho_j]\operatorname{Tr}\left[\rho_i^2\right] + O(N/\mu^2)\,.
\end{aligned} \tag{A.96}$$

A.7.3 Bound on $V_1 + V_2$

We start by observing that

$$\begin{aligned}
0 \quad &\leq \operatorname{Tr}\left[(\rho_i\sqrt{\rho_j} - \rho_k\sqrt{\rho_j})^\dagger(\rho_i\sqrt{\rho_j} - \rho_k\sqrt{\rho_j})\right] \\
&\implies 2\operatorname{Tr}[\rho_i\rho_j\rho_k] \leq \operatorname{Tr}\left[\rho_i^2\rho_j\right] + \operatorname{Tr}\left[\rho_k^2\rho_j\right]\,.
\end{aligned} \tag{A.97}$$

Applying Eq. (A.97) to the sum and summing

$$\sum_{i\neq j\neq k} \frac{p_i p_j p_k}{\mu} \operatorname{Tr}[\rho_i\rho_j\rho_k] \leq 2\sum_{i\neq j} \frac{p_i p_j (1-p_i-p_j)\operatorname{Tr}\left[\rho_i\rho_j^2\right]}{\mu}. \tag{A.98}$$

Combining Eqs. (A.88), (A.96) and using (A.98) we have

$$
\begin{aligned}
V_1 + V_2 &= O\left(\frac{N}{\mu^2}\right) + 16\sum_i \frac{p_i(1-p_i)^2}{\mu}\operatorname{Tr}\left[\rho_i^3\right] + 8\sum_{i\neq j}\frac{p_i p_j^2 \operatorname{Tr}\left[\rho_i\rho_j^2\right] + p_j p_i^2 \operatorname{Tr}\left[\rho_j\rho_i^2\right]}{\mu} \\
&- 32\sum_{i\neq j}\frac{(1-p_i)p_i p_j}{\mu}\left(\operatorname{Tr}\left[\rho_i^2\rho_j\right]\right) + 8\sum_{i\neq j\neq k}\frac{p_i p_j p_k}{\mu}(\operatorname{Tr}[\rho_i\rho_j\rho_k]) + O(N/\mu^2) \\
&\leq O\left(\frac{N}{\mu^2}\right) + 16\left(\sum_i \frac{p_i(1-p_i)^2}{\mu}\operatorname{Tr}\left[\rho_i^3\right]\right. \\
&\left. + \sum_{i\neq j}\frac{p_i p_j[(p_j+1-p_i-p_j)\operatorname{Tr}\left[\rho_i\rho_j^2\right] - 2(1-p_i)\operatorname{Tr}\left[\rho_i^2\rho_j\right]]}{\mu}\right) \\
&= O\left(\frac{N}{\mu^2}\right) + \frac{16}{\mu}\sum_{i\neq j} p_i p_j \operatorname{Tr}\left[((1-p_i)\rho_i)(\rho_i-\rho_j)^2\right] \\
&\leq \sum_{i\neq j} p_i p_j \operatorname{Tr}\left[\|(1-p_i)\rho_i\|_\infty(\rho_i-\rho_j)^2\right] \\
&\leq O\left(\frac{N}{\mu^2}\right) + \frac{16}{\mu}\sum_{i\neq j} p_i p_j \operatorname{Tr}\left[(\rho_i-\rho_j)^2\right] \\
&= O\left(\frac{N}{\mu^2}\right) + \frac{16}{\mu}\sum_{i\neq j} p_i p_j D_{HS}^2(\rho_i,\rho_j) = O\left(\frac{N}{\mu^2}\right) + \frac{16\mathcal{M}_{HS}^2}{\mu}.
\end{aligned}
\tag{A.99}
$$

Appendix B

Appendix: miscellanea on Gaussian states and channels

In this appendix we collect computations which we needed in the main text, in Chapters 8, 9.

B.1 Degradability of flagged additive gaussian noise

In this section we show that Λ^e_β is degradable. We consider the following Stinespring representation, realized with a Gaussian unitary $\hat{S}$ acting on a space of five modes. The input state of the channel is a state with covariance matrix $\boldsymbol{\sigma}_A = \begin{pmatrix} x & z \\ z & p \end{pmatrix}$. The joint input state to $\hat{S}$ has a covariance matrix

$$\boldsymbol{\sigma}_{AE} := \begin{pmatrix} x & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ z & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{\beta} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\beta}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{\beta} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\beta}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{\beta} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\beta}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{\beta} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\beta}{2} \end{pmatrix}. \tag{B.1}$$

$\hat{S}$ acts as the symplectic matrix

$$S = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 & 1 \end{pmatrix}. \tag{B.2}$$

It is verified that this dilation produces the correct channel output covariance matrix

$$\boldsymbol{\sigma}_{AF} = \begin{pmatrix} \frac{2}{\beta} + x & z & 0 & 0 & 0 & -\frac{1}{\beta} \\ z & \frac{2}{\beta} + p & 0 & \frac{1}{\beta} & 0 & 0 \\ 0 & 0 & \frac{2}{\beta} & 0 & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 & \frac{\beta}{2} + \frac{1}{2\beta} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{\beta} & 0 \\ -\frac{1}{\beta} & 0 & 0 & 0 & 0 & \frac{\beta}{2} + \frac{1}{2\beta} \end{pmatrix} \tag{B.3}$$

and the complementary channel output is

$$\boldsymbol{\sigma}_{F'} = \begin{pmatrix} \frac{2}{\beta} & 0 & 0 & -\frac{2}{\beta} \\ 0 & \frac{\beta}{2} + \frac{1}{2\beta} + p & 0 & z \\ 0 & 0 & \frac{2}{\beta} & 0 \\ -\frac{2}{\beta} & z & 0 & \frac{\beta}{2} + \frac{5}{2\beta} + x \end{pmatrix}. \tag{B.4}$$

The degrading map is obtained applying the symplectic Gaussian $\hat{S}'$ corresponding to

$$S' = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 & 0 & -1 \end{pmatrix} \tag{B.5}$$

and discarding the first mode. This is an explicit degrading Gaussian map and we can apply Theorem 4.4.3.

B.2 Coherent information of flagged additive gaussian noise

To compute the coherent information of the flagged additive noise we have to find the covariance matrix V_M of $\Lambda^e_\beta[\hat\rho_M]$ and the covariance matrix V'_M of $(\Lambda^e_\beta\otimes\mathcal{I})[|\rho_M\rangle\rangle\langle\langle\rho_M|]$ where $\hat\rho_M$ is the thermal state with average photon number M and $|\rho_M\rangle\rangle$ is its purification, which can be taken to be the two-mode squeezed state $|\tau\rangle$.

We obtain

$$V_M = \begin{pmatrix} 2M+1+\frac{2}{\beta} & 0 & 0 & 0 & 0 & -\frac{1}{\beta} \\ 0 & 2M+1+\frac{2}{\beta} & 0 & \frac{1}{\beta} & 0 & 0 \\ 0 & 0 & \frac{2}{\beta} & 0 & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 & \frac{\beta}{2}+\frac{1}{2\beta} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{\beta} & 0 \\ -\frac{1}{\beta} & 0 & 0 & 0 & 0 & \frac{\beta}{2}+\frac{1}{2\beta} \end{pmatrix} \tag{B.6}$$

$$V'_M = \begin{pmatrix} 2M+1+\frac{2}{\beta} & 0 & 0 & 0 & 0 & -\frac{1}{\beta} & 2\sqrt{M(M+1)} & 0 \\ 0 & 2M+1+\frac{2}{\beta} & 0 & \frac{1}{\beta} & 0 & 0 & 0 & -2\sqrt{M(M+1)} \\ 0 & 0 & \frac{2}{\beta} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\beta} & 0 & \frac{\beta}{2}+\frac{1}{2\beta} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{\beta} & 0 & 0 & 0 \\ -\frac{1}{\beta} & 0 & 0 & 0 & 0 & \frac{\beta}{2}+\frac{1}{2\beta} & 0 & 0 \\ 2\sqrt{M(M+1)} & 0 & 0 & 0 & 0 & 0 & 2M+1 & 0 \\ 0 & -2\sqrt{M(M+1)} & 0 & 0 & 0 & 0 & 0 & 2M+1 \end{pmatrix}. \tag{B.7}$$

The eigenvalues of $i\Omega V_M$ are

$$\pm\, 2M + O(1), \qquad \pm\frac{\sqrt{1+\beta^2}}{\beta} + O(1/M)\,, \qquad \pm\frac{\sqrt{1+\beta^2}}{\beta} + O(1/M)\,, \tag{B.8}$$

while the eigenvalues if $i\Omega V'_M$ are

$$\pm\, 2\frac{1}{\beta^{1/2}}\sqrt{M} + O(1), \qquad \pm 2\frac{1}{\beta^{1/2}}\sqrt{M} + O(1)\,, \qquad \pm 1\,, \qquad \pm 1\,. \tag{B.9}$$

Therefore we have

$$Q(\Lambda^e_\beta) = \lim_{M\to\infty} S(\Lambda^e_\beta[\rho_M]) - S((\Lambda^e_\beta\otimes\mathcal{I})[|\tau\rangle\langle\tau|_M]) = \log_2\beta - 1/\log 2 + 2h\left(\frac{\sqrt{1+\beta^2}}{\beta}\right), \tag{B.10}$$

as indicated in Eq. (8.72).

B.3 Coherent information of extended thermal attenuator

To compute the coherent information of the extended thermal attenuator $\mathcal{E}^e_{\eta,N}$ we have to find the covariance matrix V_M of $\mathcal{E}^e_{\eta,N}[\hat\rho_M]$ and the covariance matrix V'_M of the

complementary channel $\mathcal{E}^{e,c}_{\eta,N}[\hat{\rho}_M] = \mathcal{E}^{e}_{1-\eta,N}[\hat{\rho}_M]$ where $\hat{\rho}_M$ is again the thermal state with average photon number M. We obtain

$$V_M = \begin{pmatrix} \eta(2M+1)+(1-\eta)\eta(2N+1) & 0 & (1-\eta)2\sqrt{N(N+1)} & 0 \\ 0 & \eta(2M+1)+(1-\eta)(2N+1) & 0 & -(1-\eta)2\sqrt{N(N+1)} \\ (1-\eta)2\sqrt{N(N+1)} & 0 & \eta+(1-\eta)(2N+1) & 0 \\ 0 & -(1-\eta)2\sqrt{N(N+1)} & 0 & \eta+(1-\eta)(2N+1) \end{pmatrix}. \tag{B.11}$$

while V'_M is obtained from the above expression by exchanging $\eta \to 1-\eta$. The eigenvalues if $i\Omega V_M$ are hence

$$\pm\,\eta M + O(1), \qquad \pm(\eta + (1-\eta)(2N+1)) + O(1/M)\,, \tag{B.12}$$

while the eigenvalues if $i\Omega V'_M$ are

$$\pm\,(1-\eta)M + O(1), \qquad \pm((1-\eta) + \eta(2N+1)) + O(1/M)\,. \tag{B.13}$$

Therefore we have

$$\begin{aligned} Q(\mathcal{E}^{e}_{\eta,N}) &= \lim_{M\to\infty} S(\mathcal{E}^{e}_{\eta,N}[\hat{\rho}_M]) - S(\mathcal{E}^{e}_{1-\eta,N}[\hat{\rho}_M]) \\ &= -\log_2\left(\tfrac{\eta}{1-\eta}\right) + h(\eta + (1-\eta)(2N+1)) - h((1-\eta)+\eta(2N+1))\,, \end{aligned}$$

as indicated in Eq. (8.83).

B.4 Decomposition into irreducible representations of U(m)

In this section we determine the decomposition into irreducible representations of U(m) of the Hilbert space of m modes. In the following we switch to a complex notation for coherent states, i.e., $|\alpha\rangle := e^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}}\,|0\rangle$.

Since coherent states are an overcomplete set, we can first restrict to study the action of U(m) on coherent states and then straightforwardly extend the result to arbitrary bosonic states via the decomposition

$$\hat{\rho} = \int d^{2m}\vec{\alpha}\; \mathcal{P}_\rho(\vec{\alpha})\,|\vec{\alpha}\rangle\,\langle\vec{\alpha}|\,, \tag{B.14}$$

where $\mathcal{P}_\rho(\vec{\alpha})$ is the Glauber-Sudarshan P-representation [Gla63; Ser17] of the m-mode bosonic state $\hat{\rho}$.

We will make use of a crucial property that connects coherent states of an infinite-dimensional system with spin-coherent states of finite dimension [Per72; ZFG90]. First, note that an m-mode coherent state can be decomposed as

$$|\vec{\alpha}\rangle = \sum_{n=0}^{\infty} \sqrt{\mathbf{P}^{(s)}(n)}\,|\psi_n(\vec{\alpha})\rangle\,, \tag{B.15}$$

where $s := |\vec{\alpha}|^2$ the mean energy of the state, and $\hat{\Pi}_n |\vec{\alpha}\rangle = \sqrt{\mathbf{P}^{(s)}(n)}\, |\psi_n(\vec{\alpha})\rangle$. Explicitly

$$|\psi_n(\vec{\alpha})\rangle = \sum_{\sum_{i=1}^m n_i = n} \sqrt{\binom{n}{\{n_i\}}} \prod_{i=1}^m u_i^{n_i} |\vec{n}\rangle$$

where we have introduced the multi-mode Fock state $|\vec{n}\rangle = |n_1\rangle \otimes \cdots \otimes |n_m\rangle$, and $\vec{u} := \frac{\vec{\alpha}}{|\vec{\alpha}|}$. Now observe that each $|\psi_n(\vec{\alpha})\rangle$ lives in a finite-dimensional subspace and it can be mapped to the state of n copies of a m-level system state with coefficients $\vec{u}$:

$$\begin{aligned}(\sum_{i=1}^m u_i |i\rangle)^{\otimes n} &= \sum_{\sum_{i=1}^m n_i = n} \prod_{i=1}^m u_i^{n_i} \sum_{\sigma \in S_n} U(\sigma) \left|\vec{n}^{(m)}\right\rangle \\ &\cong |\psi_n(\vec{u})\rangle\,, \end{aligned} \tag{B.16}$$

where $\left|\vec{n}^{(m)}\right\rangle$ is the tensor-product state $\left|\vec{n}^{(m)}\right\rangle = |\underbrace{1, \cdots, 1}_{n_1}, \cdots \underbrace{m, \cdots, m}_{n_m}\rangle$, with n_i repetitions of the i-th basis element, $U(\sigma)$ is a permutation of the m-level systems and the isomorphism is defined on the basis of permutation-symmetric states $\binom{n}{\{n_i\}}^{-1/2} \sum_{\sigma \in S_n} U(\sigma) \left|\vec{n}^{(m)}\right\rangle \to |\vec{n}\rangle$. Finally, thanks to this mapping, the action of an energy-preserving Gaussian unitary $\hat{U}$ corresponding to $U \in \mathrm{U}(m)$ in phase space, can also be written as

$$\hat{U} |\vec{\alpha}\rangle = |U\vec{\alpha}\rangle = \sum_{n=0}^{\infty} \sqrt{\mathbf{P}^{(s)}(n)} \hat{d}_U^{(n,m)} |\psi_n(\vec{u})\rangle\,, \tag{B.17}$$

where $\hat{d}_U^{(n,m)}$ is the image of U with respect to the irreducible representation of $\mathrm{U}(m)$ on the permutation-symmetric subspace of n m-level systems. This is enough to conclude that each block with total photon number n hosts the irreducible representation of $\mathrm{U}(m)$ corresponding to the Young diagram of one row of length n, which has dimension $\binom{n+m-1}{m-1}$[Hay17b].

By Schur's lemma it then follows that the Haar average decoheres blocks with different total photon numbers and, inside each block with fixed total photon number, it acts as a $\mathrm{U}(m)$-twirling:

$$\begin{aligned}&\int_{\mathrm{U}(m)} dU \hat{U} |\vec{\alpha}\rangle \langle\vec{\alpha}| \hat{U}^\dagger = \sum_{n=0}^{\infty} \mathbf{P}^{(s)}(n) \int_{\mathrm{U}(m)} dU \hat{d}_U^{(n,m)} |\psi_n(\vec{u})\rangle \langle\psi_n(\vec{u})| \hat{d}_U^{(n,m)\dagger} \\ &= \sum_{n=0}^{\infty} \mathbf{P}^{(s)}(n) \frac{\hat{\Pi}_n}{\binom{n+m-1}{m-1}}. \end{aligned} \tag{B.18}$$

This result can then be applied to each coherent-state term in the decomposition of Eq. (B.14), obtaining Eq. ((9.7)) of the main text.

B.5 Pure-state ensembles are always optimal among Gaussian encodings

Consider an ensemble comprising general Gaussian states of the form $\hat{\rho}_{\mathrm{G}} = \hat{U}\hat{S}(\vec{r})\hat{D}(\vec{\alpha})\hat{\rho}_{\mathrm{th}}\hat{D}^{\dagger}(\vec{\alpha})\hat{S}^{\dagger}(\vec{r})\hat{U}^{\dagger}$, where $\hat{\rho}_{\mathrm{th}}$ is an m-mode thermal state , $\hat{U}$ is an m-mode PI and $\hat{D}(\vec{\alpha})$, $\hat{S}(\vec{r})$ are the tensor product of single-mode displacement operators $\hat{D}(\alpha_i) = \exp\Big(\alpha_i\hat{a}_i^{\dagger} - \alpha_i^{*}\hat{a}_i\Big)$ and squeezing operators $\hat{S}(r_i) = \exp\Big(\frac{r_i}{2}(\hat{a}_i^{2} - \hat{a}_i^{\dagger 2})\Big)$, respectively. Now recall that any thermal state can be decomposed as a mixture of coherent states with Gaussian weights, i.e., $\hat{\rho}_{\mathrm{th}} = \int d^{2m}\vec{\beta}\; p_{\mathrm{G}}(\vec{\beta})|\vec{\beta}\rangle\langle\vec{\beta}|$ [Ser17] and hence every Gaussian state $\hat{\rho}_{\mathrm{G}}$ can be written as a mixture of pure Gaussian states with Gaussian weight. Then for any mixed-state Gaussian ensemble $\mathcal{E}_{\mathrm{G}} := \{q(x), \hat{\rho}_{\mathrm{G}}(x)\}$, respecting the mean-energy constraint, one can consider an equivalent pure-state Gaussian ensemble $\tilde{\mathcal{E}}_{\mathrm{G}} := \{q(x)p_{\mathrm{G}}(\vec{\beta}|x), \hat{\Psi}_{G}(\vec{\beta}, x)\}$, comprising all the pure states $\hat{\Psi}(\vec{\beta}, x) = \Big|\psi(\vec{\beta}, x)\Big\rangle\Big\langle\psi(\vec{\beta}, x)\Big|$, with $\Big|\psi(\vec{\beta}, x)\Big\rangle = \hat{U}_x\hat{S}(\vec{r}_x)\hat{D}(\vec{\alpha}_x)\Big|\vec{\beta}\Big\rangle$, that take part in the decomposition of some $\hat{\rho}_{\mathrm{G}}(x)$, with proper weights. Then by the equivalence of these two ensembles and the concavity of the entropy we obtain, for any channel Φ acting on m bosonic modes,

$$\begin{aligned}\chi(\Phi, \mathcal{E}_{\mathrm{G}}) &= S\left(\int dx\, q(x)\Phi(\hat{\rho}_{\mathrm{G}}(x))\right) - \int dx\, q(x) S\left(\Phi(\hat{\rho}_{\mathrm{G}}(x))\right) \\ &\leq S\left(\int dx\, d^{2m}\vec{\beta}\, q(x)p_{\mathrm{G}}(\vec{\beta}|x)\Phi(\hat{\Psi}(\vec{\beta}, x))\right) \\ &\quad - \int dx\, d^{2m}\vec{\beta}\, q(x)p_{\mathrm{G}}(\vec{\beta}|x) S\left(\Phi(\hat{\Psi}(\vec{\beta}, x))\right) \\ &= \chi(\Phi, \tilde{\mathcal{E}}_{\mathrm{G}}).\end{aligned} \tag{B.19}$$

This implies that, when optimizing the Holevo quantity over Gaussian encodings, one can always restrict to pure states.

B.6 Communicate with phase reference

Consider now the scenario where Alice and Bob use a fraction of the total available energy xE to prepare a single mode state suitable for estimating the phase of the channel and $(1-x)E$ is the average energy of the ensemble of coherent states on the remaining $m-1$ modes. The input states have thus the form $|\psi\rangle \otimes |\vec{\alpha}\rangle$, with $|\vec{\alpha}\rangle = \otimes_{i=2}^{m}|\alpha_i\rangle$, $\langle\psi|\,\hat{n}_1\,|\psi\rangle = xE$, $\langle\vec{\alpha}|\,\hat{\sum}_{i=2}^{m}\hat{n}_i\,|\vec{\alpha}\rangle = |\alpha|^2$. Since Φ_m commutes with energy-preserving Gaussian unitaries on the last $m-1$ modes, one can adapt the argument in the main text to obtain an optimal rate

$$\chi_{\rm c}^{\rm ph}(\Phi_m, E, x) = \Bigg[S\left(\int dp(\vec{\alpha}) \Phi_m(|\psi\rangle\langle\psi| \otimes \int dU\, \hat{U} |\vec{\alpha}\rangle\langle\vec{\alpha}| \hat{U}) \right)$$
$$- \int dp(\vec{\alpha}) \int dU S(\Phi_m(|\psi\rangle\langle\psi| \otimes \hat{U} |\vec{\alpha}\rangle\langle\vec{\alpha}| \hat{U})) \Bigg]$$
$$= \Bigg[S\left(\int dp(\vec{\alpha}) \Phi_m(|\psi\rangle\langle\psi| \otimes \sum_{n=0}^{\infty} \mathbf{P}^{(|\alpha|^2)}(n) \frac{\hat{\Pi}_n^{(m-1)}}{\binom{n+m-2}{m-2}}) \right)$$
$$- \int dp(\vec{\alpha}) S(\Phi_m(|\psi\rangle\langle\psi| \otimes |\vec{\alpha}\rangle\langle\vec{\alpha}|)) \Bigg]. \tag{B.20}$$

where $\hat{\Pi}_n^{(m-1)}$ is the projector on the space of $m-1$ modes with total photon number n. The first term is the entropy of

$$\Phi_m(|\psi\rangle\langle\psi| \otimes \sum_{n=0}^{\infty} \mathbf{P}^{(|\alpha|^2)}(n) \frac{\hat{\Pi}_n^{(m-1)}}{\binom{n+m-2}{m-2}})$$
$$= \sum_{l=0}^{\infty} \mathbf{q}^{(xE)}(l) |l\rangle\langle l| \otimes \sum_{n=0}^{\infty} \mathbf{P}^{(|\alpha|^2)}(n) \frac{\hat{\Pi}_n^{(m-1)}}{\binom{n+m-2}{m-2}}, \tag{B.21}$$

where $\mathbf{q}^{(xE)}(n) := \mathrm{tr}\left[\hat{\Pi}_n |\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|\right]$, and the second term can be computed by noting that $\mathrm{tr}\left[\hat{\Pi}_n |\psi\rangle\langle\psi| \otimes |\vec{\alpha}\rangle\langle\vec{\alpha}|\right] = \sum_{l=0}^{n} \mathbf{q}^{(xE)}(l) \mathbf{P}^{(|\alpha|^2)}(n-l)$. Therefore, denoting $\mathbf{P}^{(s,E)}$ the probability distributions such that $\mathbf{P}^{(s,xE)}(n) = \sum_{l=0}^{n} \mathbf{q}^{(xE)}(l) \mathbf{P}^{(s)}(n-l)$, the rate is

$$\chi_{\rm c}^{\rm ph}(\Phi_m, E, x) = S[\mathbf{q}^{(xE)}] + S[\int_0^{\infty} ds\, p(s)\, \mathbf{P}^{(s)}]$$
$$+ \sum_{n=0}^{\infty} \int_0^{\infty} ds\, p(s)\, \mathbf{P}^{(s)}(n) \log \binom{n+m-2}{m-2} - \int_0^{\infty} ds\, p(s)\, S[\mathbf{P}^{(s,xE)}] \tag{B.22}$$

Using a coherent state as reference, and coding with a thermal ensemble for $m-1$ modes, at high energies we obtain, for fixed x, $0 < x < 1$,

$$\chi_{\rm c}^{\rm ph}(\Phi_m, E, x) = (m-1) \log E + \frac{1}{2} \log E - \frac{1}{2} \log E + O(1) \tag{B.23}$$

which is already sufficient to reach the upper bound at leading order. Other phase reference states are not useful at this level.

B.7 Squeezed-coherent encodings

The photon-number distribution of a coherent squeezed state $\hat{S}(r)\hat{D}(\alpha)\left|0\right\rangle$, $r \in \mathbb{R}$ and $\alpha \in \mathbb{C}$, is given by $p(n|r,\alpha) = |c(n|r,\alpha)|^2$, where [Yue76; GA90]

$$\begin{aligned} c(n|r,\alpha) &= (n!\cosh(r))^{-\frac{1}{2}} \left(\frac{1}{2}\tanh(r)\right)^{n/2} \\ &\times H_n\left[\alpha \sinh(2r)^{-1/2}\right] \exp\left[-\frac{1}{2}|\alpha|^2 - \frac{1}{2}\tanh(r)\alpha^2\right] \end{aligned} \tag{B.24}$$

and $H_n(\gamma)$ is the Hermite polynomial of order n. Taking $\alpha \in \mathbb{R}$, the average energy of the state is $E + \frac{1}{2} = \frac{1}{2}\cosh(2r) + e^{-2r}\alpha^2$. Substituting for α we then obtain

$$\begin{aligned} c(n|r,E) &= (n!\cosh(r))^{-\frac{1}{2}} \left(\frac{1}{2}\tanh(r)\right)^{n/2} \\ &\times H_n\left[\sqrt{\frac{(2E+1-\cosh(2r))e^{2r}}{2\sinh(2r)}}\right] \\ &\times \exp\left[-\frac{(2E+1-\cosh(2r))e^{2r}}{4}(1+\tanh(r))\right]. \end{aligned} \tag{B.25}$$

An achievable rate using these states for the encoding is obtained via the following on/off modulation:

$$\begin{aligned} \mathcal{E}_{p,\vec{r},\vec{\alpha},U} = \Big\{ &(1-p)\left|0\right\rangle\left\langle 0\right|^{\otimes m}, \\ &p\, dU\, \hat{U}\hat{S}(\vec{r})\hat{D}(\vec{\alpha})\left|0\right\rangle\left\langle 0\right|^{\otimes m}\hat{D}(\vec{\alpha})^\dagger\hat{S}(\vec{r})^\dagger\hat{U}^\dagger \Big\}, \end{aligned} \tag{B.26}$$

where the vacuum state is sent with probability $(1-p)$, while a pulse is sent with probability p. The latter is generated by a product of displacements and single-mode squeezing with fixed parameters on each mode, $\vec{r}, \vec{\alpha} \in \mathbb{R}^m$, followed by a Haar-random passive Gaussian unitary $\hat{U}$ on the m modes. All the parameters $p, \vec{r}, \vec{\alpha}$ are chosen so as to satisfy an average-energy constraint for the ensemble and the total photon number distribution is $\mathbf{Q}^{(\vec{r},\vec{\alpha})}$, with probabilities

$$\mathbf{Q}^{(\vec{r},\vec{\alpha})}(n) = \sum_{\sum_i n_i = n} \prod_{i=1}^{m} p(n_i|r_i,\alpha_i). \tag{B.27}$$

Following the same reasoning leading to Eq. (9.29), the rate achievable with this encoding is

$$\begin{aligned} R(\Phi_m, E, \vec{\alpha}, \vec{r}, p) &= p\sum_{n=1}^{\infty} \mathbf{Q}^{(\vec{r},\vec{\alpha})}(n) \log\binom{n+m-1}{m-1} \\ &+ \eta\left(1-p+p\mathbf{Q}^{(\vec{r},\vec{\alpha})}(0)\right) + \left(1-\mathbf{Q}^{(\vec{r},\vec{\alpha})}(0)\right)\eta(p) \\ &- p\,\eta\left(\mathbf{Q}^{(\vec{r},\vec{\alpha})}(0)\right). \end{aligned} \tag{B.28}$$

The rates of covariant ensembles generated from ternary encodings can be obtained by Eq. (9.9) simply putting $\mathbf{Q}^{(\vec{r},\vec{\alpha})}(n)$ as total photon-number distribution.

B.8 Photon number distribution of single mode Gaussian states

Single-mode Gaussian states $\hat{\rho}$ are characterized by the vector $\mathbf{m} = (x, p)$ and the covariance matrix $\boldsymbol{\sigma}$.

Their photon number distribution can be computed as follows [XHF10]. Redefining $\hat{x}$ and $\hat{p}$ with a rotation in the phase space, $\boldsymbol{\sigma}$ can be put to diagonal form. Define the variables:

$$2\tau_1^2 := \boldsymbol{\sigma}_{11} + 1 \qquad 2\tau_2^2 := \boldsymbol{\sigma}_{22} + 1, \tag{B.29}$$

$$A = \frac{1}{2\tau_1^2} + \frac{1}{2\tau_1^2} \qquad B = \frac{x}{\sqrt{2}\tau_2^2} + \frac{ip}{\sqrt{2}\tau_2^2}, \tag{B.30}$$

$$C = -\frac{1}{4\tau_2^2} + \frac{1}{4\tau_2^2} \qquad D = -\frac{x^2}{2\tau_2^2} - \frac{p^2}{2\tau_2^2}. \tag{B.31}$$

Then the probability of photon number n is

$$\begin{aligned} p(n, \mathbf{m}, V) &:= \mathrm{tr}[\hat{\rho}\,|n\rangle\langle n|] \\ &= \frac{e^D}{\tau_1\tau_2} \sum_{i=0}^{n} \frac{n!(1-A)^i|C|^{n-i}}{i![(n-i)!]^2} \left| H_m\left(\frac{iB}{2\sqrt{C}}\right)\right|^2. \end{aligned} \tag{B.32}$$

Bibliography

[Aar07] Scott Aaronson. "The learnability of quantum states". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 463.2088 (2007), pp. 3089–3114. DOI: 10.1098/rspa.2007.0113. arXiv: quant-ph/0608142 [quant-ph].

[Aar18] Scott Aaronson. "Shadow tomography of quantum states". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. New York, NY, USA: ACM, 2018, pp. 325–338. DOI: 10.1145/3188745.3188802. arXiv: 1711.01053.

[ABC16] Gerardo Adesso, Thomas R. Bromley, and Marco Cianciaruso. "Measures and applications of quantum correlations". In: *Journal of Physics A: Mathematical and Theoretical* 49.47 (2016), p. 473001. DOI: 10.1088/1751-8113/49/47/473001. arXiv: 1605.00806.

[ABG06] Esma Aïmeur, Gilles Brassard, and Sébastien Gambs. "Machine learning in a quantum world". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 4013 LNAI. Springer Verlag, 2006, pp. 431–442. DOI: 10.1007/11766247_37.

[Abr+19] Héctor Abraham, AduOffei, Rochisha Agarwal, Ismail Yunus Akhalwaya, Gadi Aleksandrowicz, Thomas Alexander, et al. *Qiskit: An Open-source Framework for Quantum Computing*. 2019. DOI: 10.5281/zenodo.2562110.

[AC97] C. Adami and N. J. Cerf. "von Neumann capacity of noisy quantum channels". In: *Physical Review A* 56.5 (1997), pp. 3470–3483. DOI: 10.1103/PhysRevA.56.3470. arXiv: quant-ph/9609024 [quant-ph].

[AD15] Jayadev Acharya and Constantinos Daskalakis. "Testing Poisson Binomial Distributions". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015, pp. 1829–1840. DOI: 10.1137/1.9781611973730.122. arXiv: 1507.05952.

[AFD19] Francesco Albarelli, Jamie F. Friel, and Animesh Datta. "Evaluating the Holevo Cramér-Rao Bound for Multiparameter Quantum Metrology". In: *Physical Review Letters* 123.20 (2019), p. 200503. DOI: 10.1103/PhysRevLett.123.200503. arXiv: 1906.05724.

[AG15] L. Alonso and T. Gorin. "Joint probability distributions for projection probabilities of random orthonormal states". In: *Journal of Physics A: Mathematical and Theoretical* 49.14 (2015). DOI: 10.1088/1751-8113/49/14/145004. arXiv: 1510.05333.

[AH11] Daiki Akimoto and Masahito Hayashi. "Discrimination of the change point in a quantum setting". In: *Physical Review A* 83.5 (2011), p. 052328. DOI: 10.1103/PhysRevA.83.052328. arXiv: 1102.2555.

[AKG19] Anirudh Acharya, Theodore Kypraios, and Madalin Guta. "A comparative study of estimation methods in quantum tomography". In: *Journal of Physics A: Mathematical and Theoretical* 52.23 (2019), p. 234001. DOI: 10.1088/1751-8121/ab1958. arXiv: 1901.07991.

[ALY10] José A. Adell, Alberto Lekuona, and Yaming Yu. "Sharp Bounds on the Entropy of the Poisson Law and Related Quantities". In: *IEEE Transactions on Information Theory* 56.5 (2010), pp. 2299–2306. DOI: 10.1109/TIT.2010.2044057. arXiv: arXiv:1001.2897.

[App05] D. M. Appleby. "Symmetric informationally complete–positive operator valued measures and the extended Clifford group". In: *Journal of Mathematical Physics* 46.5 (2005), p. 052107. DOI: 10.1063/1.1896384. arXiv: quant-ph/0412001 [arXiv:quant-ph].

[AR19] Scott Aaronson and Guy N. Rothblum. "Gentle measurement of quantum states and differential privacy". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. New York, NY, USA: ACM, 2019, pp. 322–333. DOI: 10.1145/3313276.3316378. arXiv: 1904.08747.

[ARS88] Robert Alicki, Slawomir Rudnicki, and Slawomir Sadowski. "Symmetry properties of product states for the system of N n-level atoms". In: *Journal of Mathematical Physics* 29.5 (1988), pp. 1158–1162. DOI: 10.1063/1.527958.

[ATP19] Hamza Adnane, Berihu Teklu, and Matteo G. A. Paris. "Quantum phase communication channels assisted by non-deterministic noiseless amplifiers". In: *Journal of the Optical Society of America B* 36.11 (2019), p. 2938. DOI: 10.1364/JOSAB.36.002938. arXiv: 1909.07138.

[Aud+08] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete. "Asymptotic Error Rates in Quantum Hypothesis Testing". In: *Communications in*

Mathematical Physics 279.1 (2008), pp. 251–283. DOI: 10.1007/s00220-008-0417-5. arXiv: 0708.4282.

[Ban+20] Konrad Banaszek, Ludwig Kunz, Michal Jachura, and Marcin Jarzyna. "Quantum Limits in Optical Communications". In: *Journal of Lightwave Technology* 38.10 (2020), pp. 2741–2754. DOI: 10.1109/JLT.2020.2973890. arXiv: 2002.05766.

[Ban+97] Masashi Ban, Keiko Kurokawa, Rei Momose, and Osamu Hirota. "Optimum measurements for discrimination among symmetric quantum states and parameter estimation". In: *International Journal of Theoretical Physics* 36.6 (1997), pp. 1269–1288. DOI: 10.1007/BF02435921.

[Bar01] Stephen M. Barnett. "Minimum-error discrimination between multiply symmetric states". In: *Physical Review A* 64.3 (2001), p. 030303. DOI: 10.1103/PhysRevA.64.030303.

[Bar+08] Lucie Bartůšková, Antonín Černoch, Jan Soubusta, and Miloslav Dušek. "Programmable discriminator of coherent states: Experimental realization". In: *Physical Review A* 77.3 (2008), p. 034306. DOI: 10.1103/PhysRevA.77.034306. arXiv: 0711.4712.

[Bat+01] Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. "Testing random variables for independence and identity". In: *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. IEEE, 2001, pp. 442–451. DOI: 10.1109/SFCS.2001.959920.

[BB14] Charles H. Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". In: *Theor. Comput. Sci.* 560.P1 (2014), pp. 7–11. DOI: 10.1016/j.tcs.2014.05.025. arXiv: arXiv:2003.06557.

[BC09] Stephen M. Barnett and Sarah Croke. "Quantum state discrimination". In: *Advances in Optics and Photonics* 1.2 (2009), p. 238. DOI: 10.1364/AOP.1.000238.

[BCH06] Dave Bacon, Isaac L. Chuang, and Aram W. Harrow. "Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms". In: *Physical Review Letters* 97.17 (2006), p. 170502. DOI: 10.1103/PhysRevLett.97.170502. arXiv: quant-ph/0407082 [quant-ph].

[BCL20] Sebastien Bubeck, Sitan Chen, and Jerry Li. *Entanglement is Necessary for Optimal Quantum Property Testing*. 2020. DOI: 10.1109/FOCS46700.2020.00070. arXiv: 2004.07869.

[BD10] Francesco Buscemi and Nilanjana Datta. "The Quantum Capacity of Channels With Arbitrarily Correlated Noise". In: *IEEE Transactions on Information Theory* 56.3 (2010), pp. 1447–1460. DOI: 10.1109/TIT.2009.2039166. arXiv: 0902.0158.

[BD11] Fernando G. S. L. Brandao and Nilanjana Datta. "One-Shot Rates for Entanglement Manipulation Under Non-entangling Maps". In: *IEEE Transactions on Information Theory* 57.3 (2011), pp. 1754–1760. DOI: 10.1109/TIT.2011.2104531. arXiv: 0905.2673.

[BDS97] Charles H. Bennett, David P. DiVincenzo, and John A. Smolin. "Capacities of Quantum Erasure Channels". In: *Physical Review Letters* 78.16 (1997), pp. 3217–3220. DOI: 10.1103/PhysRevLett.78.3217. arXiv: quant-ph/9701015 [quant-ph].

[Ben+02] C.H. Bennett, P. W. Shor, J.A. Smolin, and A.V. Thapliyal. "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem". In: *IEEE Transactions on Information Theory* 48.10 (2002), pp. 2637–2655. DOI: 10.1109/TIT.2002.802612. arXiv: quant-ph/0106052 [quant-ph].

[Ben+14] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. "The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels". In: *IEEE Transactions on Information Theory* 60.5 (2014), pp. 2926–2959. DOI: 10.1109/TIT.2014.2309968. arXiv: arXiv:0912.5537.

[Ben+19] Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam, and Alejandro Perdomo-Ortiz. "A generative modeling approach for benchmarking and training shallow quantum circuits". In: *npj Quantum Information* 5.1 (2019), p. 45. DOI: 10.1038/s41534-019-0157-8. arXiv: 1801.07686.

[Ben+93] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels". In: *Physical Review Letters* 70.13 (1993), pp. 1895–1899. DOI: 10.1103/PhysRevLett.70.1895.

[Ben+96] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. "Mixed-state entanglement and quantum error correction". In: *Physical Review A* 54.5 (1996), pp. 3824–3851. DOI: 10.1103/PhysRevA.54.3824. arXiv: quant-ph/9604024 [quant-ph].

[Ben+99] Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. "Entanglement-Assisted Classical Capacity of Noisy Quantum Channels". In: *Physical Review Letters* 83.15 (1999), pp. 3081–3084. DOI: 10.1103/PhysRevLett.83.3081. arXiv: quant-ph/9904023 [quant-ph].

[Ber10] János A. Bergou. "Discrimination of quantum states". In: *Journal of Modern Optics* 57.3 (2010), pp. 160–180. DOI: 10.1080/09500340903477756.

[Ber+13] Mario Berta, Fernando G. S. L. Brandao, Matthias Christandl, and Stephanie Wehner. "Entanglement Cost of Quantum Channels". In: *IEEE Transactions on Information Theory* 59.10 (2013), pp. 6779–6795. DOI: 10.1109/TIT.2013.2268533. arXiv: 1108.5357.

[BFH06] János A. Bergou, Edgar Feldman, and Mark Hillery. "Optimal unambiguous discrimination of two subspaces as a case in mixed-state discrimination". In: *Physical Review A* 73.3 (2006), p. 032107. DOI: 10.1103/PhysRevA.73.032107. arXiv: quant-ph/0602093 [quant-ph].

[BH05] János A. Bergou and Mark Hillery. "Universal programmable quantum state discriminator that is optimal for unambiguously distinguishing between unknown states". In: *Physical Review Letters* 94.16 (2005), p. 160501. DOI: 10.1103/PhysRevLett.94.160501. arXiv: quant-ph/0504201 [arXiv:quant-ph].

[BH09] Jop Briët and Peter Harremoës. "Properties of classical and quantum Jensen-Shannon divergence". In: *Physical Review A* 79.5 (2009), p. 052311. DOI: 10.1103/PhysRevA.79.052311. arXiv: 0806.4472.

[Bia+17] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. "Quantum machine learning". In: *Nature* 549.7671 (2017), pp. 195–202. DOI: 10.1038/nature23474. arXiv: 1611.09347.

[BIMT06] E. Bagan, S. Iblisdir, and R. Muñoz-Tapia. "Relative states, quantum axes, and quantum references". In: *Physical Review A* 73.2 (2006), p. 022341. DOI: 10.1103/PhysRevA.73.022341. arXiv: quant-ph/0508187 [quant-ph].

[Bis+10] Alessandro Bisio, Giulio Chiribella, Giacomo Mauro D'Ariano, Stefano Facchini, and Paolo Perinotti. "Optimal quantum learning of a unitary transformation". In: *Physical Review A* 81.3 (2010), p. 032324. DOI: 10.1103/PhysRevA.81.032324. arXiv: 0903.0543.

[BK02] H. Barnum and E. Knill. "Reversing quantum dynamics with near-optimal quantum and classical fidelity". In: *Journal of Mathematical Physics* 43.5 (2002), p. 2097. DOI: 10.1063/1.1459754. arXiv: quant-ph/0004088 [quant-ph].

[BK15] Joonwoo Bae and Leong-Chuan Kwek. "Quantum state discrimination and its applications". In: *Journal of Physics A: Mathematical and Theoretical* 48.8 (2015), p. 083001. DOI: 10.1088/1751-8113/48/8/083001. arXiv: 1707.02571.

[BKN00] H. Barnum, E. Knill, and M.A. Nielsen. "On quantum fidelities and channel capacities". In: *IEEE Transactions on Information Theory* 46.4 (2000),

pp. 1317–1329. DOI: 10.1109/18.850671. arXiv: quant-ph/9809010 [quant-ph].

[BL19] Johannes Bausch and Felix Leditzky. *Error Thresholds for Arbitrary Pauli Noise*. 2019. arXiv: 1910.00471.

[BL20] Johannes Bausch and Felix Leditzky. "Quantum codes from neural networks". In: *New Journal of Physics* 22.2 (2020), p. 023005. DOI: 10.1088/1367-2630/ab6cdd. arXiv: 1806.08781.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities*. Oxford University Press, 2013. DOI: 10.1093/acprof:oso/9780199535255.001.0001.

[BOW19] Costin Bădescu, Ryan O'Donnell, and John Wright. "Quantum state certification". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. New York, NY, USA: ACM, 2019, pp. 503–514. DOI: 10.1145/3313276.3316344. arXiv: 1708.06002.

[BR82] Jacob Burbea and C. Rao. "On the convexity of some divergence measures based on entropy functions". In: *IEEE Transactions on Information Theory* 28.3 (1982), pp. 489–495. DOI: 10.1109/TIT.1982.1056497.

[BRS04] Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens. "Optimal measurements for relative quantum information". In: *Physical Review A* 70.3 (2004), p. 032321. DOI: 10.1103/PhysRevA.70.032321. arXiv: quant-ph/0310009 [quant-ph].

[BRS07] Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens. "Reference frames, superselection rules, and quantum information". In: *Reviews of Modern Physics* 79.2 (2007), pp. 555–609. DOI: 10.1103/RevModPhys.79.555. arXiv: quant-ph/0610030 [quant-ph].

[Bru+98] Dagmar Bruß, David P. DiVincenzo, Artur Ekert, Christopher A. Fuchs, Chiara Macchiavello, and John A. Smolin. "Optimal universal and state-dependent quantum cloning". In: *Physical Review A* 57.4 (1998), pp. 2368–2378. DOI: 10.1103/PhysRevA.57.2368. arXiv: quant-ph/9705038 [quant-ph].

[Buh+01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. "Quantum Fingerprinting". In: *Physical Review Letters* 87.16 (2001), p. 167902. DOI: 10.1103/PhysRevLett.87.167902. arXiv: quant-ph/0102001 [quant-ph].

[BW00] D. W. Berry and H. M. Wiseman. "Optimal States and Almost Optimal Adaptive Measurements for Quantum Interferometry". In: *Physical Review Letters* 85.24 (2000), pp. 5098–5101. DOI: 10.1103/PhysRevLett.85.5098. arXiv: quant-ph/0009117 [quant-ph].

[BW18] Mario Berta and Mark M. Wilde. "Amortization does not enhance the max-Rains information of a quantum channel". In: *New Journal of Physics* 20.5 (2018), p. 053044. DOI: 10.1088/1367-2630/aac153. arXiv: 1709.04907.

[BW92] Charles H. Bennett and Stephen J. Wiesner. "Communication via one- and two-particle operators on Einstein-Podolsky- Rosen states". In: *Physical Review Letters* 69.20 (1992), pp. 2881–2884. DOI: 10.1103/PhysRevLett.69.2881.

[Can20] Clement L. Canonne. "A Survey on Distribution Testing: Your Data is Big. But is it Blue?" In: *Theory of Computing* 1.1 (2020), pp. 1–100. DOI: 10.4086/toc.gs.2020.009.

[Car+14] Filippo Caruso, Vittorio Giovannetti, Cosmo Lupo, and Stefano Mancini. "Quantum channels and memory effects". In: *Rev. Mod. Phys.* 86.4 (2014), pp. 1203–1259. DOI: 10.1103/RevModPhys.86.1203. arXiv: 1207.5435.

[Cav81] Carlton M. Caves. "Quantum-mechanical noise in an interferometer". In: *Physical Review D* 23.8 (1981), pp. 1693–1708. DOI: 10.1103/PhysRevD.23.1693.

[CB08] Luca Chirolli and Guido Burkard. "Decoherence in solid-state qubits". In: *Advances in Physics* 57.3 (2008), pp. 225–285. DOI: 10.1080/00018730802218067. arXiv: 0809.4716.

[CB98] Anthony Chefles and Stephen M Barnett. "Optimum unambiguous discrimination between linearly independent symmetric states". In: *Physics Letters A* 250.4-6 (1998), pp. 223–229. DOI: 10.1016/S0375-9601(98)00827-5. arXiv: quant-ph/9807023 [quant-ph].

[CCP14] Gianfranco Cariolaro, Roberto Corvaja, and Gianfranco Pierobon. "Gaussian states and geometrically uniform symmetry". In: *Physical Review A* 90.4 (2014), p. 042309. DOI: 10.1103/PhysRevA.90.042309. arXiv: arXiv:1410.5282.

[CD94] Carlton M. Caves and P. D. Drummond. "Quantum limits on bosonic communication rates". In: *Reviews of Modern Physics* 66.2 (1994), pp. 481–537. DOI: 10.1103/RevModPhys.66.481.

[CDP08] Giulio Chiribella, Giacomo M. D'Ariano, and Paolo Perinotti. "Memory Effects in Quantum Channel Discrimination". In: *Physical Review Letters* 101.18 (2008), p. 180501. DOI: 10.1103/PhysRevLett.101.180501. arXiv: 0803.3237.

[Cer00] Nicolas J. Cerf. "Pauli Cloning of a Quantum Bit". In: *Physical Review Letters* 84.19 (2000), pp. 4497–4500. DOI: 10.1103/PhysRevLett.84.4497. arXiv: quant-ph/9803058 [quant-ph].

[CFG19] Stefano Chessa, Marco Fanizza, and Vittorio Giovannetti. "Quantum-capacity bounds in spin-network communication channels". In: *Physical Review A* 100.3 (2019), p. 032311. DOI: 10.1103/PhysRevA.100.032311. arXiv: 1905.11920.

[CG06] Filippo Caruso and Vittorio Giovannetti. "Degradability of Bosonic Gaussian channels". In: *Physical Review A* 74.6 (2006), p. 062307. DOI: 10.1103/PhysRevA.74.062307. arXiv: quant-ph/0603257 [quant-ph].

[CGH06] F Caruso, V Giovannetti, and A S Holevo. "One-mode bosonic Gaussian channels: a full weak-degradability classification". In: *New Journal of Physics* 8.12 (2006), pp. 310–310. DOI: 10.1088/1367-2630/8/12/310. arXiv: quant-ph/0609013 [quant-ph].

[CGZ11] Hye Won Chung, Saikat Guha, and Lizhong Zheng. "On capacity of optical channels with coherent detection". In: *IEEE International Symposium on Information Theory - Proceedings*. IEEE, 2011, pp. 284–288. DOI: 10.1109/ISIT.2011.6034095.

[CH03] C.-L. Chou and L. Y. Hsu. "Minimum-error discrimination between symmetric mixed quantum states". In: *Physical Review A* 68.4 (2003), p. 042305. DOI: 10.1103/PhysRevA.68.042305. arXiv: quant-ph/0304117 [arXiv:quant-ph].

[Cha+14] Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. "Optimal Algorithms for Testing Closeness of Discrete Distributions". In: *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2014, pp. 1193–1203. DOI: 10.1137/1.9781611973402.88. arXiv: 1308.3946.

[Cha+18] Ulysse Chabaud, Eleni Diamanti, Damian Markham, Elham Kashefi, and Antoine Joux. "Optimal quantum-programmable projective measurement with linear optics". In: *Physical Review A* 98.6 (2018), p. 062318. DOI: 10.1103/PhysRevA.98.062318. arXiv: 1805.02546.

[Che98] Anthony Chefles. "Unambiguous discrimination between linearly independent quantum states". In: *Physics Letters A* 239.6 (1998), pp. 339–347. DOI: 10.1016/S0375-9601(98)00064-4.

[Chr06] Matthias Christandl. *The Structure of Bipartite Quantum States - Insights from Group Theory and Cryptography*. 2006. arXiv: quant-ph/0604183 [quant-ph].

[Cil+17] Carlo Ciliberto, Mark Herbster, Alessandro Davide Ialongo, Massimiliano Pontil, Andrea Rocchetto, Simone Severini, et al. "Quantum machine learning: a classical perspective". In: *Proceedings of the Royal Society A:*

Mathematical, Physical and Engineering Sciences 474.2209 (2017). DOI: 10.1098/rspa.2017.0551. arXiv: 1707.08561.

[Cin+18] Lukasz Cincio, Yiğit Subaşi, Andrew T. Sornborger, and Patrick J. Coles. "Learning the quantum algorithm for state overlap". In: *New Journal of Physics* 20.11 (2018). DOI: 10.1088/1367-2630/aae94a. arXiv: 1803.04114.

[CLO21] Sitan Chen, Jerry Li, and Ryan O'Donnell. *Toward Instance-Optimal State Certification With Incoherent Measurements*. 2021. arXiv: 2102.13098.

[CM06] Matthias Christandl and Graeme Mitchison. "The spectra of quantum states and the Kronecker coefficients of the symmetric group". In: *Communications in Mathematical Physics* 261.3 (2006), pp. 789–797. DOI: 10.1007/s00220-005-1435-1. arXiv: quant-ph/0409016 [quant-ph].

[CMH17] Matthias Christandl and Alexander Müller-Hermes. "Relative Entropy Bounds on Quantum, Private and Repeater Capacities". In: *Communications in Mathematical Physics* 353.2 (2017), pp. 821–852. DOI: 10.1007/s00220-017-2885-y. arXiv: 1604.03448.

[Col12] A. J. T. Colin. "Programmed discrimination of multiple sets of qbits with added classical information". In: *The European Physical Journal D* 66.7 (2012), p. 185. DOI: 10.1140/epjd/e2012-20618-3.

[COP18] Giovanni Chesi, Stefano Olivares, and Matteo G. A. Paris. "Squeezing-enhanced phase-shift-keyed binary communication in noisy channels". In: *Physical Review A* 97.3 (2018), p. 032315. DOI: 10.1103/PhysRevA.97.032315. arXiv: 1710.09577.

[CR19] Mahdi Cheraghchi and Joao Ribeiro. "Improved Upper Bounds and Structural Results on the Capacity of the Discrete-Time Poisson Channel". In: *IEEE Transactions on Information Theory* 65.7 (2019), pp. 4052–4068. DOI: 10.1109/TIT.2019.2896931.

[Cro+06] Sarah Croke, Erika Andersson, Stephen M. Barnett, Claire R. Gilson, and John Jeffers. "Maximum Confidence Quantum Measurements". In: *Physical Review Letters* 96.7 (2006), p. 070401. DOI: 10.1103/PhysRevLett.96.070401. arXiv: quant-ph/0604026 [arXiv:quant-ph].

[CRS08] Toby S. Cubitt, Mary Beth Ruskai, and Graeme Smith. "The structure of degradable quantum channels". In: *Journal of Mathematical Physics* 49.10 (2008), p. 102104. DOI: 10.1063/1.2953685. arXiv: 0802.1360.

[CT05] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. 2005. DOI: 10.1002/047174882X.

[Cub+15] Toby Cubitt, David Elkouss, William Matthews, Maris Ozols, David Pérez-García, and Sergii Strelchuk. "Unbounded number of channel uses

may be required to detect quantum capacity". In: *Nature Communications* 6.1 (2015), p. 6739. DOI: 10.1038/ncomms7739. arXiv: 1408.5115.

[CWY04] N. Cai, A. Winter, and R. W. Yeung. "Quantum privacy and quantum wiretap channels". In: *Problems of Information Transmission* 40.4 (2004), pp. 318–336. DOI: 10.1007/s11122-005-0002-x.

[Dat09] Nilanjana Datta. "Min- and Max-Relative Entropies and a New Entanglement Monotone". In: *IEEE Transactions on Information Theory* 55.6 (2009), pp. 2816–2826. DOI: 10.1109/TIT.2009.2018325. arXiv: 0803.2770.

[Day+19] Alexandre G.R. Day, Marin Bukov, Phillip Weinberg, Pankaj Mehta, and Dries Sels. "Glassy Phase of Optimal Quantum Control". In: *Physical Review Letters* 122.2 (2019), p. 020601. DOI: 10.1103/PhysRevLett.122.020601. arXiv: 1803.10856.

[DB02] Miloslav Dušek and Vladimír Bužek. "Quantum-controlled measurement device for quantum-state discrimination". In: *Physical Review A* 66.2 (2002), p. 022112. DOI: 10.1103/PhysRevA.66.022112.

[DB18] Vedran Dunjko and Hans J. Briegel. "Machine learning & artificial intelligence in the quantum domain: A review of recent progress". In: *Reports on Progress in Physics* 81.7 (2018), p. 74001. DOI: 10.1088/1361-6633/aab406. arXiv: 1709.02779.

[dBe04] J. Niel de Beaudrap. "One-qubit fingerprinting schemes". In: *Physical Review A* 69.2 (2004), p. 022307. DOI: 10.1103/PhysRevA.69.022307. arXiv: quant-ph/0309036 [arXiv:quant-ph].

[dBe13] Niel de Beaudrap. "A linearized stabilizer formalism for systems of finite dimension". In: *Quantum Information and Computation* 13.1-2 (2013), pp. 0073–0115. DOI: 10.26421/qic13.1-2-6. arXiv: 1102.3354.

[Dev05] Igor Devetak. "The Private Classical Capacity and Quantum Capacity of a Quantum Channel". In: *IEEE Transactions on Information Theory* 51.1 (2005), pp. 44–55. DOI: 10.1109/TIT.2004.839515. arXiv: quant-ph/0304127 [quant-ph].

[DHW04] Igor Devetak, Aram W. Harrow, and Andreas Winter. "A Family of Quantum Protocols". In: *Physical Review Letters* 93.23 (2004), p. 230504. DOI: 10.1103/PhysRevLett.93.230504. arXiv: quant-ph/0308044 [quant-ph].

[Die88] D. Dieks. "Overlap and distinguishability of quantum states". In: *Physics Letters A* 126.5-6 (1988), pp. 303–306. DOI: 10.1016/0375-9601(88)90840-7.

[DiM+19] M. T. DiMario, L. Kunz, K. Banaszek, and F. E. Becerra. "Optimized communication strategies with binary coherent states over phase noise

channels". In: *npj Quantum Information* 5.1 (2019), p. 65. DOI: 10.1038/s41534-019-0177-4. arXiv: 1907.12515.

[DK16] Ilias Diakonikolas and Daniel M. Kane. "A New Approach for Testing Properties of Discrete Distributions". In: *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2016, pp. 685–694. DOI: 10.1109/FOCS.2016.78. arXiv: 1601.05557.

[DKN15] Ilias Diakonikolas, Daniel M. Kane, and Vladimir Nikishkin. "Testing Identity of Structured Distributions". In: *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. Vol. 2015-Janua. January. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015, pp. 1841–1854. DOI: 10.1137/1.9781611973730.123.

[DL15] Eleni Diamanti and Anthony Leverrier. "Distributing Secret Keys with Quantum Continuous Variables: Principle, Security and Implementations". In: *Entropy* 17.12 (2015), pp. 6072–6092. DOI: 10.3390/e17096072. arXiv: 1506.02888.

[Don+18] Pietro Donà, Marco Fanizza, Giorgio Sarno, and Simone Speziale. "SU(2) graph invariants, Regge actions and polytopes". In: *Classical and Quantum Gravity* 35.4 (2018), p. 045011. DOI: 10.1088/1361-6382/aaa53a. arXiv: 1708.01727.

[Don+19] Pietro Donà, Marco Fanizza, Giorgio Sarno, and Simone Speziale. "Numerical study of the Lorentzian Engle-Pereira-Rovelli-Livine spin foam amplitude". In: *Physical Review D* 100.10 (2019), p. 106003. DOI: 10.1103/PhysRevD.100.106003. arXiv: 1903.12624.

[Don+20] Pietro Dona, Marco Fanizza, Pierre Martin-Dussaud, and Simone Speziale. *Asymptotics of* $\mathrm{SL}(2,\mathbb{C})$ *coherent invariant tensors*. 2020. arXiv: 2011.13909.

[DRC17] C. L. Degen, F. Reinhard, and P. Cappellaro. "Quantum sensing". In: *Reviews of Modern Physics* 89.3 (2017), p. 035002. DOI: 10.1103/RevModPhys.89.035002. arXiv: 1611.02427.

[DS05] I. Devetak and P. W. Shor. "The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information". In: *Communications in Mathematical Physics* 256.2 (2005), pp. 287–303. DOI: 10.1007/s00220-005-1317-6. arXiv: quant-ph/0311131 [quant-ph].

[DSS98] David P. DiVincenzo, Peter W. Shor, and John A. Smolin. "Quantum-channel capacity of very noisy channels". In: *Physical Review A* 57.2 (1998), pp. 830–839. DOI: 10.1103/PhysRevA.57.830. arXiv: quant-ph/9706061 [quant-ph].

[EF01] Y.C. Eldar and G.D. Forney. "On quantum detection and the square-root measurement". In: *IEEE Transactions on Information Theory* 47.3

(2001), pp. 858–872. DOI: 10.1109/18.915636. arXiv: quant-ph/0005132 [quant-ph].

[Eke+02] Artur K. Ekert, Carolina Moura Alves, Daniel K. L. Oi, Michał Horodecki, Paweł Horodecki, and L. C. Kwek. "Direct Estimations of Linear and Nonlinear Functionals of a Quantum State". In: *Physical Review Letters* 88.21 (2002), p. 217901. DOI: 10.1103/PhysRevLett.88.217901. arXiv: quant-ph/0203016 [quant-ph].

[Eld03a] Y.C. Eldar. "A semidefinite programming approach to optimal unambiguous discrimination of quantum states". In: *IEEE Transactions on Information Theory* 49.2 (2003), pp. 446–456. DOI: 10.1109/TIT.2002.807291. arXiv: quant-ph/0206093 [quant-ph].

[Eld03b] Yonina C. Eldar. "Mixed-quantum-state detection with inconclusive results". In: *Physical Review A* 67.4 (2003), p. 042309. DOI: 10.1103/PhysRevA.67.042309. arXiv: quant-ph/0211121 [arXiv:quant-ph].

[Ele14] Neven Elezović. "Asymptotic expansions of central binomial coefficients and catalan numbers". In: *Journal of Integer Sequences* 17.2 (2014).

[Ell71] David Elliott. *Uniform Asymptotic Expansions of the Jacobi Polynomials and an Associated Function*. Tech. rep. 1971.

[EMV03] Y.C. Eldar, Alexandre Megretski, and G.C. Verghese. "Designing optimal quantum detectors via semidefinite programming". In: *IEEE Transactions on Information Theory* 49.4 (2003), pp. 1007–1012. DOI: 10.1109/TIT.2003.809510. arXiv: quant-ph/0205178v1 [quant-ph].

[EMV04] Y.C. Eldar, Alexandre Megretski, and G.C. Verghese. "Optimal Detection of Symmetric Mixed Quantum States". In: *IEEE Transactions on Information Theory* 50.6 (2004), pp. 1198–1207. DOI: 10.1109/TIT.2004.828070. arXiv: quant-ph/0211111 [quant-ph].

[ES03] D.M. Endres and J.E. Schindelin. "A new metric for probability distributions". In: *IEEE Transactions on Information Theory* 49.7 (2003), pp. 1858–1860. DOI: 10.1109/TIT.2003.813506.

[ES15] David Elkouss and Sergii Strelchuk. "Superadditivity of Private Information for Any Number of Uses of the Channel". In: *Physical Review Letters* 115.4 (2015), p. 040501. DOI: 10.1103/PhysRevLett.115.040501. arXiv: 1502.05326.

[Fan+20a] M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti. "Beyond the Swap Test: Optimal Estimation of Quantum State Overlap". In: *Physical Review Letters* 124.6 (2020), p. 060503. DOI: 10.1103/PhysRevLett.124.060503. arXiv: 1906.10639.

[Fan+20b] Marco Fanizza, Matteo Rosati, Michalis Skotiniotis, John Calsamiglia, and Vittorio Giovannetti. *Classical capacity of quantum Gaussian codes without a phase reference: when squeezing helps*. 2020. arXiv: 2006.06522.

[FD04] Jaromír Fiurášek and Miloslav Dušek. "Probabilistic quantum multimeters". In: *Physical Review A - Atomic, Molecular, and Optical Physics* 69.3 (2004), p. 032302. DOI: 10.1103/PhysRevA.69.032302. arXiv: quant-ph/0201097 [quant-ph].

[FDF02] Jaromír Fiurášek, Miloslav Dušek, and Radim Filip. "Universal Measurement Apparatus Controlled by Quantum Software". In: *Physical Review Letters* 89.19 (2002), p. 190401. DOI: 10.1103/PhysRevLett.89.190401. arXiv: quant-ph/0202152 [quant-ph].

[FDY04] Yuan Feng, Runyao Duan, and Mingsheng Ying. "Unambiguous discrimination between mixed quantum states". In: *Physical Review A* 70.1 (2004), p. 012308. DOI: 10.1103/PhysRevA.70.012308. arXiv: quant-ph/0403147 [arXiv:quant-ph].

[Fen+02] Yuan Feng, Shengyu Zhang, Runyao Duan, and Mingsheng Ying. "Lower bound on inconclusive probability of unambiguous discrimination". In: *Physical Review A* 66.6 (2002), p. 062313. DOI: 10.1103/PhysRevA.66.062313.

[Fey82] Richard P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21.6-7 (1982), pp. 467–488. DOI: 10.1007/BF02650179.

[FJ03] Jaromír Fiurášek and Miroslav Ježek. "Optimal discrimination of mixed quantum states involving inconclusive results". In: *Physical Review A* 67.1 (2003), p. 012321. DOI: 10.1103/PhysRevA.67.012321. arXiv: quant-ph/0208126 [arXiv:quant-ph].

[FKG20] Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. "Quantum Flags and New Bounds on the Quantum Capacity of the Depolarizing Channel". In: *Physical Review Letters* 125.2 (2020), p. 020503. DOI: 10.1103/PhysRevLett.125.020503. arXiv: 1911.01977.

[FKG21] Marco Fanizza, Farzad Kianvash, and Vittorio Giovannetti. *Estimating Quantum and Private capacities of Gaussian channels via degradable extensions*. 2021. arXiv: 2103.09569.

[Fla+12] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. "Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators". In: *New Journal of Physics* 14.9 (2012), p. 095022. DOI: 10.1088/1367-2630/14/9/095022. arXiv: 1205.2300.

[FMG19] Marco Fanizza, Andrea Mari, and Vittorio Giovannetti. "Optimal Universal Learning Machines for Quantum State Discrimination". In: *IEEE*

Transactions on Information Theory 65.9 (2019), pp. 5931–5944. DOI: 10.1109/TIT.2019.2916646. arXiv: 1805.03477.

[FSG21] Marco Fanizza, Raffaele Salvia, and Vittorio Giovannetti. *Testing identity of collections of quantum states: sample complexity analysis*. 2021. arXiv: 2103.14511.

[FW08] Jesse Fern and K. Birgitta Whaley. "Lower bounds on the nonzero capacity of Pauli channels". In: *Physical Review A* 78.6 (2008), p. 062335. DOI: 10.1103/PhysRevA.78.062335. arXiv: 0708.1597.

[GA90] J. J. Gong and P. K. Aravind. "Expansion coefficients of a squeezed coherent state in the number state basis". In: *Am. J. Phys.* 58.10 (1990), pp. 1003–1006. DOI: 10.1119/1.16337.

[GB10] Xavier Glorot and Yoshua Bengio. *Understanding the difficulty of training deep feedforward neural networks*. Tech. rep. 2010, pp. 249–256.

[GE08] D. Gross and J. Eisert. "Quantum margulis expanders". In: *Quantum Information and Computation* 8.8-9 (2008), pp. 0722–0733. DOI: 10.26421/qic8.8-9-3. arXiv: 0710.0651.

[GF05] Vittorio Giovannetti and Rosario Fazio. "Information-capacity description of spin-chain correlations". In: *Physical Review A* 71.3 (2005), p. 032314. DOI: 10.1103/PhysRevA.71.032314. arXiv: quant-ph/0405110 [quant-ph].

[GHGP15] V. Giovannetti, A. S. Holevo, and R. García-Patrón. "A Solution of Gaussian Optimizer Conjecture for Quantum Channels". In: *Communications in Mathematical Physics* 334.3 (2015), pp. 1553–1571. DOI: 10.1007/s00220-014-2150-6. arXiv: arXiv:1312.2251.

[GI06] N. Gisin and S. Iblisdir. "Quantum relative states". In: *The European Physical Journal D* 39.2 (2006), pp. 321–327. DOI: 10.1140/epjd/e2006-00097-y. arXiv: quant-ph/0507118 [quant-ph].

[Gio+03] Vittorio Giovannetti, Seth Lloyd, Lorenzo Maccone, and Peter W. Shor. "Entanglement Assisted Capacity of the Broadband Lossy Channel". In: *Physical Review Letters* 91.4 (2003), p. 047901. DOI: 10.1103/PhysRevLett.91.047901. arXiv: quant-ph/0304020 [quant-ph].

[Gio+04] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen. "Classical Capacity of the Lossy Bosonic Channel: The Exact Solution". In: *Physical Review Letters* 92.2 (2004), p. 027902. DOI: 10.1103/PhysRevLett.92.027902. arXiv: quant-ph/0308012 [quant-ph].

[Gio+14] V. Giovannetti, R. García-Patrón, N. J. Cerf, and A. S. Holevo. "Ultimate classical communication rates of quantum optical channels". In: *Nature Photonics* 8.10 (2014), pp. 796–800. DOI: 10.1038/nphoton.2014.216.

[Git] *quantum-learning-machines, https://github.com/fanizzamarco/quantum-learning-machines.*

[GJL18] Li Gao, Marius Junge, and Nicholas LaRacuente. "Capacity bounds via operator space methods". In: *Journal of Mathematical Physics* 59.12 (2018), p. 122202. DOI: 10.1063/1.5058692. arXiv: arXiv:1509.07294.

[GK10] Mădălin Guta and Wojciech Kotłowski. "Quantum learning: asymptotically optimal classification of qubit states". In: *New Journal of Physics* 12.12 (2010), p. 123032. DOI: 10.1088/1367-2630/12/12/123032. arXiv: 1004.2468.

[GL20] András Gilyén and Tongyang Li. "Distributional property testing in a quantum world". In: *Leibniz International Proceedings in Informatics, LIPIcs* 151 (2020). DOI: 10.4230/LIPIcs.ITCS.2020.25. arXiv: 1902.00814.

[Gla63] Roy J. Glauber. "Coherent and incoherent states of the radiation field". In: *Physical Review* 131.6 (1963), pp. 2766–2788. DOI: 10.1103/PhysRev.131.2766.

[GM90] J. P. Gordon and L. F. Mollenauer. "Phase noise in photonic communications systems using linear amplifiers". In: *Optics Letters* 15.23 (1990), p. 1351. DOI: 10.1364/ol.15.001351.

[GNW21] David Gross, Sepehr Nezami, and Michael Walter. "Schur–Weyl Duality for the Clifford Group with Applications: Property Testing, a Robust Hudson Theorem, and de Finetti Representations". In: *Communications in Mathematical Physics* 385.3 (2021), pp. 1325–1393. DOI: 10.1007/s00220-021-04118-7. arXiv: 1712.08628.

[Gol17a] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. DOI: 10.1017/9781108135252.

[Gol17b] Oded Goldreich. "Testing Properties of Distributions". In: *Introduction to Property Testing* 9.1 (2017), pp. 304–347. DOI: 10.1017/9781108135252.013.

[Got10] Daniel Gottesman. "An introduction to quantum error correction and fault-tolerant quantum computation". In: 2010, pp. 13–58. DOI: 10.1090/psapm/068/2762145. arXiv: 0904.2557.

[GP+09] Raúl García-Patrón, Stefano Pirandola, Seth Lloyd, and Jeffrey H. Shapiro. "Reverse coherent information". In: *Physical Review Letters* 102.21 (2009). DOI: 10.1103/PhysRevLett.102.210501. arXiv: arXiv:0808.0210.

[GR11] Oded Goldreich and Dana Ron. "On Testing Expansion in Bounded-Degree Graphs". In: *Lecture Notes in Computer Science (including sub-*

series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2011, pp. 68–75. DOI: 10.1007/978-3-642-22670-0_9.

[Gro06] D. Gross. "Hudson's theorem for finite-dimensional quantum systems". In: *Journal of Mathematical Physics* 47.12 (2006), p. 122107. DOI: 10.1063/1.2393152. arXiv: quant-ph/0602001 [arXiv:quant-ph].

[Gro+10] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. "Quantum State Tomography via Compressed Sensing". In: *Physical Review Letters* 105.15 (2010), p. 150401. DOI: 10.1103/PhysRevLett.105.150401. arXiv: 0909.3304.

[Gro96] Lov K. Grover. "A fast quantum mechanical algorithm for database search". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC '96*. Vol. Part F1294. New York, New York, USA: ACM Press, 1996, pp. 212–219. DOI: 10.1145/237814.237866. arXiv: quant-ph/9605043 [quant-ph].

[GS08] Gilad Gour and Robert W. Spekkens. "The resource theory of quantum reference frames: manipulations and monotones". In: *New Journal of Physics* 10.3 (2008), p. 033023. DOI: 10.1088/1367-2630/10/3/033023. arXiv: 0711.0043.

[Gut+18] Madalin Guta, Jonas Kahn, Richard Kueng, and Joel A Tropp. "Fast state tomography with optimal error bounds". In: *Journal of Physics A: Mathematical and Theoretical* 53.20 (2018), p. 204001. DOI: 10.1088/1751-8121/ab8111. arXiv: 1809.11162.

[GW09] Roe Goodman and Nolan R. Wallach. *Symmetry, Representations, and Invariants*. Vol. 255. Graduate Texts in Mathematics. New York, NY: Springer New York, 2009. DOI: 10.1007/978-0-387-79852-3.

[Haa+17] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. "Sample-optimal tomography of quantum states". In: *IEEE Transactions on Information Theory* 63.9 (2017), pp. 1–1. DOI: 10.1109/TIT.2017.2719044. arXiv: 1508.01797.

[Hal15] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations*. Vol. 222. Graduate Texts in Mathematics. Cham: Springer International Publishing, 2015. DOI: 10.1007/978-3-319-13467-3.

[Har05a] Aram W. Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. 2005. arXiv: quant-ph/0512255 [quant-ph].

[Har05b] Aram W. Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. 2005. arXiv: quant-ph/0512255 [quant-ph].

[Har+10] Aram W. Harrow, Avinatan Hassidim, Debbie W. Leung, and John Watrous. "Adaptive versus nonadaptive strategies for quantum channel discrimination". In: *Physical Review A* 81.3 (2010), p. 032339. DOI: 10.1103/PhysRevA.81.032339. arXiv: 0909.0256.

[Has09] M. B. Hastings. "Superadditivity of communication capacity using entangled inputs". In: *Nature Physics* 5.4 (2009), pp. 255–257. DOI: 10.1038/nphys1224. arXiv: 0809.3972.

[Hav+19] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, et al. *Supervised learning with quantum-enhanced feature spaces*. 2019. DOI: 10.1038/s41586-019-0980-2. arXiv: 1804.11326.

[Hay02] Masahito Hayashi. "Optimal sequence of quantum measurements in the sense of Stein's lemma in quantum hypothesis testing". In: *Journal of Physics A: Mathematical and General* 35.50 (2002), pp. 10759–10773. DOI: 10.1088/0305-4470/35/50/307. arXiv: quant-ph/0208020 [quant-ph].

[Hay07] Masahito Hayashi. "Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding". In: *Physical Review A* 76.6 (2007), p. 062301. DOI: 10.1103/PhysRevA.76.062301. arXiv: quant-ph/0611013 [quant-ph].

[Hay08] Masahito Hayashi. "Universal coding for classical-quantum channel". In: *Communications in Mathematical Physics* 289.3 (2008), pp. 1087–1098. DOI: 10.1007/s00220-009-0825-1. arXiv: 0805.4092.

[Hay+08] Patrick Hayden, Michał Horodecki, Andreas Winter, and Jon Yard. "A Decoupling Approach to the Quantum Capacity". In: *Open Systems & Information Dynamics* 15.01 (2008), pp. 7–19. DOI: 10.1142/S1230161208000043. arXiv: quant-ph/0702005 [quant-ph].

[Hay17a] Masahito Hayashi. *A Group Theoretic Approach to Quantum Information*. Cham: Springer International Publishing, 2017. DOI: 10.1007/978-3-319-45241-8.

[Hay17b] Masahito Hayashi. *Group Representation for Quantum Theory*. Cham: Springer International Publishing, 2017. DOI: 10.1007/978-3-319-44906-7.

[Hay17c] Masahito Hayashi. *Quantum Information Theory*. Ed. by Springer. Graduate Texts in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017. DOI: 10.1007/978-3-662-49725-8.

[Hay97] Masahito Hayashi. "Asymptotic estimation theory for a finite dimensional pure state model". In: *Journal of Physics A: Mathematical and General* 31.20 (1997), pp. 4633–4655. DOI: 10.1088/0305-4470/31/20/006. arXiv: quant-ph/9704041 [quant-ph].

[HB07] Bing He and János A. Bergou. "Programmable unknown quantum-state discriminators with multiple copies of program and data: A Jordan-basis approach". In: *Physical Review A* 75.3 (2007), p. 032316. DOI: 10.1103/PhysRevA.75.032316. arXiv: quant-ph/0610226 [arXiv:quant-ph].

[HB08] Ulrike Herzog and János A. Bergou. "Optimum unambiguous identification of d unknown pure qudit states". In: *Physical Review A* 78.3 (2008), p. 032320. DOI: 10.1103/PhysRevA.78.032320. arXiv: 0809.4884.

[Hel69] Carl W. Helstrom. "Quantum detection and estimation theory". In: *Journal of Statistical Physics* 1.2 (1969), pp. 231–252. DOI: 10.1007/BF01007479.

[HG12] A. S. Holevo and V. Giovannetti. "Quantum channels and their entropic characteristics". In: *Reports on Progress in Physics* 75.4 (2012), p. 046001. DOI: 10.1088/0034-4885/75/4/046001.

[HHH05a] A. Hayashi, M. Horibe, and T. Hashimoto. "Quantum pure-state identification". In: *Physical Review A* 72.5 (2005), p. 052306. DOI: 10.1103/PhysRevA.72.052306. arXiv: quant-ph/0507237 [arXiv:quant-ph].

[HHH05b] A. Hayashi, M. Horibe, and T. Hashimoto. "Unambiguous pure state identification without classical knowledge". In: *Physical Review A - Atomic, Molecular, and Optical Physics* 73.1 (2005). DOI: 10.1103/PhysRevA.73.012328. arXiv: quant-ph/0510015 [quant-ph].

[HHH98] Pawel Horodecki, Michal Horodecki, and Ryszard Horodecki. "General teleportation channel, singlet fraction and quasi-distillation". In: *Physical Review A - Atomic, Molecular, and Optical Physics* 60.3 (1998), pp. 1888–1898. arXiv: quant-ph/9807091 [quant-ph].

[Hil+10] Mark Hillery, Erika Andersson, Stephen M. Barnett, and Daniel Oi. "Decision problems with quantum black boxes". In: *Journal of Modern Optics* 57.3 (2010), pp. 244–252. DOI: 10.1080/09500340903203129. arXiv: 1109.4823.

[HJ12] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". In: *Nature Physics* 16.10 (2020), pp. 1050–1057. DOI: 10.1038/s41567-020-0932-7. arXiv: 2002.08953.

[HM02] Masahito Hayashi and Keiji Matsumoto. "Quantum universal variable-length source coding". In: *Physical Review A* 66.2 (2002), p. 022311. DOI: 10.1103/PhysRevA.66.022311. arXiv: quant-ph/0202001 [quant-ph].

[HM10] Aram W. Harrow and Ashley Montanaro. "An Efficient Test for Product States with Applications to Quantum Merlin-Arthur Games". In: *2010*

IEEE 51st Annual Symposium on Foundations of Computer Science 60.1 (2010), pp. 633–642. DOI: 10.1109/FOCS.2010.66. arXiv: 1001.0017.

[Hol08] A. S. Holevo. "Entanglement-breaking channels in infinite dimensions". In: *Problems of Information Transmission* 44.3 (2008), pp. 171–184. DOI: 10.1134/S0032946008030010. arXiv: 0802.0235.

[Hol11a] A. S. Holevo. "The Choi–Jamiolkowski forms of quantum Gaussian channels". In: *Journal of Mathematical Physics* 52.4 (2011), p. 042202. DOI: 10.1063/1.3581879. arXiv: 1004.0196.

[Hol11b] Alexander Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. Pisa: Edizioni della Normale, 2011. DOI: 10.1007/978-88-7642-378-9.

[Hol19] Alexander S. Holevo. *Quantum Systems, Channels, Information*. De Gruyter, 2019. DOI: 10.1515/9783110642490.

[Hol73] A. S. Holevo. "Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel". In: *Probl. Peredachi Inf.* 9.3 (1973).

[Hol98] A. S. Holevo. "The capacity of the quantum channel with general signal states". In: *IEEE Transactions on Information Theory* 44.1 (1998), pp. 269–273. DOI: 10.1109/18.651037. arXiv: quant-ph/9611023 [quant-ph].

[HP91] Fumio Hiai and Dénes Petz. "The proper formula for relative entropy and its asymptotics in quantum probability". In: *Communications in Mathematical Physics* 143.1 (1991), pp. 99–114. DOI: 10.1007/BF02100287.

[HT16] Masahito Hayashi and Marco Tomamichel. "Correlation detection and an operational interpretation of the Rényi mutual information". In: *Journal of Mathematical Physics* 57.10 (2016), p. 102201. DOI: 10.1063/1.4964755. arXiv: 1408.6894.

[HV01] L. Henderson and V. Vedral. "Classical, quantum and total correlations". In: *Journal of Physics A: Mathematical and General* 34.35 (2001), pp. 6899–6905. DOI: 10.1088/0305-4470/34/35/315. arXiv: quant-ph/0105028 [quant-ph].

[HW01] A. Holevo and R. Werner. "Evaluating capacities of bosonic Gaussian channels". In: *Physical Review A* 63.3 (2001), p. 032312. DOI: 10.1103/PhysRevA.63.032312. arXiv: quant-ph/9912067 [quant-ph].

[HW08] Patrick Hayden and Andreas Winter. "Counterexamples to the Maximal p-Norm Multiplicativity Conjecture for all $p > 1$". In: *Communications in Mathematical Physics* 284.1 (2008), pp. 263–280. DOI: 10.1007/s00220-008-0624-0. arXiv: 0807.4753.

[HW94] Paul Hausladen and William K Wootters. "A 'Pretty Good' Measurement for Distinguishing Quantum States". In: *Journal of Modern Optics* 41.12 (1994), pp. 2385–2390. DOI: 10.1080/09500349414552221.

[Ibm] *IBM Quantum*.

[Iva87] I. D. Ivanovic. "How to differentiate between non-orthogonal states". In: *Physics Letters A* 123.6 (1987), pp. 257–259. DOI: 10.1016/0375-9601(87)90222-2.

[Jar+16] Marcin Jarzyna, Victoria Lipińska, Aleksandra Klimek, Konrad Banaszek, and Matteo G. A. Paris. "Phase noise in collective binary phase shift keying with Hadamard words". In: *Optics Express* 24.2 (2016), p. 1693. DOI: 10.1364/OE.24.001693. arXiv: 1509.00009.

[JBDD14] Marcin Jarzyna, Konrad Banaszek, and Rafał Demkowicz-Dobrzański. "Dephasing in coherent communication with weak signal states". In: *Journal of Physics A: Mathematical and Theoretical* 47.27 (2014), p. 275302. DOI: 10.1088/1751-8113/47/27/275302. arXiv: arXiv:1307.6871.

[JS95] Gregg Jaeger and Abner Shimony. "Optimal distinction between two non-orthogonal quantum states". In: *Physics Letters A* 197.2 (1995), pp. 83–87. DOI: 10.1016/0375-9601(94)00919-G.

[KDK17] Niraj Kumar, Eleni Diamanti, and Iordanis Kerenidis. "Efficient quantum communications with coherent state fingerprints over multiple channels". In: *Physical Review A* 95.3 (2017), p. 032337. DOI: 10.1103/PhysRevA.95.032337. arXiv: arXiv:1609.09600.

[Key06] M. Keyl. "Quantum state estimation and large deviations". In: *Reviews in Mathematical Physics* 18.1 (2006), pp. 19–60. DOI: 10.1142/S0129055X06002565.

[KFG19] Farzad Kianvash, Marco Fanizza, and Vittorio Giovannetti. "Optimal quantum subtracting machine". In: *Physical Review A* 99.5 (2019), p. 052319. DOI: 10.1103/PhysRevA.99.052319. arXiv: 1811.07187.

[KFG20] Farzad Kianvash, Marco Fanizza, and Vittorio Giovannetti. *Bounding the quantum capacity with flagged extensions*. 2020. arXiv: 2008.02461.

[KG08] Jonas Kahn and Madalin Guta. "Local asymptotic normality for finite dimensional quantum systems". In: *Communications in Mathematical Physics* 289.2 (2008), pp. 597–652. DOI: 10.1007/s00220-009-0787-3. arXiv: 0804.3876.

[Kha+19] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T. Sornborger, and Patrick J. Coles. "Quantum-assisted quantum compiling". In: *Quantum* 3 (2019), p. 140. DOI: 10.22331/q-2019-05-13-140. arXiv: 1807.00800.

[Kin02] Christopher King. "Additivity for unital qubit channels". In: *Journal of Mathematical Physics* 43.10 (2002), p. 4641. DOI: 10.1063/1.1500791. arXiv: quant-ph/0103156 [quant-ph].

[Kin03] Christopher King. "The capacity of the quantum depolarizing channel". In: *IEEE Transactions on Information Theory* 49.1 (2003), pp. 221–229. DOI: 10.1109/TIT.2002.806153. arXiv: quant-ph/0204172 [quant-ph].

[Kin+05] Christopher King, Keiji Matsumoto, Michael Nathanson, and Mary Beth Ruskai. "Properties of Conjugate Channels with Applications to Additivity and Multiplicativity". In: *Markov Process and Related Fields* 13 (2005), pp. 391–423. arXiv: quant-ph/0509126 [quant-ph].

[Kin06] Christopher King. "An application of the Lieb-Thirring inequality in quantum information theory". In: *XIVth International Congress on Mathematical Physics: Lisbon, 28 July - 2 August 2003*. WORLD SCIENTIFIC, 2006, pp. 486–490. DOI: 10.1142/9789812704016_0047. arXiv: quant-ph/0412046 [quant-ph].

[Kit97] A Yu Kitaev. "Quantum computations: algorithms and error correction". In: *Russian Mathematical Surveys* 52.6 (1997), pp. 1191–1249. DOI: 10.1070/RM1997v052n06ABEH002155.

[Kna86] Anthony W. Knapp. *Representation Theory of Semisimple Groups*. Princeton University Press, 1986. DOI: 10.1515/9781400883974.

[Kor+19] Kamil Korzekwa, Zbigniew Puchała, Marco Tomamichel, and Karol Życzkowski. "Encoding classical information into quantum resources". In: (2019). arXiv: 1911.12373.

[KPB18] Ludwig Kunz, Matteo G. A. Paris, and Konrad Banaszek. "Noisy propagation of coherent states in a lossy Kerr medium". In: *Journal of the Optical Society of America B* 35.2 (2018), p. 214. DOI: 10.1364/JOSAB.35.000214. arXiv: 1707.09196.

[KR21] Martin Kliesch and Ingo Roth. "Theory of Quantum System Certification". In: *PRX Quantum* 2.1 (2021), p. 010201. DOI: 10.1103/PRXQuantum.2.010201. arXiv: 2010.05925.

[Kro19] Hari Krovi. "An efficient high dimensional quantum Schur transform". In: *Quantum* 3 (2019), p. 122. DOI: 10.22331/q-2019-02-14-122. arXiv: 1804.00055.

[KRT17] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. "Low rank matrix recovery from rank one measurements". In: *Applied and Computational Harmonic Analysis* 42.1 (2017), pp. 88–116. DOI: 10.1016/j.acha.2015.07.007. arXiv: 1410.6913.

[KSW20] Sumeet Khatri, Kunal Sharma, and Mark M. Wilde. "Information-theoretic aspects of the generalized amplitude-damping channel". In: *Phys-*

ical Review A 102.1 (2020), p. 012401. DOI: 10.1103/PhysRevA.102.012401. arXiv: 1903.07747.

[KW01] M. Keyl and R. F. Werner. "Estimating the spectrum of a density operator". In: *Physical Review A* 64.5 (2001), p. 052311. DOI: 10.1103/PhysRevA.64.052311. arXiv: quant-ph/0102027 [quant-ph].

[Lam+08] P. W. Lamberti, A. P. Majtey, A. Borras, M. Casas, and A. Plastino. "Metric character of the quantum Jensen-Shannon divergence". In: *Physical Review A* 77.5 (2008), p. 052311. DOI: 10.1103/PhysRevA.77.052311. arXiv: arXiv:0801.1586.

[Lap+08] Amos Lapidoth, Ligong Wang, Jeffrey H. Shapiro, and Vinodh Venkatesan. "The poisson channel at low input powers". In: *IEEE Conv. Electr. Electron. Eng. Isr. Proc.* 2008, pp. 654–658. DOI: 10.1109/EEEI.2008.4736614. arXiv: 0810.3564.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444. DOI: 10.1038/nature14539.

[LDS18] Felix Leditzky, Nilanjana Datta, and Graeme Smith. "Useful States and Entanglement Distillation". In: *IEEE Transactions on Information Theory*. Vol. 64. 7. 2018, pp. 4689–4708. DOI: 10.1109/TIT.2017.2776907. arXiv: 1701.03081.

[Led+18] Felix Leditzky, Eneet Kaur, Nilanjana Datta, and Mark M. Wilde. "Approaches for approximate additivity of the Holevo information of quantum channels". In: *Physical Review A* 97.1 (2018), p. 012332. DOI: 10.1103/PhysRevA.97.012332. arXiv: 1709.01111.

[Li+09] Ke Li, Andreas Winter, Xubo Zou, and Guangcan Guo. "Private Capacity of Quantum Channels is Not Additive". In: *Physical Review Letters* 103.12 (2009), p. 120501. DOI: 10.1103/PhysRevLett.103.120501. arXiv: 0903.4308.

[Li16] Ke Li. "Discriminating quantum states: The multiple Chernoff distance". In: *The Annals of Statistics* 44.4 (2016), pp. 1661–1679. DOI: 10.1214/16-AOS1436. arXiv: 1508.06624.

[Lin75] Göran Lindblad. "Completely positive maps and entropy inequalities". In: *Communications in Mathematical Physics* 40.2 (1975), pp. 147–151. DOI: 10.1007/BF01609396.

[Lin91] Jianhua Lin. "Divergence measures based on the Shannon entropy". In: *IEEE Transactions on Information Theory* 37.1 (1991), pp. 145–151. DOI: 10.1109/18.61115.

[Llo97] Seth Lloyd. "Capacity of the noisy quantum channel". In: *Physical Review A* 55.3 (1997), pp. 1613–1622. DOI: 10.1103/PhysRevA.55.1613. arXiv: quant-ph/9604015 [quant-ph].

[LLS18a] Felix Leditzky, Debbie Leung, and Graeme Smith. "Dephrasure Channel and Superadditivity of Coherent Information". In: *Physical Review Letters* 121.16 (2018), p. 160501. DOI: 10.1103/PhysRevLett.121.160501. arXiv: 1806.08327.

[LLS18b] Felix Leditzky, Debbie Leung, and Graeme Smith. "Quantum and Private Capacities of Low-Noise Channels". In: *Physical Review Letters* 120.16 (2018), p. 160503. DOI: 10.1103/PhysRevLett.120.160503. arXiv: 1705.04335.

[LM09] Amos Lapidoth and Stefan M. Moser. "On the Capacity of the Discrete-Time Poisson Channel". In: *IEEE Transactions on Information Theory* 55.1 (2009), pp. 303–322. DOI: 10.1109/TIT.2008.2008121.

[LMR13] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. "Quantum algorithms for supervised and unsupervised machine learning". In: (2013). arXiv: 1307.0411.

[LR18] Nana Liu and Patrick Rebentrost. "Quantum machine learning for quantum anomaly detection". In: *Physical Review A* 97.4 (2018), p. 042315. DOI: 10.1103/PhysRevA.97.042315. arXiv: 1710.07405.

[LR73a] Elliott H. Lieb and Mary Beth Ruskai. "A fundamental property of quantum-mechanical entropy". In: *Physical Review Letters* 30.10 (1973), pp. 434–436. DOI: 10.1103/PhysRevLett.30.434.

[LR73b] Elliott H. Lieb and Mary Beth Ruskai. "Proof of the strong subadditivity of quantum-mechanical entropy". In: *Journal of Mathematical Physics* 14.12 (1973), pp. 1938–1941. DOI: 10.1063/1.1666274.

[LS09] Debbie Leung and Graeme Smith. "Continuity of Quantum Channel Capacities". In: *Communications in Mathematical Physics* 292.1 (2009), pp. 201–215. DOI: 10.1007/s00220-009-0833-1. arXiv: 0810.4931.

[LSB06] Netanel H. Lindner, Petra F. Scudo, and Dagmar Bruß. "Quantum estimation of relative information". In: *International Journal of Quantum Information* 4.1 (2006), pp. 131–149. DOI: 10.1142/S0219749906001657. arXiv: quant-ph/0506223 [quant-ph].

[Lü21] Xin Lü. "Lower bounds on the failure probability of unambiguous discrimination". In: *Physical Review A* 103.2 (2021), p. 022216. DOI: 10.1103/PhysRevA.103.022216.

[Mar07] Alfonso Martinez. "Spectral efficiency of optical direct detection". In: *Journal of the Optical Society of America B* 24.4 (2007), p. 739. DOI: 10.1364/josab.24.000739.

[Mat10] Keiji Matsumoto. "Reverse Test and Characterization of Quantum Relative Entropy". In: (2010). arXiv: 1010.1030.

[MC19] Yin Mo and Giulio Chiribella. "Quantum-enhanced learning of rotations about an unknown direction". In: *New Journal of Physics* 21.11 (2019), p. 113003. DOI: 10.1088/1367-2630/ab4d9a. arXiv: 1906.01300.

[McC+18] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. "Barren plateaus in quantum neural network training landscapes". In: *Nature Communications* 9.1 (2018), p. 4812. DOI: 10.1038/s41467-018-07090-4. arXiv: 1803.11173.

[MD19] Pierre Martin-Dussaud. "A primer of group theory for Loop Quantum Gravity and spin-foams". In: *General Relativity and Gravitation* 51.9 (2019), p. 110. DOI: 10.1007/s10714-019-2583-5. arXiv: 1902.08439.

[MGH14] Andrea Mari, Vittorio Giovannetti, and A. S. Holevo. "Quantum state majorization at the output of bosonic Gaussian channels". In: *Nature Communications* 5.1 (2014), p. 3826. DOI: 10.1038/ncomms4826. arXiv: 1312.3545.

[MH04] Keiji Matsumoto and Masahito Hayashi. "Universal distortion-free entanglement concentration". In: *IEEE International Symposium on Information Theory - Proceedings* (2004), p. 323. DOI: 10.1109/isit.2004.1365360. arXiv: quant-ph/0209030 [quant-ph].

[MHRW16] Alexander Müller-Hermes, David Reeb, and Michael M. Wolf. "Positivity of linear maps under tensor powers". In: *Journal of Mathematical Physics* 57.1 (2016), p. 015202. DOI: 10.1063/1.4927070. arXiv: 1502.05630.

[MKB05] Florian Mintert, Marek Kuś, and Andreas Buchleitner. "Concurrence of Mixed Multipartite Quantum States". In: *Physical Review Letters* 95.26 (2005), p. 260502. DOI: 10.1103/PhysRevLett.95.260502. arXiv: quant-ph/0411127 [quant-ph].

[MLP05] A. P. Majtey, P. W. Lamberti, and D. P. Prato. "Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states". In: *Physical Review A* 72.5 (2005), p. 052310. DOI: 10.1103/PhysRevA.72.052310. arXiv: quant-ph/0508138 [quant-ph].

[Mon06] Alex Monras. "Optimal phase measurements with pure Gaussian states". In: *Physical Review A* 73.3 (2006), p. 033821. DOI: 10.1103/PhysRevA.73.033821. arXiv: quant-ph/0509018 [quant-ph].

[Mon07] Ashley Montanaro. "On the Distinguishability of Random Quantum States". In: *Communications in Mathematical Physics* 273.3 (2007), pp. 619–636. DOI: 10.1007/s00220-007-0221-7. arXiv: quant-ph/0607011 [quant-ph].

[Mon08] Ashley Montanaro. "A lower bound on the probability of error in quantum state discrimination". In: *2008 IEEE Information Theory Workshop*. IEEE, 2008, pp. 378–380. DOI: 10.1109/ITW.2008.4578690. arXiv: 0711.2012.

[Mon19] Ashley Montanaro. "Pretty simple bounds on quantum state discrimination". In: (2019). arXiv: 1908.08312.

[MS13] Iman Marvian and Robert W. Spekkens. "The theory of manipulations of pure state asymmetry: I. Basic tools, equivalence classes and single copy transformations". In: *New Journal of Physics* 15.3 (2013), p. 033001. DOI: 10.1088/1367-2630/15/3/033001. arXiv: 1104.0018.

[MS14a] Iman Marvian and Robert W. Spekkens. "A Generalization of Schur–Weyl Duality with Applications in Quantum Estimation". In: *Communications in Mathematical Physics* 331.2 (2014), pp. 431–475. DOI: 10.1007/s00220-014-2059-0. arXiv: 1112.0638.

[MS14b] Iman Marvian and Robert W. Spekkens. "Modes of asymmetry: The application of harmonic analysis to symmetric quantum dynamics and quantum reference frames". In: *Physical Review A* 90.6 (2014), p. 062110. DOI: 10.1103/PhysRevA.90.062110. arXiv: 1312.0680.

[MSW16] Alex Monràs, Gael Sentís, and Peter Wittek. "Inductive supervised quantum learning". In: *Physical Review Letters* 118.19 (2016). DOI: 10.1103/PhysRevLett.118.190503. arXiv: 1605.07541.

[MW16] Ashley Montanaro and Ronald de Wolf. "A survey of quantum property testing". In: *Theory of Computing* 1.1 (2016), pp. 1–81. DOI: 10.4086/toc.gs.2016.007. arXiv: 1310.2035.

[Nag06] Hiroshi Nagaoka. *The Converse Part of The Theorem for Quantum Hoeffding Bound*. 2006. arXiv: quant-ph/0611289 [quant-ph].

[NAJ19] Kyungjoo Noh, Victor V. Albert, and Liang Jiang. "Quantum Capacity Bounds of Gaussian Thermal Loss Channels and Achievable Rates with Gottesman-Kitaev-Preskill Codes". In: *IEEE Transactions on Information Theory* 65.4 (2019), pp. 2563–2582. DOI: 10.1109/TIT.2018.2873764. arXiv: 1801.07271.

[NPJ20] Kyungjoo Noh, Stefano Pirandola, and Liang Jiang. "Enhanced energy-constrained quantum communication over bosonic Gaussian channels". In: *Nature Communications* 11.1 (2020), p. 457. DOI: 10.1038/s41467-020-14329-6. arXiv: 1811.06988.

[NS09] Michael Nussbaum and Arleta Szkoła. "The Chernoff lower bound for symmetric quantum hypothesis testing". In: *The Annals of Statistics* 37.2 (2009), pp. 1040–1057. DOI: 10.1214/08-AOS593. arXiv: quant-ph/0607216 [quant-ph].

[NUK18] Kenji Nakahira, Tsuyoshi Sasaki Usuda, and Kentaro Kato. "Upper and lower bounds on optimal success probability of quantum state discrimination with and without inconclusive results". In: *Physical Review A* 97.1

(2018), p. 012103. DOI: 10.1103/PhysRevA.97.012103. arXiv: 1601.06231.

[OH04] Tomohiro Ogawa and Masahito Hayashi. "On Error Exponents in Quantum Hypothesis Testing". In: *IEEE Transactions on Information Theory* 50.6 (2004), pp. 1368–1372. DOI: 10.1109/TIT.2004.828155. arXiv: quant-ph/0206151 [quant-ph].

[Oli+13] Stefano Olivares, Simone Cialdi, Fabrizio Castelli, and Matteo G A Paris. "Homodyne detection as a near-optimum receiver for phase-shift-keyed binary communication in the presence of phase diffusion". In: *Physical Review A* 87.5 (2013), p. 050303. DOI: 10.1103/PhysRevA.87.050303. arXiv: 1305.4201.

[ON05] Tomohiro Ogawa and Hiroshi Nagaoka. "Strong Converse and Stein's Lemma in Quantum Hypothesis Testing". In: *Asymptotic Theory of Quantum Statistical Inference*. WORLD SCIENTIFIC, 2005, pp. 28–42. DOI: 10.1142/9789812563071_0003. arXiv: quant-ph/9906090 [quant-ph].

[Ouy14] Yingkai Ouyang. "Channel covariance, twirling, contraction, and some upper bounds on the quantum capacity". In: *Quantum Information and Computation* 14.11-12 (2014), pp. 917–936. DOI: 10.26421/qic14.11-12-2. arXiv: 1106.2337.

[OW15] Ryan O'Donnell and John Wright. "Quantum Spectrum Testing". In: *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*. Vol. 14-17-June. New York, NY, USA: ACM, 2015, pp. 529–538. DOI: 10.1145/2746539.2746582. arXiv: 1501.05028.

[OW16] Ryan O'Donnell and John Wright. "Efficient quantum tomography". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. Vol. 19-21-June. New York, NY, USA: ACM, 2016, pp. 899–912. DOI: 10.1145/2897518.2897544. arXiv: 1508.01907.

[OW17] Ryan O'Donnell and John Wright. "Efficient quantum tomography II". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. New York, NY, USA: ACM, 2017, pp. 962–974. DOI: 10.1145/3055399.3055454. arXiv: 1612.00034.

[OZ01] Harold Ollivier and Wojciech H Zurek. "Quantum Discord: A Measure of the Quantumness of Correlations". In: *Physical Review Letters* 88.1 (2001), p. 017901. DOI: 10.1103/PhysRevLett.88.017901. arXiv: quant-ph/0105072 [quant-ph].

[Pan08] Liam Paninski. "A Coincidence-Based Test for Uniformity Given Very Sparsely Sampled Discrete Data". In: *IEEE Transactions on Information Theory* 54.10 (2008), pp. 4750–4755. DOI: 10.1109/TIT.2008.928987.

[Par09] Matteo G.A. Paris. "Quantum estimation for quantum technology". In: *International Journal of Quantum Information*. Vol. 7. SUPPL. 2009, pp. 125–137. DOI: 10.1142/S0219749909004839. arXiv: 0804.2981.

[Per71] S. Personick. "Application of quantum estimation theory to analog communication over quantum channels". In: *IEEE Transactions on Information Theory* 17.3 (1971), pp. 240–246. DOI: 10.1109/TIT.1971.1054643.

[Per72] A. M. Perelomov. "Coherent states for arbitrary Lie group". In: *Communications in Mathematical Physics* 26.3 (1972), pp. 222–236. DOI: 10.1007/BF01645091. arXiv: math-ph/0203002 [math-ph].

[Per88] Asher Peres. "How to differentiate between non-orthogonal states". In: *Physics Letters A* 128.1-2 (1988), p. 19. DOI: 10.1016/0375-9601(88)91034-1.

[Pir+17] Stefano Pirandola, Riccardo Laurenza, Carlo Ottaviani, and Leonardo Banchi. "Fundamental limits of repeaterless quantum communications". In: *Nature Communications* 8.1 (2017), p. 15043. DOI: 10.1038/ncomms15043. arXiv: 1510.08863.

[Pir+19] Stefano Pirandola, Riccardo Laurenza, Cosmo Lupo, and Jason L. Pereira. "Fundamental limits to quantum channel discrimination". In: *npj Quantum Information* 5.1 (2019), p. 50. DOI: 10.1038/s41534-019-0162-y. arXiv: 1803.02834.

[Pre18] John Preskill. "Quantum Computing in the NISQ era and beyond". In: *Quantum* 2 (2018), p. 79. DOI: 10.22331/q-2018-08-06-79. arXiv: 1801.00862.

[PW17] Daniel Puzzuoli and John Watrous. "Ancilla Dimension in Quantum Channel Discrimination". In: *Annales Henri Poincaré* 18.4 (2017), pp. 1153–1184. DOI: 10.1007/s00023-016-0537-y. arXiv: 1604.08197.

[Qi+15] Bing Qi, Pavel Lougovski, Raphael Pooser, Warren Grice, and Miljko Bobrek. "Generating the Local Oscillator "Locally" in Continuous-Variable Quantum Key Distribution Based on Coherent Detection". In: *Physical Review X* 5.4 (2015), p. 041009. DOI: 10.1103/PhysRevX.5.041009. arXiv: 1503.00662.

[Qiu08] Daowen Qiu. "Minimum-error discrimination between mixed quantum states". In: *Physical Review A* 77.1 (2008), p. 012328. DOI: 10.1103/PhysRevA.77.012328. arXiv: 0707.3970.

[Rai01] E. M. Rains. "A semidefinite program for distillable entanglement". In: *IEEE Transactions on Information Theory* 47.7 (2001), pp. 2921–2933. DOI: 10.1109/18.959270. arXiv: quant-ph/0008047 [quant-ph].

[RG16] Matteo Rosati and Vittorio Giovannetti. "Achieving the Holevo bound via a bisection decoding protocol". In: *Journal of Mathematical Physics* 57.6 (2016), p. 62204. DOI: 10.1063/1.4953690. arXiv: 1506.04999.

[RMG16] Matteo Rosati, Andrea Mari, and Vittorio Giovannetti. "Multiphase Hadamard receivers for classical communication on lossy bosonic channels". In: *Physical Review A* 94.6 (2016), p. 062325. DOI: 10.1103/PhysRevA.94.062325. arXiv: 1610.05676.

[RMG17] Matteo Rosati, Andrea Mari, and Vittorio Giovannetti. "Capacity of coherent-state adaptive decoders with interferometry and single-mode detectors". In: *Physical Review A* 96.1 (2017), p. 012317. DOI: 10.1103/PhysRevA.96.012317. arXiv: 1703.05701.

[RMG18] Matteo Rosati, Andrea Mari, and Vittorio Giovannetti. "Narrow bounds for the quantum capacity of thermal attenuators". In: *Nature Communications* 9.1 (2018), p. 4339. DOI: 10.1038/s41467-018-06848-0. arXiv: 1801.04731.

[RML14] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. "Quantum Support Vector Machine for Big Data Classification". In: *Physical Review Letters* 113.13 (2014), p. 130503. DOI: 10.1103/PhysRevLett.113.130503. arXiv: 1307.0471.

[Roa+11] Luis Roa, Carla Hermann-Avigliano, R. Salazar, and A. B. Klimov. "Conclusive discrimination among N equidistant pure states". In: *Physical Review A* 84.1 (2011), p. 014302. DOI: 10.1103/PhysRevA.84.014302.

[RST03] Terry Rudolph, Robert W. Spekkens, and Peter Shipley Turner. "Unambiguous discrimination of mixed states". In: *Physical Review A* 68.1 (2003), p. 010301. DOI: 10.1103/PhysRevA.68.010301. arXiv: quant-ph/0303071 [quant-ph].

[ŠAF15] Dominik Šafránek, Mehdi Ahmadi, and Ivette Fuentes. "Quantum parameter estimation with imperfect reference frames". In: *New Journal of Physics* 17.3 (2015), p. 033012. DOI: 10.1088/1367-2630/17/3/033012. arXiv: 1404.6421.

[Sag01] Bruce E. Sagan. *The Symmetric Group*. Vol. 203. Graduate Texts in Mathematics. New York, NY: Springer New York, 2001. DOI: 10.1007/978-1-4757-6804-6.

[SBD16] Magdalena Szczykulska, Tillmann Baumgratz, and Animesh Datta. "Multi-parameter quantum metrology". In: *Advances in Physics: X* 1.4 (2016), pp. 621–639. DOI: 10.1080/23746149.2016.1230476. arXiv: 1604.02615.

[SBZ19] Michal Sedlák, Alessandro Bisio, and Mário Ziman. "Optimal Probabilistic Storage and Retrieval of Unitary Channels". In: *Physical Review Letters*

122.17 (2019), p. 170502. DOI: 10.1103/PhysRevLett.122.170502. arXiv: 1809.04552.

[SC02] Masahide Sasaki and Alberto Carlini. "Quantum learning and universal quantum matching machine". In: *Physical Review A* 66.2 (2002), p. 022303. DOI: 10.1103/PhysRevA.66.022303. arXiv: quant-ph/0202173 [quant-ph].

[Sch95] Benjamin Schumacher. "Quantum coding". In: *Physical Review A* 51.4 (1995), pp. 2738–2747. DOI: 10.1103/PhysRevA.51.2738.

[SCJ01] Masahide Sasaki, Alberto Carlini, and Richard Jozsa. "Quantum template matching". In: *Physical Review A* 64.2 (2001), p. 022317. DOI: 10.1103/PhysRevA.64.022317. arXiv: quant-ph/0102020 [quant-ph].

[SCMT17] Gael Sentís, John Calsamiglia, and Ramon Muñoz-Tapia. "Exact Identification of a Quantum Change Point". In: *Physical Review Letters* 119.14 (2017). DOI: 10.1103/PhysRevLett.119.140506. arXiv: 1707.07769.

[Sed+07] Michal Sedlák, Mário Ziman, Ondřej Přibyla, Vladimír Bužek, and Mark Hillery. "Unambiguous identification of coherent states: Searching a quantum database". In: *Physical Review A* 76.2 (2007), p. 022326. DOI: 10.1103/PhysRevA.76.022326.

[Sed+09] Michal Sedlák, Mário Ziman, Vladimír Bužek, and Mark Hillery. "Unambiguous identification of coherent states. II. Multiple resources". In: *Physical Review A* 79.6 (2009), p. 062305. DOI: 10.1103/PhysRevA.79.062305. arXiv: 0901.3206.

[Sen+10] G. Sentís, E. Bagan, J. Calsamiglia, and R. Muñoz-Tapia. "Multicopy programmable discrimination of general qubit states". In: *Physical Review A* 82.4 (2010), p. 042312. DOI: 10.1103/PhysRevA.82.042312. arXiv: arXiv:1007.5497.

[Sen+12] G. Sentís, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan. "Quantum learning without quantum memory". In: *Scientific Reports* 2 (2012). DOI: 10.1038/srep00708. arXiv: 1106.2742.

[Sen+13] G. Sentís, E. Bagan, J. Calsamiglia, and R. Muñoz-Tapia. "Programmable discrimination with an error margin". In: *Physical Review A - Atomic, Molecular, and Optical Physics* 88.5 (2013). DOI: 10.1103/PhysRevA.88.052304. arXiv: 1308.1378.

[Sen+16] Gael Sentís, Emilio Bagan, John Calsamiglia, Giulio Chiribella, and Ramon Muñoz-Tapia. "Quantum change point". In: *Physical Review Letters* 117.15 (2016). DOI: 10.1103/PhysRevLett.117.150502. arXiv: 1605.01916.

[Sen+19] Gael Sentís, Alex Monràs, Ramon Muñoz-Tapia, John Calsamiglia, and Emilio Bagan. "Unsupervised Classification of Quantum Data". In: *Phys-*

ical Review X 9.4 (2019), p. 041029. DOI: 10.1103/PhysRevX.9.041029. arXiv: 1903.01391.

[Ser17] Alessio Serafini. *Quantum continuous variables: A primer of theoretical methods*. Boca Raton: CRC Press, 2017, pp. 1–349. DOI: 10.1201/9781315118727.

[SG21] Vikesh Siddhu and Robert B. Griffiths. "Positivity and Nonadditivity of Quantum Capacities Using Generalized Erasure Channels". In: *IEEE Transactions on Information Theory* 67.7 (2021), pp. 4533–4545. DOI: 10.1109/TIT.2021.3080819. arXiv: 2003.00583.

[Sha+18] Kunal Sharma, Mark M. Wilde, Sushovit Adhikari, and Masahiro Takeoka. "Bounding the energy-constrained quantum and private capacities of phase-insensitive bosonic Gaussian channels". In: *New Journal of Physics* 20.6 (2018), p. 063025. DOI: 10.1088/1367-2630/aac11a. arXiv: 1708.07257.

[Sha48] C. E. Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27.4 (1948), pp. 623–656. DOI: 10.1002/j.1538-7305.1948.tb00917.x.

[Shi17] M. E. Shirokov. "Tight uniform continuity bounds for the quantum conditional mutual information, for the Holevo quantity, and for capacities of quantum channels". In: *Journal of Mathematical Physics* 58.10 (2017), p. 102202. DOI: 10.1063/1.4987135. arXiv: 1512.09047.

[Sho02a] Peter W. Shor. "Additivity of the classical capacity of entanglement-breaking quantum channels". In: *Journal of Mathematical Physics* 43.9 (2002), pp. 4334–4340. DOI: 10.1063/1.1498000. arXiv: quant-ph/0201149 [quant-ph].

[Sho02b] Peter W. Shor. "The quantum channel capacity and coherent information". In: *lecture notes, MSRI Workshop on Quantum Computation (Quantum Information and Cryptography)*. 2002. DOI: www.msri.org/workshops/203/schedules/1181.

[Sho95] Peter W. Shor. "Scheme for reducing decoherence in quantum computer memory". In: *Physical Review A* 52.4 (1995), R2493–R2496. DOI: 10.1103/PhysRevA.52.R2493.

[Sho96] P.W. Shor. "Fault-tolerant quantum computation". In: *Proceedings of 37th Conference on Foundations of Computer Science*. IEEE Comput. Soc. Press, 1996, pp. 56–65. DOI: 10.1109/SFCS.1996.548464. arXiv: quant-ph/9605011 [quant-ph].

[Sho97] Peter W. Shor. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer". In: *SIAM*

Journal on Computing 26.5 (1997), pp. 1484–1509. DOI: 10.1137/S0097539795293172. arXiv: quant-ph/9508027 [quant-ph].

[SHW20] Farzin Salek, Masahito Hayashi, and Andreas Winter. *When are Adaptive Strategies in Asymptotic Quantum Channel Discrimination Useful?* 2020. arXiv: 2011.06569.

[Sid20a] Vikesh Siddhu. *Leaking information to gain entanglement.* 2020. arXiv: 2011.15116.

[Sid20b] Vikesh Siddhu. *Log-singularities for studying capacities of quantum channels.* 2020. arXiv: 2003.10367.

[Sin+14] L. C. Sinclair, F. R. Giorgetta, W. C. Swann, E. Baumann, I. Coddington, and N. R. Newbury. "Optical phase noise from atmospheric fluctuations and its impact on optical time-frequency transfer". In: *Physical Review A* 89.2 (2014), p. 023805. DOI: 10.1103/PhysRevA.89.023805.

[Smi08] Graeme Smith. "Private classical capacity with a symmetric side channel and its application to quantum cryptography". In: *Physical Review A* 78.2 (2008), p. 022306. DOI: 10.1103/PhysRevA.78.022306. arXiv: 0705.3838.

[SMVMT18] Gael Sentís, Esteban Martínez-Vargas, and Ramon Muñoz-Tapia. "Online strategies for exactly identifying a quantum change point". In: *Physical Review A* 98.5 (2018), p. 052305. DOI: 10.1103/PhysRevA.98.052305. arXiv: 1802.00280.

[SOP15] Carlo Sparaciari, Stefano Olivares, and Matteo G. A. Paris. "Bounds to precision for quantum interferometry with Gaussian states and operations". In: *Journal of the Optical Society of America B* 32.7 (2015), p. 1354. DOI: 10.1364/JOSAB.32.001354.

[Sra21] Suvrit Sra. "Metrics induced by Jensen-Shannon and related divergences on positive definite matrices". In: *Linear Algebra and its Applications* 616 (2021), pp. 125–138. DOI: 10.1016/j.laa.2020.12.023. arXiv: 1911.02643.

[SRS08] Graeme Smith, Joseph M. Renes, and John A. Smolin. "Structured Codes Improve the Bennett-Brassard-84 Quantum Key Rate". In: *Physical Review Letters* 100.17 (2008), p. 170502. DOI: 10.1103/PhysRevLett.100.170502. arXiv: quant-ph/0607018 [quant-ph].

[SS07] Graeme Smith and John A. Smolin. "Degenerate Quantum Codes for Pauli Channels". In: *Physical Review Letters* 98.3 (2007), p. 030501. DOI: 10.1103/PhysRevLett.98.030501. arXiv: quant-ph/0604107 [quant-ph].

[SS08] Graeme Smith and John A. Smolin. "Additive extensions of a quantum channel". In: *2008 IEEE Information Theory Workshop*. IEEE, 2008, pp. 368–372. DOI: 10.1109/ITW.2008.4578688. arXiv: 0712.2471.

[SS91] Jeffrey H. Shapiro and Scott R. Shepard. "Quantum phase measurement: A system-theory perspective". In: *Physical Review A* 43.7 (1991), pp. 3795–3818. DOI: 10.1103/PhysRevA.43.3795.

[SS96] Peter W. Shor and John A. Smolin. *Quantum Error-Correcting Codes Need Not Completely Reveal the Error Syndrome*. 1996. arXiv: quant-ph/9604006 [quant-ph].

[SSBD13] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Vol. 9781107057. Cambridge University Press, 2013, pp. 1–397. DOI: 10.1017/CBO9781107298019.

[SSM18] Atsushi Shimbo, Akihito Soeda, and Mio Murao. *Equivalence determination of unitary operations*. 2018. arXiv: 1803.11414.

[SSM21] Akihito Soeda, Atsushi Shimbo, and Mio Murao. *Comparison protocol is optimal for quantum discrimination of single-qubit unitary gates–two candidates, one quantum sample per candidate*. 2021. arXiv: 2103.08208.

[SSP14] M. Schuld, I. Sinayskiy, and F. Petruccione. "An introduction to quantum machine learning". In: *Contemporary Physics* 56.2 (2014), pp. 172–185. DOI: 10.1080/00107514.2014.964942. arXiv: 1409.3097.

[SSW08] Graeme Smith, John A. Smolin, and Andreas Winter. "The quantum capacity with symmetric side channels". In: *IEEE Transactions on Information Theory* 54.9 (2008), pp. 4208–4217. DOI: 10.1109/TIT.2008.928269. arXiv: quant-ph/0607039 [quant-ph].

[SSY11] Graeme Smith, John A. Smolin, and Jon Yard. "Quantum communication with Gaussian channels of zero quantum capacity". In: *Nature Photonics* 5.10 (2011), pp. 624–627. DOI: 10.1038/nphoton.2011.203. arXiv: 1102.4580.

[ST87] B. E. A. Saleh and M. C. Teich. "Can the channel capacity of a light-wave communication system be increased by the use of photon-number – squeezed light?" In: *Physical Review Letters* 58.25 (1987), pp. 2656–2659. DOI: 10.1103/PhysRevLett.58.2656.

[Sut+17] David Sutter, Volkher B. Scholz, Andreas Winter, and Renato Renner. "Approximate Degradable Quantum Channels". In: *IEEE Transactions on Information Theory* 63.12 (2017), pp. 7832–7844. DOI: 10.1109/TIT.2017.2754268. arXiv: 1412.0980.

[SV16] Igal Sason and Sergio Verdu. "f -Divergence Inequalities". In: *IEEE Transactions on Information Theory* 62.11 (2016), pp. 5973–6006. DOI: 10.1109/TIT.2016.2603151. arXiv: 1508.00335.

[SW97] Benjamin Schumacher and Michael D. Westmoreland. "Sending classical information via noisy quantum channels". In: *Physical Review A* 56.1

(1997), pp. 131–138. DOI: 10.1103/PhysRevA.56.131. arXiv: quant-ph/9604023 [quant-ph].

[SY08] Graeme Smith and Jon Yard. "Quantum Communication with Zero-Capacity Channels". In: *Science* 321.5897 (2008), pp. 1812–1815. DOI: 10.1126/science.1162242. arXiv: 0807.4935.

[SYH20] Jun Suzuki, Yuxiang Yang, and Masahito Hayashi. "Quantum state estimation with nuisance parameters". In: *Journal of Physics A: Mathematical and Theoretical* 53.45 (2020), p. 453001. DOI: 10.1088/1751-8121/ab8b78. arXiv: 1911.02790.

[Sze59] Gábor Szego. "Orthogonal polynomials". In: *Amer. Math. Soc. Colloquium, 1959*. 1959.

[TGW14] Masahiro Takeoka, Saikat Guha, and Mark M. Wilde. "The Squashed Entanglement of a Quantum Channel". In: *IEEE Transactions on Information Theory* 60.8 (2014), pp. 4987–4998. DOI: 10.1109/TIT.2014.2330313. arXiv: 1310.0129.

[TWW17] Marco Tomamichel, Mark M. Wilde, and Andreas Winter. "Strong Converse Rates for Quantum Communication". In: *IEEE Transactions on Information Theory* 63.1 (2017), pp. 715–727. DOI: 10.1109/TIT.2016.2615847. arXiv: 1406.2946.

[Vah+16] Henning Vahlbruch, Moritz Mehmet, Karsten Danzmann, and Roman Schnabel. "Detection of 15 dB Squeezed States of Light and their Application for the Absolute Calibration of Photoelectric Quantum Efficiency". In: *Physical Review Letters* 117.11 (2016), p. 110801. DOI: 10.1103/PhysRevLett.117.110801.

[Vap98] Vladimir Vapnik. *Statistical Learning Theory*. Ed. by Wiley. 1998.

[Vir21] Dániel Virosztek. "The metric property of the quantum Jensen-Shannon divergence". In: *Advances in Mathematics* 380 (2021), p. 107595. DOI: 10.1016/j.aim.2021.107595. arXiv: 1910.10447.

[VMK88] D A Varshalovich, A N Moskalev, and V K Khersonskii. *Quantum Theory of Angular Momentum*. WORLD SCIENTIFIC, 1988. DOI: 10.1142/0270.

[VR94] A. Vourdas and J. R. F. da Rocha. "Pulse Position Modulation and Extended Pulse Position Modulation with Squeezed Light". In: *Journal of Modern Optics* 41.12 (1994), pp. 2291–2299. DOI: 10.1080/09500349414552141.

[VV14] Gregory Valiant and Paul Valiant. "An Automatic Inequality Prover and Instance Optimal Identity Testing". In: *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*. IEEE, 2014, pp. 51–60. DOI: 10.1109/FOCS.2014.14.

[Wal+07] S. P. Walborn, P. H. Souto Ribeiro, L. Davidovich, F. Mintert, and A. Buchleitner. "Experimental determination of entanglement by a projective measurement". In: *Physical Review A* 75.3 (2007), p. 032338. DOI: 10.1103/PhysRevA.75.032338.

[Wan21] Xin Wang. "Pursuing the fundamental limits for quantum communication". In: *IEEE Transactions on Information Theory* 67.7 (2021), pp. 4524–4532. DOI: 10.1109/TIT.2021.3068818. arXiv: 1912.00931.

[Wan92] K H Wanser. "Fundamental phase noise limit in optical fibres due to temperature fluctuations". In: *Electron. Lett.* 28.1 (1992), pp. 53–54. DOI: 10.1049/el:19920033.

[Wat09] John Watrous. *Semidefinite programs for completely bounded norms*. 2009. arXiv: 0901.4709.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: 10.1017/9781316848142.

[Wee+12] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, et al. "Gaussian quantum information". In: *Reviews of Modern Physics* 84.2 (2012), pp. 621–669. DOI: 10.1103/RevModPhys.84.621. arXiv: 1110.3234.

[WEH18] Stephanie Wehner, David Elkouss, and Ronald Hanson. "Quantum internet: A vision for the road ahead". In: *Science* 362.6412 (2018), eaam9288. DOI: 10.1126/science.aam9288.

[Wey47] Hermann Weyl. *The Classical Groups*. Princeton University Press, 1947. DOI: 10.1515/9781400883905.

[WFD19] Xin Wang, Kun Fang, and Runyao Duan. "Semidefinite Programming Converse Bounds for Quantum Communication". In: *IEEE Transactions on Information Theory* 65.4 (2019), pp. 2583–2592. DOI: 10.1109/TIT.2018.2874031. arXiv: 1709.00200.

[Wil17] Mark M. Wilde. *Quantum Information Theory*. Cambridge: Cambridge University Press, 2017, pp. 1–757. DOI: 10.1017/9781316809976.

[Wil+20] Mark M. Wilde, Mario Berta, Christoph Hirche, and Eneet Kaur. "Amortized channel divergence for asymptotic quantum channel discrimination". In: *Letters in Mathematical Physics* 110.8 (2020), pp. 2277–2336. DOI: 10.1007/s11005-020-01297-7. arXiv: 1808.01498.

[Win16] Andreas Winter. "Tight Uniform Continuity Bounds for Quantum Entropies: Conditional Entropy, Relative Entropy Distance and Energy Constraints". In: *Communications in Mathematical Physics* 347.1 (2016), pp. 291–313. DOI: 10.1007/s00220-016-2609-8. arXiv: 1507.07775.

[Wit14] Peter Wittek. *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. Elsevier Inc., 2014, pp. 1–163. DOI: 10.1016/C2013-0-19170-2.

[WKS16] Nathan Wiebe, Ashish Kapoor, and Krysta M. Svore. "Quantum deep learning". In: *Quantum Information and Computation* 16.7-8 (2016), pp. 541–587. DOI: 10.26421/qic16.7-8-1. arXiv: 1412.3489.

[Wol18] Inc. Wolfram Research. *Mathematica*. Champaign, IL, 2018.

[Woo87] William K. Wootters. "A Wigner-function formulation of finite-state quantum mechanics". In: *Annals of Physics* 176.1 (1987). DOI: 10.1016/0003-4916(87)90176-X.

[WPG07] Michael M. Wolf and David Pérez-García. "Quantum capacities of channels with small environment". In: *Physical Review A* 75.1 (2007), p. 012303. DOI: 10.1103/PhysRevA.75.012303. arXiv: quant-ph/0607070 [quant-ph].

[WPGG07] Michael M. Wolf, David Pérez-García, and Geza Giedke. "Quantum Capacities of Bosonic Channels". In: *Physical Review Letters* 98.13 (2007), p. 130501. DOI: 10.1103/PhysRevLett.98.130501. arXiv: quant-ph/0606132 [quant-ph].

[WQ16] Mark M. Wilde and Haoyu Qi. "Energy-constrained private and quantum capacities of quantum channels". In: *IEEE Transactions on Information Theory* 64.12 (2016), pp. 7802–7827. DOI: 10.1109/TIT.2018.2854766. arXiv: 1609.01997.

[WR12] Ligong Wang and Renato Renner. "One-Shot Classical-Quantum Capacity and Hypothesis Testing". In: *Physical Review Letters* 108.20 (2012), p. 200501. DOI: 10.1103/PhysRevLett.108.200501. arXiv: 1007.5456.

[WTB17] Mark M. Wilde, Marco Tomamichel, and Mario Berta. "Converse bounds for private communication over quantum channels". In: *IEEE Transactions on Information Theory* 63.3 (2017), pp. 1792–1817. DOI: 10.1109/TIT.2017.2648825. arXiv: 1602.08898.

[WW14] Ligong Wang and Gregory W. Wornell. "A Refined Analysis of the Poisson Channel in the High-Photon-Efficiency Regime". In: *IEEE Transactions on Information Theory* 60.7 (2014), pp. 4299–4311. DOI: 10.1109/TIT.2014.2320718. arXiv: arXiv:1401.5767.

[WW19a] Xin Wang and Mark M. Wilde. "Resource theory of asymmetric distinguishability". In: *Physical Review Research* 1.3 (2019), p. 033170. DOI: 10.1103/PhysRevResearch.1.033170. arXiv: 1905.11629.

[WW19b] Xin Wang and Mark M. Wilde. "Resource theory of asymmetric distinguishability for quantum channels". In: *Physical Review Research* 1.3

(2019), p. 033169. DOI: 10.1103/PhysRevResearch.1.033169. arXiv: 1907.06306.

[XHF10] Xue-xiang Xu, Li-yun Hu, and Hong-yi Fan. "Photon-added squeezed thermal states: Statistical properties and its decoherence in a photon-loss channel". In: *Optics Communications* 283.9 (2010), pp. 1801–1809. DOI: 10.1016/j.optcom.2009.12.043.

[YCH19] Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi. "Attaining the Ultimate Precision Limit in Quantum State Estimation". In: *Communications in Mathematical Physics* 368.1 (2019), pp. 223–293. DOI: 10.1007/s00220-019-03433-4. arXiv: 1802.07587.

[YHD08] Jon Yard, Patrick Hayden, and Igor Devetak. "Capacity theorems for quantum multiple-access channels: classical-quantum and quantum-quantum capacity regions". In: *IEEE Transactions on Information Theory* 54.7 (2008), pp. 3091–3113. DOI: 10.1109/TIT.2008.924665. arXiv: quant-ph/0501045 [quant-ph].

[YKL75] H. Yuen, R. Kennedy, and Melvin Lax. "Optimum testing of multiple hypotheses in quantum detection theory". In: *IEEE Transactions on Information Theory* 21.2 (1975), pp. 125–134. DOI: 10.1109/TIT.1975.1055351.

[YO93] Horace P. Yuen and Masanao Ozawa. "Ultimate information carrying limit of quantum systems". In: *Physical Review Letters* 70.4 (1993), pp. 363–366. DOI: 10.1103/PhysRevLett.70.363.

[Yon+12] Hidehiro Yonezawa, Daisuke Nakane, Trevor A. Wheatley, Kohjiro Iwasawa, Shuntaro Takeda, Hajime Arao, et al. "Quantum-Enhanced Optical-Phase Tracking". In: *Science* 337.6101 (2012), pp. 1514–1517. DOI: 10.1126/science.1225258. arXiv: 1209.4716.

[YS78] H. Yuen and J. Shapiro. "Optical communication with two-photon coherent states–Part I: Quantum-state propagation and quantum-noise". In: *IEEE Transactions on Information Theory* 24.6 (1978), pp. 657–668. DOI: 10.1109/TIT.1978.1055958.

[Yu19] Nengkun Yu. *Quantum Closeness Testing: A Streaming Algorithm and Applications*. 2019. arXiv: 1904.03218.

[Yu20] Nengkun Yu. *Sample optimal Quantum identity testing via Pauli Measurements*. 2020. arXiv: 2009.11518.

[Yu+20] Shang Yu, Yu Meng, Raj B. Patel, Yi-Tao Wang, Zhi-Jin Ke, Wei Liu, et al. "Experimental Observation of Coherent-Information Superadditivity in a Dephrasure Channel". In: *Physical Review Letters* 125.6 (2020), p. 060502. DOI: 10.1103/PhysRevLett.125.060502. arXiv: 2003.13000.

[Yue04] Horace P. Yuen. "Communication and Measurement with Squeezed States". In: 2004, pp. 227–261. DOI: 10.1007/978-3-662-09645-1_7. arXiv: quant-ph/0109054 [quant-ph].

[Yue76] Horace P. Yuen. "Two-photon coherent states of the radiation field". In: *Physical Review A* 13.6 (1976), pp. 2226–2243. DOI: 10.1103/PhysRevA.13.2226.

[ZFF19] Zhikuan Zhao, Jack K. Fitzsimons, and Joseph F. Fitzsimons. "Quantum-assisted Gaussian process regression". In: *Physical Review A* 99.5 (2019), p. 052331. DOI: 10.1103/PhysRevA.99.052331. arXiv: 1512.03929.

[ZFG90] Wei Min Zhang, Da Hsuan Feng, and Robert Gilmore. "Coherent states: Theory and some applications". In: *Rev. Mod. Phys.* 62.4 (1990), pp. 867–927. DOI: 10.1103/RevModPhys.62.867.

[Zha+01] Shengyu Zhang, Yuan Feng, Xiaoming Sun, and Mingsheng Ying. "Upper bound for the success probability of unambiguous discrimination among quantum states". In: *Physical Review A* 64.6 (2001), p. 062103. DOI: 10.1103/PhysRevA.64.062103.

[Zha+21] Y. Zhang, M. Menotti, K. Tan, V. D. Vaidya, D. H. Mahler, L. G. Helt, et al. "Squeezed light from a nanophotonic molecule". In: *Nature Communications* 12.1 (2021), p. 2233. DOI: 10.1038/s41467-021-22540-2. arXiv: 2001.09474.

[Zho14] Tao Zhou. "Success probabilities for universal unambiguous discriminators between unknown pure states". In: *Physical Review A - Atomic, Molecular, and Optical Physics* 89.1 (2014), p. 014301. DOI: 10.1103/PhysRevA.89.014301. arXiv: arXiv:1308.0707.

[Zhu+19] Elton Yechao Zhu, Quntao Zhuang, Min Hsiu Hsieh, and Peter W. Shor. "Superadditivity in Trade-Off Capacities of Quantum Channels". In: *IEEE Transactions on Information Theory* 65.6 (2019), pp. 3973–3989. DOI: 10.1109/TIT.2018.2889082. arXiv: 1708.04314.

[ZP20a] Quntao Zhuang and Stefano Pirandola. "Ultimate Limits for Multiple Quantum Channel Discrimination". In: *Physical Review Letters* 125.8 (2020), p. 080505. DOI: 10.1103/PhysRevLett.125.080505. arXiv: 1909.05826.

[ZP20b] Quntao Zhuang and Stefano Pirandola. "Ultimate Limits for Multiple Quantum Channel Discrimination". In: *Physical Review Letters* 125.8 (2020), p. 080505. DOI: 10.1103/PhysRevLett.125.080505. arXiv: 2007.14566.

[ZYQ06] Chi Zhang, Mingsheng Ying, and Bo Qiao. "Universal programmable devices for unambiguous discrimination". In: *Physical Review A* 74.4 (2006), p. 042308. DOI: 10.1103/PhysRevA.74.042308.

[ZZS17] Elton Yechao Zhu, Quntao Zhuang, and Peter W. Shor. "Superadditivity of the Classical Capacity with Limited Entanglement Assistance". In: *Physical Review Letters* 119.4 (2017), p. 040503. DOI: 10.1103/PhysRevLett.119.040503. arXiv: 1704.06955.