SCUOLA
NORMALE
SUPERIORE

Classe di Scienze

Corso di perfezionamento in
Nanoscienze

XXXVII ciclo

Quantum learning and Shannon theory with bosonic systems and beyond

Settore Scientifico Disciplinare

Candidato
dr. Francesco Anna MELE

Relatori

Prof. Vittorio GIOVANNETTI

Prof. Ludovico LAMI

SCUOLA
NORMALE
SUPERIORE

Anno accademico 2025/2026

Abstract

How to learn a quantum state of light? How to distribute quantum information across an optical link? Is it possible to overcome the challenge of regularisation in quantum Shannon theory? This thesis investigates all three of these questions.

The first part of this thesis investigates the intersection between the two important fields of quantum learning theory and continuous-variable (CV) systems. Quantum learning theory addresses the question of how to extract classical information from quantum systems as efficiently as possible. CV systems are ubiquitous in nature and in quantum technologies, as they model bosonic systems and quantum optical platforms. This intersection raises many interesting open questions, some of which are addressed here. A first natural question is: What are the ultimate achievable limits of tomography for CV systems? We answer this question by proving that: (i) tomography of non-Gaussian states is extremely inefficient; (ii) tomography of Gaussian states is efficient; and (iii) the sample complexity of CV tomography grows exponentially with the degree of non-Gaussianity of the unknown state. Other natural questions we address in related follow-up works include: How to efficiently learn Gaussian processes? And how to efficiently test whether an unknown CV state is Gaussian or far from the set of Gaussian states? As a by-product of our analyses, we establish several bounds on the trace distance between CV states in terms of their covariance matrices, which provide a useful toolbox for research in quantum learning theory with CV systems and potentially beyond.

The second part of this thesis focuses on quantum Shannon theory with CV systems. One of the main goals of quantum Shannon theory is to understand how to efficiently transmit classical or quantum information through quantum channels. We begin by determining the maximum level of noise compatible with quantum key distribution and entanglement distribution over phase-insensitive bosonic Gaussian channels, which are standard models for optical fibres and free-space links. We do so by establishing state-of-the-art lower bounds on the secret-key capacity and the two-way quantum capacity in the high-noise regime. We then study bosonic quantum communication under several realistic constraints, including the combined effect of photon loss and dephasing and the presence of non-Markovian (memory) noise; in related follow-up work, we also consider the one-shot regime of finite channel uses and nonzero error tolerance.

The final part of this thesis shifts to finite-dimensional systems and addresses a central obstacle in quantum Shannon theory: the problem of regularisations, which makes it computationally intractable to gain a precise quantitative understanding of the ultimate efficiency of key operational tasks such as entanglement manipulation. Here we consider the problem of computing the asymptotic entanglement cost of preparing noisy quantum states exactly under PPT operations. Building on a previous characterisation of the PPT entanglement cost in terms of a regularised formula, we construct instead a hierarchy of semi-definite programs that bypasses the issue of regularisation altogether, and exponentially converges to the true asymptotic value of the entanglement cost. To our knowledge, this is the first time that an asymptotic entanglement measure is shown to be efficiently computable despite no closed-form single-letter formula being available.

Acknowledgements

I am extremely grateful to my supervisor, Vittorio Giovannetti, for more reasons than I can properly list here. First of all, his lectures on Quantum Information Theory at SNS are what really got me passionate about the subject in the first place. Despite his many commitments, he was always welcoming whenever I visited his office—whether I had a research question or needed more delicate advice about my academic path. Many of those chats turned into long, intense research discussions, and I was often left impressed by his physical intuition. From my first-year exams at SNS all the way to the end of my PhD, he has been a constant source of support. During my PhD, he gave me a lot of freedom to focus on what I actually enjoyed working on, and he encouraged (and funded) opportunities that ended up being crucial for my career: conferences, research visits, and long stays abroad—three different universities, almost two years in total away from Pisa. Even when we were far apart, his guidance was always there and played a huge role in shaping me into the researcher I am today.

During my Master's thesis, once I had my first results, Vittorio told me he wanted to introduce me to a researcher who would be excited about them. It took me a few years to realise that this was probably the luckiest moment of my academic career so far. That researcher was Ludovico Lami, now the co-supervisor of this thesis. I honestly don't have enough words to say how important Ludovico has been for me. I learned an enormous amount from him—from his mathematical precision to his care in writing clear, clean papers—and this incredibly shaped my development as a researcher. He taught me countless mathematical tricks, and we spent infinitely many hours at the blackboard together. Throughout my PhD, he was always there, and his feedback was consistently sharp and incredibly useful; many times I was impressed by his brilliant ideas. I have never stopped learning from Ludovico, and I still have a lot to learn from him. I am confident that our collaboration will continue well beyond the end of my PhD. The Giovannetti–Lami combo has been perfect for my PhD, and I truly cannot imagine a better configuration of supervisors. Both of them have been examples to me, scientifically and personally, and these few lines will never suffice to express my gratitude to them.

Another special person whom I unofficially consider one of my supervisors is Jens Eisert. Jens welcomed me for a semester at the Freie Universität Berlin, where we started some amazing research together. He has always treated me as a member of his group and actively involved me in group events, even after I left Berlin. I feel that my time in Berlin has been a key part of my PhD.

Many other people have been crucial to my PhD research. Antonio Anna Mele—my twin brother—is definitely one of them. Our shared passion for quantum information constantly feeds off itself, to the point that even our parents now know what bosons, fermions, and trace distance are. People often ask: "Is it weird to have a twin brother in the same research field?" Yes, it is. "Is it stressful to write papers together?" Absolutely. We have zero boundaries, we are brutally honest with each other all the time, and at the same time we always support each other and are constantly available to give feedback on each other's research. Our first paper together turned out to be a real success.

Another person who has been crucial to my PhD is SFE (Salvatore F. E. Oliviero). I really owe a lot to him, both as a collaborator and as a friend. He was my office mate at SNS, and we did great research together. I will never forget his precious advice. Whenever I have a question about some aspect of academia, SFE is there—and he knows the answer.

Most of the time, when I am not able to answer a quantum information question, I text a message to Marco Fanizza. Since the start of my PhD, Marco has been incredibly helpful and available throughout my path, and I owe a lot to him as well.

I also owe a great deal to Lennart Bittel, who has been an amazing collaborator, always with clever ideas and helpful technical insights. Important to my PhD has also been Lorenzo Leone, both as a great collaborator and as a friend. A great thank you also goes to Farzin Salek, for being such an incredibly supportive and helpful collaborator. I also deeply thank Ulysse Chabaud, for being such an amazing collaborator and very helpful with advice, and also as one of the people I go to when I have questions about CV systems. A special mention also goes to my friend and collaborator Filippo Girardi: we have been doing amazing collaborations together only since last year, but I am sure we will work a lot more in the future.

I also deeply thank the other amazing collaborators with whom I had the honour to write a paper together. Apart from the ones mentioned above, there are: Amira Abbas, Giovanni Barbarino, Vasco Cavina, Nunzia Cerrato, Senrui Chen, Giacomo De Palma, David Gross, Dmitry Grinko, Francisco E. Gutiérrez, Vishnu V. Iyer, Junseo Lee, Pengcheng Liao, Bartosz Regula, Frank E. Q. Rodriguez, Pulkit Sinha, Salvatore Tirone, Varun Upreti, Michael Walter, Freek Witteveen, Haimeng Zhao, Xiaobin Zhao, and Quntao Zhuang.

A big thank you goes to the special people I met in the SNS group in Pisa. Among them are Alessandro, Andrea, Antonio, Ashkan, Dario, Fabio, Farzad, Federico, Francesco, Filippo, Frank, Gaetano, Peppe (Giuseppe Catalano), Giuseppe De Riso, Jason, Lorenzo, Luca, Paolo, Roberto, Salvatore Raia, Salvatore Tirone, Stefano, Samrat, SFE, Silvia, Sofia, Stefano, Vasco, and Vito.

A special mention goes to Andrea, Federico, Luca, Nunzia, and Peppe. I truly don't have enough words to describe how special these years have been thanks to you. Andrea, you have been a special friend for a long time now, ever since our first years of the Bachelor's, and you are always there to enjoy life and go on adventures together. Federico, you have been a key presence since the very beginning of my PhD, from your precious suggestions on our first day sharing an office to the two years we spent together as office mates, during which we became real friends. Luca, thank you for being such a nice person—we had an amazing time when we were in Amsterdam, where you, Nunzia, and I were ironically known as the "Italian Mafia." Nunzia, over these years you have basically become my sister; we became each other's therapists and shared countless adventures together; thank you so much for always being there. Peppe, we had so much fun over these years, and you always manage to spread positive vibes.

A special thanks go to the colleagues I met during my stay at the Freie Universität Berlin from March to July 2023, in particular Daniel, Jacopo, Ryotaro, Sumeet, and Yihui. Another special thank you goes to the colleagues I met during my stay at QuSoft and the University of Amsterdam from January to July 2024, in particular Amira, Dmitry, Fran, Léo, Llorenç, Randy, and Yaroslav.

I also want to warmly thank the people I met during my research visit at Caltech from May to August 2025. In particular, I am deeply grateful to Hsin-Yuan (Robert) Huang for kindly organising my visit and for being such a great collaborator. I also thank John Preskill for welcoming me into his group. A special thank you goes to Aditya, Bora, Haimeng, Liang, Nadine, Saleh, Senrui, and Xinui for such a warm welcome. A special mention goes to Nadine Meister, an amazing collaborator: we spent countless hours together at the blackboard, and she would always come up with new combinatorial inequalities that solved the problems we were struggling with. Another special mention goes to Senrui Chen, with whom I am also collaborating extensively, and without whom my stay at Caltech would not have been the same.

I am especially grateful to my friends from my hometown—Alessandro De Capua, Alessandro Palazzo, Andrea, Crishta, Cucca, Fabio, Gio', Pietro—who have always been there to have fun together and enjoy life.

Finally, this thesis would not have been possible without the immense love, care, and support of my family. Everything I have achieved is thanks to them, and I thank my parents from the bottom of my heart.

Contents

List of publications and preprints vii

1 Introduction 1
- 1.1 Part I: Quantum learning theory with continuous-variable systems . . . 1
- 1.2 Part II: Quantum Shannon theory with continuous-variable systems . . . 5
- 1.3 Part III: Quantum Shannon theory with finite dimensional systems . . . 7

I Quantum learning theory with continuous-variable systems 14

2 Learning quantum states of continuous-variable systems 15
- 2.1 Introduction . . . 15
- 2.2 Energy-constrained states . . . 16
- 2.3 Gaussian states . . . 18
- 2.4 t-doped Gaussian states . . . 20
- 2.5 Discussion . . . 22
- 2.6 Methods . . . 22
- 2.7 Supplementary Material . . . 28
  - 2.7.1 Preliminaries . . . 28
  - 2.7.2 Tomography of moment-constrained states . . . 46
  - 2.7.3 Tomography of bosonic Gaussian states . . . 63
  - 2.7.4 t-doped Gaussian states and unitaries . . . 87

3 Optimal estimates of trace distance between bosonic Gaussian states and applications to learning 109
- 3.1 Introduction . . . 109
- 3.2 Prior work . . . 111
- 3.3 Notation . . . 112
- 3.4 Tight upper bound . . . 112
- 3.5 Application to quantum state tomography . . . 113
- 3.6 Discussion and conclusions . . . 114
- 3.7 Technical tools . . . 114
  - 3.7.1 Alternative bounds . . . 116
- 3.8 Supplementary Material . . . 117
  - 3.8.1 Preliminaries on continuous variable systems . . . 117
  - 3.8.2 Proof of our results . . . 120
  - 3.8.3 Differentiability of Gaussian states . . . 136
  - 3.8.4 Alternative approach based on the Gaussian noise channel . . . 144
  - 3.8.5 Sample complexity for estimating the covariance matrix using heterodyne measurements . . . 150

4 Review on bosonic trace distance bounds 163

II Quantum Shannon Theory with continuous variable systems 167

5 Maximum tolerable excess noise in continuous-variable quantum key distribution and improved lower bound on two-way capacities 168
5.1 Introduction . . . 168
5.2 Results . . . 169
5.2.1 Maximum tolerable excess noise in CV-QKD . . . 169
5.2.2 Improved lower bound on two-way capacities . . . 172
5.3 Supplementary material . . . 178
5.3.1 Notation and preliminaries . . . 178
5.3.2 Results . . . 191

6 Quantum communication on the bosonic loss-dephasing channel 232
6.1 Introduction . . . 232
6.2 Preliminaries . . . 233
6.3 Anti-degradability . . . 234
6.4 Two-way quantum communication . . . 236
6.5 Degradability . . . 238
6.6 Conclusion . . . 238
6.7 Supplementary material . . . 239
6.7.1 Preliminaries and notation . . . 239
6.7.2 Anti-degradability and Degradability of bosonic loss-dephasing channel . . . 248
6.7.3 Generalisation of our methods to general bosonic dephasing channels . . . 261
6.7.4 Coherence preservation of the bosonic loss-dephasing channel . . . 262
6.7.5 Technical lemmas . . . 265
6.7.6 Hadamard maps . . . 265

7 Optical fibres with memory effects and their quantum communication capacities 276
7.1 Introduction . . . 276
7.1.1 Delocalised Interaction Model . . . 280
7.1.2 Capacities of the model . . . 282
7.1.3 Beam splitter and thermal attenuator . . . 289
7.2 Quantum capacity, two-way quantum capacity, and secret-key capacity . . . 290
7.2.1 Quantum capacity, two-way quantum capacity, and secret-key capacity of the pure-loss channel and thermal attenuator . . . 291
7.3 A novel model of optical fibre with memory effects: the Delocalised Interaction Model 291
7.4 Problems of the "Localised" interaction model solved by the "Delocalised" one . . . 299
7.5 Avram–Parter theorem and its consequences . . . 301
7.6 Capacities of the Delocalised Interaction Model . . . 304
7.7 Capacities of the Localised Interaction Model . . . 310
7.8 Transversal attenuation . . . 311

III Quantum Shannon Theory with finite-dimensional systems 319

8 Computable entanglement cost under PPT operations 320
8.1 Introduction . . . 320
8.2 Discussion and conclusions . . . 325
8.3 Methods . . . 325
8.4 Supplementary Material . . . 328
8.4.1 Introduction . . . 328
8.4.2 Violations of additivity of E_κ . . . 331

8.4.3 New entanglement monotones: $E_{\chi,p}$ and $E_{\kappa,q}$ 332
8.4.4 Main results 342
8.4.5 From zero error to very small error 351

IV Conclusion **359**

List of publications and preprints

1. **F. A. Mele**, L. Lami, V. Giovannetti,
Maximum tolerable excess noise in continuous-variable quantum key distribution and improved lower bound on two-way capacities,
Nature Photonics 19, 329–334 (2025)
Accepted at **TQC 2023**

2. **F. A. Mele**, A. A. Mele, L. Bittel, J. Eisert, V. Giovannetti, L. Lami, L. Leone, S. F. E. Oliviero,
Learning quantum states of continuous variable systems,
Nature Physics 21, 2002–2008 (2025)
Accepted at **QIP 2025** and **TQC 2024**

3. **F. A. Mele**, L. Lami, V. Giovannetti,
Restoring quantum communication efficiency over high loss optical fibers,
Physical Review Letters 129, 180501 (2022)

4. L. Lami, **F. A. Mele**, B. Regula **(alphabetical order)**,
Computable Entanglement Cost under Positive Partial Transpose Operations,
Physical Review Letters 134, 090202 (2025)
Accepted at **QIP 2025**

5. **F. A. Mele**, F. Salek, V. Giovannetti, L. Lami,
Quantum communication on the bosonic loss-dephasing channel,
Physical Review A 110, 012460 (2024)
Accepted at **TQC 2024**

6. L. Bittel, **F. A. Mele**, A. A. Mele, S. Tirone, L. Lami,
Optimal estimates of trace distance between bosonic Gaussian states and applications to learning,
Quantum 9, 1769 (2025)
Accepted at **QIP 2025**

7. **F. A. Mele**, L. Lami, V. Giovannetti,
Quantum optical communication in the presence of strong attenuation noise,
Physical Review A 106 (4), 042437 (2022)

8. **F. A. Mele**, G. De Palma, M. Fanizza, V. Giovannetti, L. Lami,
Optical fibres with memory effects and their quantum communication capacities,
IEEE Transactions on Information Theory 70(12), 8844–8869 (2024)

9. X. Zhao, P. Liao, **F. A. Mele**, U. Chabaud, Q. Zhuang,
Complexity of quantum tomography from genuine non-Gaussian entanglement,
Nature Communications (2025)

10. **F. A. Mele**, L. Lami,
Optimising quantum data hiding,
arXiv:2510.03538 (2025)
Accepted at **QIP 2026**

11. M. Fanizza, V. Iyer, J. Lee, A. A. Mele, **F. A. Mele (alphabetical order)**,
Efficient learning of bosonic Gaussian unitaries,
arXiv:2510.05531 (2025)
Accepted at **QIP 2026**

12. F. Girardi, F. Witteveen, **F. A. Mele**, L. Bittel, S. F. E. Oliviero, D. Gross, M. Walter,
Is it Gaussian? Testing bosonic quantum states,
arXiv:2510.07305 (2025)
Accepted at **QIP 2026**

13. **F. A. Mele**, G. Barbarino, V. Giovannetti, M. Fanizza,
Achievable rates in non-asymptotic bosonic quantum communication,
arXiv:2502.05524 (2025)

14. **F. A. Mele**, G. Barbarino, V. Giovannetti, M. Fanizza,
Non-asymptotic quantum communication on lossy transmission lines with memory,
arXiv:2503.13207 (2025)

15. **F. A. Mele**, S. F. E. Oliviero, V. Upreti, U. Chabaud,
The symplectic rank of non-Gaussian quantum states,
arXiv:2504.19319 (2025)

16. L. Bittel, **F. A. Mele**, J. Eisert, A. A. Mele,
Energy-independent tomography of Gaussian states,
arXiv:2508.14979 (2025)

17. A. Abbas, N. Cerrato, F. E. Gutiérrez, D. Grinko, **F. A. Mele**, P. S. Sinha **(alphabetical order)**,
Nearly optimal algorithms to learn sparse quantum Hamiltonians in physically motivated distances,
arXiv:2509.09813 (2025)

18. F. E. Q. Rodriguez, **F. A. Mele**, S. F. E. Oliviero, V. Giovannetti, L. Lami, V. Cavina,
Extracting energy via bosonic Gaussian operations,
arXiv:2503.21748 (2025)

19. F. Girardi, **F. A. Mele**, L. Lami,
Random purification channel made simple,
arXiv:2511.23451 (2025)

20. **F. A. Mele**, F. Girardi, S. Chen, M. Fanizza, L. Lami,
Random purification channel for passive Gaussian bosons,
arXiv:2512.16878 (2025)

21. F. Girardi $*$, **F. A. Mele** $*$, H. Zhao $*$, M. Fanizza $\diamond$, L. Lami $\diamond$, ($*, \diamond =$ **These authors contributed equally**)
Random Stinespring superchannel: converting channel queries into dilation isometry queries,
arXiv:2512.16878 (2025)

Chapter 1

Introduction

Quantum Information Theory [1] is a beautiful mixture of Physics, Mathematics, and Computer Science that investigates how to exploit the laws of quantum physics to improve technology. This thesis, based on Refs. [2–22], investigates quantum information theory primarily in the setting of *continuous-variable systems* [23], while also including selected results for qubit and qudit systems.

Continuous-variable (CV) systems [23], such as bosonic and quantum optical systems, are ubiquitous both in nature and in technology. The main mathematical difference between CV systems and the more widely studied qubit or qudit systems [1] is that the former are associated with infinite-dimensional Hilbert spaces, whereas the latter are associated with finite-dimensional ones, thus rendering the analysis of CV systems sometimes trickier. A CV system can be thought of as consisting of n qudits (i.e., d-level systems) with $d = \infty$, where n is referred to as the number of modes. However, the distinction between continuous-variable and finite-dimensional quantum information is not merely mathematical. Rather, it is deeply rooted in the set of physical states and operations that can be realistically implemented and found in nature. In the CV setting, the class of experimentally accessible and technologically mature states and operations is largely restricted to *Gaussian* states and *Gaussian operations*, which include linear optical transformations, squeezing, and displacement operations. This operational constraint gives rise to a framework that is fundamentally different from that of qudit-based quantum information. Gaussian and non-Gaussian operations define a natural and physically motivated computational model, leading to a landscape of protocols, limitations, and advantages that cannot be straightforwardly inferred from finite-dimensional systems. Consequently, quantum information theory for continuous-variable systems should not be regarded as a simple extension of its qudit counterpart, but rather as a distinct paradigm with its own structural features and theoretical challenges.

This thesis is organised into three parts. Parts I and II focus on continuous-variable systems: Part I is devoted to quantum learning theory [24], while Part II addresses quantum Shannon theory [25, 26]. Part III then shifts to quantum Shannon theory in finite-dimensional (qudit) systems.

1.1 Part I: Quantum learning theory with continuous-variable systems

Quantum learning theory [24] addresses the problem of how to extract classical information from quantum systems as efficiently as possible. Over the past decade, this field has undergone rapid development in the context of finite-dimensional quantum systems [24]. Among the central questions that have been extensively investigated are the following: how many copies of an unknown quantum state are required to reconstruct a classical description of the state to a prescribed accuracy [27, 28], a task commonly referred to as quantum state tomography or quantum state learning; how many copies are necessary to learn an unknown quantum process [29–31], known

as quantum process tomography; and how structural promises on the unknown state—such as membership in a restricted and highly symmetric class like stabilizer states [32]—can dramatically reduce the resources required for learning. Closely related to these problems is the task of quantum property testing, which asks how many copies of a quantum state are needed to efficiently determine whether it satisfies a given property, for example whether an unknown state is a stabilizer state or not [33]. Despite this substantial progress in finite-dimensional settings, the development of quantum learning theory for continuous-variable systems has long remained limited. The infinite-dimensional nature of continuous-variable quantum systems, combined with physically motivated constraints such as bounded energy, renders many of the techniques developed for qubits and qudits inapplicable or insufficient. As a result, prior to the works on which this thesis is based, only a small number of contributions [34–37] addressed learning-theoretic questions in the continuous-variable regime, leaving many foundational problems unresolved. The intersection of quantum learning theory and continuous-variable quantum information—here referred to as *quantum learning theory with continuous variables*—has so far received comparatively limited attention. Nevertheless, this setting raises a number of important questions at the interface of quantum mechanics, information theory, and statistical learning theory. The works composing this thesis contribute to a systematic study of this regime: they address several of these questions, while at the same time highlighting others that remain open and constitute a promising direction for future research.

Our contributions in this direction can be summarised as follows:

- **Chapter 2 (based on Ref. [2]): Learning quantum states of continuous-variable systems.** A central problem in quantum learning theory is quantum state tomography, namely the task of reconstructing a classical description of an unknown quantum state given access to multiple copies of it. Although the notion of tomography was originally introduced in the quantum literature in the context of quantum optics, the fundamental limits of tomography for continuous-variable systems had remained largely unexplored prior to our work. In this chapter, we close this gap by determining the ultimate achievable performance of quantum state tomography for continuous-variable quantum systems.

  More precisely, we characterise the *sample complexity* of tomography, defined as the minimum number of state copies required to reconstruct an unknown state to a prescribed accuracy. While the infinite dimensionality of the underlying Hilbert space implies that tomography of arbitrary continuous-variable quantum states requires an infinite number of samples, physically relevant states are subject to energy constraints. This observation motivates our analysis of tomography for continuous-variable quantum states with bounded mean energy, where the energy bound can be interpreted as a limitation imposed by the experimental setup or by the available optical resources. Throughout this work, the reconstruction error is quantified using the trace distance, which admits a clear operational interpretation through the Holevo–Helstrom theorem. Our main results can be summarised as follows.

  - We first determine the sample complexity of tomography for arbitrary energy-constrained continuous-variable quantum states and show that, even under energy constraints, tomography is *extremely inefficient*. In particular, the sample complexity scales exponentially with the number of modes and exhibits a highly unfavourable dependence on the target estimation error.
  - We then focus on the important subclass of Gaussian states, which are ubiquitous in both natural and engineered quantum optical systems. In contrast to the generic case, we prove that tomography of Gaussian states is efficient, with a sample complexity that scales polynomially in the number of modes and in the energy. Although Gaussian states are completely characterised by their first moments and covariance matrices, a key technical challenge lies in relating estimation errors on these moments to an error

measured in trace distance. We address this challenge by deriving explicit bounds on the trace distance between Gaussian states in terms of norm distances between their first and second moments.

  - Finally, motivated by the stark contrast between the efficient learnability of Gaussian states and the extreme inefficiency observed for generic non-Gaussian states, we investigate the trade-off between tomography efficiency and the degree of non-Gaussianity of the unknown state. We do so by studying tomography of *t-doped Gaussian states*, defined as states obtained by applying arbitrary Gaussian unitaries to the vacuum followed by at most t single-mode non-Gaussian gates. We prove that the sample complexity of tomography scales exponentially with t, thereby quantitatively interpolating between the Gaussian and fully non-Gaussian regimes.

- **Chapter 3 (based on Ref. [3]): Optimal estimates of the trace distance between bosonic Gaussian states and applications to learning.** The analysis of Gaussian state tomography developed in Chapter 2 relies on upper bounds on the trace distance between two Gaussian states in terms of differences in their first and second moments. While these bounds suffice to establish the efficient learnability of Gaussian states, they are not optimal. Subsequent works, including a refinement by Holevo [38], improved upon our initial estimates. In this chapter, we derive tight upper bounds on the trace distance between two bosonic Gaussian states in terms of norm distances between their covariance matrices and first moments. These optimal bounds directly lead to improved sample-complexity guarantees for Gaussian state tomography, thereby sharpening and completing the results obtained in Chapter 2.

- **Chapter 4: Review on bosonic trace-distance bounds.** This chapter provides a brief literature review of bosonic trace-distance bounds developed in our recent works [2–5, 12]. Some of these results were already anticipated in Chapters 2 and 3, while others were derived in works [4, 5, 12] that are not fully included in this thesis. Collectively, these bounds form a useful toolbox for addressing problems in quantum learning theory with continuous-variable systems.

In addition to the works presented in the chapters of this thesis, my PhD research also contributed to further results on quantum learning theory with continuous-variable systems, which we briefly summarise below but do not include here in order to keep the thesis to a reasonable length.

- **Energy-independent tomography of Gaussian states** [4]: The sample complexity for Gaussian state tomography derived in Chapter 3 exhibits a polynomial dependence on the energy of the unknown state. This naturally raises the question of whether such an energy dependence is intrinsic, or merely an artefact of the tomography protocol employed. Indeed, the sample complexity upper bounds in Chapter 2 and Chapter 3 are achieved by a conceptually simple strategy: estimating the first moments and covariance matrix and then propagating the resulting errors to the trace distance. In Ref. [4], we show that this energy dependence can be dramatically improved. We introduce an adaptive tomography algorithm whose sample complexity depends on the energy only through a doubly logarithmic factor, and is therefore almost energy independent. The key idea is to progressively undo the squeezing present in the unknown Gaussian state before performing measurements of its moments. All the required operations are experimentally feasible, resulting in a concrete tomography protocol that is readily implementable in the laboratory. This chapter thus demonstrates that, when learning Gaussian states in trace distance, naively estimating moments is far from optimal, and that more sophisticated adaptive strategies can yield near energy-independent performance.

- **The symplectic rank of non-Gaussian quantum states** [5]: In Chapter 2, we showed that the sample complexity of CV quantum state tomography grows exponentially with the degree of non-Gaussianity of the unknown state, as quantified by the doping parameter t of a *t-doped Gaussian state*. Such states are generated by applying arbitrary Gaussian unitaries to the vacuum, followed by at most t single-mode non-Gaussian unitaries. A key structural property of t-doped Gaussian states is their *t-compressibility*: all non-Gaussian features of the state can be concentrated into at most t modes using Gaussian transformations alone. This observation naturally raises the question of whether the notion of t-compressibility can be elevated to a rigorous and intrinsic measure of non-Gaussianity that applies beyond the restricted class of t-doped states. In Ref. [5], we address this question by introducing a well-defined measure of non-Gaussianity, which we term the *symplectic rank*. The symplectic rank is defined as the number of symplectic eigenvalues of the covariance matrix that are strictly larger than one. The symplectic rank enjoys several desirable properties. Most notably, it is non-increasing under post-selected Gaussian operations, making it a natural monotone for deriving no-go theorems for quantum state conversion under Gaussian protocols. Moreover, it has direct implications for both quantum learning theory and resource-theoretic analyses of non-Gaussianity. As an additional result of independent interest, we derive an upper bound on the distance between the covariance matrices of two arbitrary non-Gaussian states in terms of the trace distance between the states themselves.

- **Efficient learning of bosonic Gaussian unitaries** [6]: In Chapter 2, we studied the problem of quantum state tomography for CV systems and showed that, while tomography of arbitrary states is highly inefficient, Gaussian states can instead be learned efficiently. A natural next question is whether a similar conclusion holds at the level of quantum dynamics: namely, can one efficiently learn *bosonic Gaussian unitaries*? In Ref. [6], we address this question and prove that learning bosonic Gaussian unitaries is indeed efficient. More precisely, we provide rigorous performance guarantees for the learning task, establishing query-complexity bounds that scale polynomially with the number of modes and with other relevant system parameters. The accuracy of the learning procedure is quantified using the *energy-constrained diamond norm* [39–41], which is the appropriate operational distance for comparing quantum channels acting on infinite-dimensional systems under physically motivated energy constraints.

- **Is it Gaussian? Testing bosonic quantum states** [7]: In the above works, we showed that assuming the underlying bosonic state to be Gaussian enables efficient learning algorithms. This naturally raises the question of how many copies are required to determine whether an unknown bosonic state is close to the set of Gaussian states, or far from it, in trace distance. In Ref. [7], we study the sample complexity of *Gaussianity testing* as a function of the number of modes. We show that for pure states, a constant number of copies suffices to decide whether the state is exactly Gaussian. In the tolerant setting, a polynomial number of copies is enough to distinguish states that are close to Gaussian from those that are far. In contrast, we prove that testing Gaussianity for general mixed states necessarily requires an exponential number of copies, revealing a fundamental limitation of certification tasks in continuous-variable systems. Our approach combines representation-theoretic techniques with recently developed tools for bounding the trace distance in continuous-variable quantum systems.

- **Complexity of quantum tomography from genuine non-Gaussian entanglement** [8]. In Ref. [8], we investigate the sample complexity of quantum state tomography for a non-trivial class of non-Gaussian states, namely *Gaussian-entanglable states*. These are states obtained by applying a Gaussian unitary to a tensor product of single-mode non-Gaussian states. We show that, despite their non-Gaussian nature, tomography of such states can

be performed efficiently, with sample complexity scaling polynomially with the number of modes.

The research area of quantum learning theory with continuous-variable systems leaves many fundamental questions open. For instance, what are the optimal sample and time complexities of tomography of Gaussian states? Can non-Gaussian protocols improve the performance of tomography even when the unknown states are Gaussian? Beyond the unitary setting, can general (non-unitary) Gaussian channels be learned efficiently under physically motivated constraints? More broadly, which structured and experimentally relevant classes of non-Gaussian states and channels are efficiently learnable? These questions provide promising directions for future work.

1.2 Part II: Quantum Shannon theory with continuous-variable systems

Quantum Shannon theory addresses the question of how to reliably transmit classical or quantum information across quantum channels as efficiently as possible. While quantum learning theorists are interested in determining the sample complexity of relevant quantum learning tasks, the main goal of quantum Shannon theorists is to estimate the *capacities* of relevant quantum channels [25, 26]. The capacities of a quantum channel quantify the maximum rate at which classical or quantum information can be transmitted reliably in the asymptotic limit of infinitely many channel uses. The relevant notion of capacity depends both on the information task—e.g. transmitting classical data, quantum states, or secret keys—and on the resources available to the communicating parties, such as pre-shared entanglement or two-way classical communication. For instance, when the goal is to generate secret keys and two-way classical communication is allowed, the task corresponds to *quantum key distribution* (QKD) [42], and the associated figure of merit is the *secret-key capacity* [26]. Over the past decade, considerable effort has been devoted to estimating the capacities of practically relevant CV channels, particularly those modelling optical fibres and free-space links. A paradigmatic example is the *pure-loss channel*, which captures the noise mechanism due to photon loss [23]. Our contributions in this direction are summarised below:

- **Chapter 5 (based on Ref. [10]): Maximum tolerable excess noise in continuous-variable quantum key distribution and improved lower bounds on two-way capacities.** We begin our journey in quantum Shannon theory with CV systems by studying the ultimate performance of QKD [42]. Since its inception in 1984 [42], QKD has transitioned from a theoretical concept to a commercially viable technology [43]. CV systems [23] offer a powerful approach to QKD, known as *CV-QKD* [43–45], that is expected to be seamlessly integrated into existing optical fibre-based Internet infrastructure in the near future [43, 46–50], making it highly practical for real-world applications. A significant challenge in this field, essential for the development of a large-scale quantum internet [43], is extending CV-QKD over long distances. In recent years, numerous experiments have been setting new distance records, achieving CV-QKD across optical fibres exceeding one hundred kilometers in length [46–50]. This progress prompts a critical question: "What is the maximum achievable distance of CV-QKD according to the laws of quantum physics?" This was identified as a "crucial" open problem by leading experts of the field [51, Section 7]. In Chapter 5, we solve this problem by determining the ultimate limitations to CV-QKD imposed by the laws of quantum physics. By leveraging this result, we have established the maximum achievable distance of CV-QKD, demonstrating that the current Internet infrastructure can in principle support CV-QKD only up to distances of approximately 990 km, which is not so far from the distance records of approximately 200 km achieved by state-of-the-art experiments [46–50]. To establish this result, we have developed a novel protocol to distribute entanglement (and hence secret keys) across noisy optical links. On a technical level, the rate of this protocol

provides state-of-the-art lower bounds on the two-way quantum capacity and secret-key capacity of phase-insensitive single-mode Gaussian channels in the high-noise regime.

- **Chapter 6 (based on Ref. [11]): Quantum communication on the bosonic loss-dephasing channel.** Quantum optical systems are typically affected by two types of noise: photon loss and dephasing. Despite extensive research on each noise process individually, a comprehensive understanding of their combined effect is still lacking. A crucial problem lies in determining the values of loss and dephasing for which the resulting loss-dephasing channel is anti-degradable, implying the absence of codes capable of correcting its effect or, alternatively, capable of enabling quantum communication. A conjecture in Ref. [52] suggested that the bosonic loss-dephasing channel is not anti-degradable if the loss is below 50%. In Chapter 6 we refute this conjecture, specifically proving that for any value of the loss, if the dephasing is above a critical value, then the bosonic loss-dephasing channel is anti-degradable. While our result identifies a large parameter region where quantum communication is not possible, we also prove that if two-way classical communication is available, then quantum communication — and thus quantum key distribution — is always achievable, even for high values of loss and dephasing.

- **Chapter 7 (based on Ref. [13]): Optical fibres with memory effects and their quantum communication capacities.** The preceding chapters, as well as much of the research on CV quantum communication, rely on an overly simplistic assumption known as the *memoryless assumption*, which posits that noise acts independently and identically on each signal. However, experimental results show that this assumption is not always satisfied. Specifically, when signals are sent in close temporal succession, this assumption fails and the noise becomes temporally correlated, displaying *non-Markovian* behaviour [53]. This reveals a significant gap in the theoretical understanding of quantum communication over non-Markovian optical links. In Chapter 7, we address this gap by determining the fundamental limits of bosonic quantum communication in such non-Markovian settings and computing the corresponding quantum capacities. Crucially, our results demonstrate that non-Markovian effects can be leveraged to enable entanglement and key distribution even in highly noisy regimes where such tasks were previously deemed impossible, opening new possibilities for the development of a large-scale quantum internet. We conducted further investigations of quantum optical communication in the presence of memory effects in Refs. [14–16], which are not included in this thesis.

Another topic I explored during my PhD, but did not include among the chapters of this thesis, concerns *non-asymptotic* bosonic quantum communication [12, 13]. Bosonic quantum communication has extensively been analysed in the *asymptotic setting*, under the idealised assumptions of infinitely many channel uses and vanishing communication errors. However, significantly fewer comprehensive analyses exist in the *non-asymptotic setting*, where only a *finite* number of channel uses is permitted. This gap is particularly critical, as practical quantum communication systems inherently operate under non-asymptotic conditions. In Refs. [12, 13], we address this gap by tackling the following key question: what is the minimum number of uses of an optical channel required to transmit k qubits, distil k Bell pairs, or generate k secret-key bits, within a given error tolerance ε? We provide easily computable bounds on this quantity in both memoryless [12] and non-Markovian [13] realistic scenarios, thus offering tools for the analysis of practical quantum communication protocols.

1.3 Part III: Quantum Shannon theory with finite dimensional systems

The third and final part of this thesis is devoted to Ref. [17], titled **Computable entanglement cost under positive partial transpose operations**, which is summarised in Chapter 8, and shifts the focus from continuous-variable systems to finite-dimensional systems.

As we saw in the Part II of this thesis, quantum Shannon theory studies the fundamental limitations on the manipulation of quantum information. Calculating those limits often involves computing certain functions that encapsulate the ultimate capabilities of information carriers. Paradigmatic examples include the various capacities of quantum channels, such as the classical [54, 55], quantum [56–58], private [58, 59], and entanglement-assisted [60, 61] capacities, but also the entanglement measures that tell us how much entanglement can be extracted from a given quantum state, i.e. the distillable entanglement [62–65], and vice versa how much entanglement must be invested to *create* that state [66, 67]. This latter quantity, the entanglement cost, is the focus of Chapter 8.

With the sole exception of the entanglement-assisted capacity, all of the above functions are expressed by *regularised formulas*, i.e. formulas that involve an explicit limit $n \to \infty$ over the number of uses of the channel or available copies of the state. For example, the entanglement cost of preparing a state $\rho = \rho_{AB}$ using local operations and classical communication (LOCC) is given by

$$E_{c,\,\mathrm{LOCC}}(\rho) = \lim_{n\to\infty} \frac{1}{n} E_f\left(\rho^{\otimes n}\right), \tag{1.1}$$

where E_f is the 'entanglement of formation' [64, 68]. Importantly, the limit $n \to \infty$ over an unbounded number of copies of the state, known as regularisation, cannot be removed, due to the phenomenon of non-additivity [69–74]. This is a problem, because the regularisation makes formulas such as (8.1) hard to control and computationally intractable. Indeed, on the one hand the underlying Hilbert space dimension is exponential in n, rendering numerical calculations quickly infeasible as n grows; on the other, there is no guarantee on the quality of the approximation obtained by stopping at the nth level of any of these hierarchies—for instance, an unbounded n may be required to even check that the quantum capacity is non-zero [75]. The regularisation thus appears to be an omnipresent curse that stifles almost every attempt to quantitatively understand the ultimate limitations of quantum information manipulation.

But is that really so? In Chapter 8, we will show how to overcome this fundamental obstacle, efficiently calculating a type of entanglement cost—expressed as a regularised quantity—on all quantum states. To this end, we look at a problem that has been studied by many authors [76–80], but for which a full solution had not been found prior to our work: namely, *zero-error asymptotic entanglement dilution with 'positive partial transpose' (PPT) operations*. We construct a solution in the form of two hierarchies of semi-definite programs that converge to the true asymptotic value of the entanglement cost from above and from below. Our main result establishes that this convergence happens exponentially fast, thus yielding an efficient algorithm that approximates the cost up to an additive error ε in time $\mathrm{poly}\big(D, \log(1/\varepsilon)\big)$, where D is the underlying Hilbert space dimension. To our knowledge, this is the first time that an asymptotic entanglement measure is shown to be efficiently computable despite no closed-form single-letter formula being available.

Bibliography

[1] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010.

[2] **Francesco Anna Mele**, Antonio A. Mele, Lennart Bittel, Jens Eisert, Vittorio Giovannetti, Ludovico Lami, Lorenzo Leone, and Salvatore F. E. Oliviero. Learning quantum states of continuous-variable systems. *Nature Physics*, 21(12):2002–2008, November 2025. ISSN 1745-2481. doi: 10.1038/s41567-025-03086-2. URL http://dx.doi.org/10.1038/s41567-025-03086-2.

[3] Lennart Bittel, **Francesco Anna Mele**, Antonio Anna Mele, Salvatore Tirone, and Ludovico Lami. Optimal estimates of trace distance between bosonic gaussian states and applications to learning. *Quantum*, 9:1769, June 2025. ISSN 2521-327X. doi: 10.22331/q-2025-06-12-1769. URL http://dx.doi.org/10.22331/q-2025-06-12-1769.

[4] Lennart Bittel, **Francesco Anna Mele**, Jens Eisert, and Antonio Anna Mele. Energy-independent tomography of gaussian states. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2508.14979.

[5] **Francesco Anna Mele**, Salvatore Francesco Emanuele Oliviero, Varun Upreti, and Ulysse Chabaud. The symplectic rank of non-gaussian quantum states. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2504.19319.

[6] Marco Fanizza, Vishnu Iyer, Junseo Lee, Antonio Anna Mele, and **Francesco Anna Mele (alphabetical order)**. Efficient learning of bosonic gaussian unitaries. *arXiv preprint arXiv:2510.05531*, 2025. URL https://arxiv.org/abs/2510.05531.

[7] Filippo Girardi, Freek Witteveen, **Francesco Anna Mele**, Lennart Bittel, Salvatore F. E. Oliviero, David Gross, and Michael Walter. Is it gaussian? testing bosonic quantum states. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2510.07305.

[8] Xiaobin Zhao, Pengcheng Liao, **Francesco Anna Mele**, Ulysse Chabaud, and Quntao Zhuang. Complexity of quantum tomography from genuine non-gaussian entanglement. *Nature Communications*, December 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-67062-3. URL http://dx.doi.org/10.1038/s41467-025-67062-3.

[9] Amira Abbas, Nunzia Cerrato, Francisco Escudero Gutiérrez, Dmitry Grinko, **Francesco Anna Mele**, and Pulkit Sinha **(alphabetical order)**. Nearly optimal algorithms to learn sparse quantum hamiltonians in physically motivated distances. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2509.09813.

[10] **Francesco Anna Mele**, Ludovico Lami, and Vittorio Giovannetti. Maximum tolerable excess noise in continuous-variable quantum key distribution and improved lower bound on two-way capacities. *Nature Photonics*, 19(3):329–334, January 2025. ISSN 1749-4893. doi: 10.1038/s41566-024-01595-9. URL http://dx.doi.org/10.1038/s41566-024-01595-9.

[11] **Francesco Anna Mele**, Farzin Salek, Vittorio Giovannetti, and Ludovico Lami. Quantum communication on the bosonic loss-dephasing channel. *Physical Review A*, 110(1):012460, Jul 2024. doi: 10.1103/PhysRevA.110.012460. URL https://link.aps.org/doi/10.1103/PhysRevA.110.012460.

[12] **Francesco Anna Mele**, Giovanni Barbarino, Vittorio Giovannetti, and Marco Fanizza. Achievable rates in non-asymptotic bosonic quantum communication. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2502.05524.

[13] **Francesco Anna Mele**, Giacomo De Palma, Marco Fanizza, Vittorio Giovannetti, and Ludovico Lami. Optical fibers with memory effects and their quantum communication capacities. *IEEE Transactions on Information Theory*, 70(12):8844–8869, 2024. doi: 10.1109/TIT.2024.3450501.

[14] **Francesco Anna Mele**, Giovanni Barbarino, Vittorio Giovannetti, and Marco Fanizza. Non-asymptotic quantum communication on lossy transmission lines with memory. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2503.13207.

[15] **Francesco Anna Mele**, Ludovico Lami, and Vittorio Giovannetti. Restoring quantum communication efficiency over high loss optical fibers. *Physical Review Letters*, 129(18), October 2022. ISSN 1079-7114. doi: 10.1103/PhysRevLett.129.180501. URL http://dx.doi.org/10.1103/PhysRevLett.129.180501.

[16] **Francesco Anna Mele**, Ludovico Lami, and Vittorio Giovannetti. Quantum optical communication in the presence of strong attenuation noise. *Physical Review A*, 106(4):042437, Oct 2022. doi: 10.1103/PhysRevA.106.042437. URL https://link.aps.org/doi/10.1103/PhysRevA.106.042437.

[17] Ludovico Lami, **Francesco Anna Mele**, and Bartosz Regula **(alphabetical order)**. Computable entanglement cost under positive partial transpose operations. *Physical Review Letters*, 134(9):090202, Mar 2025. doi: 10.1103/PhysRevLett.134.090202. URL https://link.aps.org/doi/10.1103/PhysRevLett.134.090202.

[18] **Francesco Anna Mele** and Ludovico Lami. Optimising quantum data hiding. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2510.03538.

[19] Filippo Girardi, **Francesco Anna Mele**, and Ludovico Lami. Random purification channel made simple. *arXiv preprint arXiv:2511.23451*, 2025. URL https://arxiv.org/abs/2511.23451.

[20] **Francesco Anna Mele**, Filippo Girardi, Senrui Chen, Marco Fanizza, and Ludovico Lami. Random purification channel for passive gaussian bosons. *arXiv preprint arXiv:2512.16878*, 2025. URL https://arxiv.org/abs/2512.16878.

[21] Filippo Girardi, **Francesco Anna Mele**, Haimeng Zhao, Marco Fanizza, and Ludovico Lami. Random stinespring superchannel: converting channel queries into dilation isometry queries. *arXiv preprint arXiv:2512.20599*, 2025. URL https://arxiv.org/abs/2512.20599.

[22] Frank Ernesto Quintela Rodriguez, **Francesco Anna Mele**, Salvatore Francesco Emanuele Oliviero, Vittorio Giovannetti, Ludovico Lami, and Vasco Cavina. Extracting energy via bosonic gaussian operations. *arXiv preprint*, 2025. URL https://arxiv.org/abs/2503.21748.

[23] A. Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, Taylor & Francis Group, Boca Raton, USA, 2017.

[24] Anurag Anshu and Srinivasan Arunachalam. A survey on the complexity of learning quantum states. *Nature Rev. Phys.*, 6(1):59–69, 2024. doi: 10.1038/s42254-023-00662-4. URL https://doi.org/10.1038/s42254-023-00662-4.

[25] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2nd edition, 2017. ISBN 9781316813300.

[26] S. Khatri and M. M. Wilde. Principles of quantum communication theory: A modern approach. *preprint arXiv:2011.04672*, 2020.

[27] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. In *Proc. 48th Annual ACM Symp. Theory Comput.*, STOC '16, pages 913–925, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341325. doi: 10.1145/2897518.2897585. URL https://doi.org/10.1145/2897518.2897585.

[28] Ryan O'Donnell and John Wright. Efficient quantum tomography, 2015. URL https://arxiv.org/abs/1508.01907.

[29] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance, 2023.

[30] Antonio Anna Mele and Lennart Bittel. Optimal learning of quantum channels in diamond distance. *Preprint arXiv:2512.10214*, 2025. URL https://arxiv.org/abs/2512.10214.

[31] Kean Chen, Nengkun Yu, and Zhicheng Zhang. Quantum channel tomography and estimation by local test. *arXiv preprint arXiv:2512.13614*, 2025. URL https://arxiv.org/abs/2512.13614.

[32] Ashley Montanaro. Learning stabilizer states by bell sampling, 2017.

[33] 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. *Communications in Mathematical Physics*, 385(3):1325–1393, 2021.

[34] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature Comm.*, 6(1):8498, 2015. ISSN 2041-1723. doi: 10.1038/ncomms9498.

[35] Srilekha Gandhari, Victor V. Albert, Thomas Gerrits, Jacob M. Taylor, and Michael J. Gullans. Precision bounds on continuous-variable state tomography using classical shadows, 2023.

[36] S. Becker, N. Datta, L. Lami, and C. Rouzé. Classical shadow tomography for continuous variables quantum systems. *IEEE Trans. Inf. Theory*, 70(5):3427–3452, 2024. doi: 10.1109/TIT.2024.3357972.

[37] Changhun Oh, Senrui Chen, Yat Wong, Sisi Zhou, Hsin-Yuan Huang, Jens A. H. Nielsen, Zheng-Hao Liu, Jonas S. Neergaard-Nielsen, Ulrik L. Andersen, Liang Jiang, and John Preskill. Entanglement-enabled advantage for learning a bosonic random displacement channel, 2024.

[38] A. S. Holevo. On estimates of trace-norm distance between quantum gaussian states, 2024. URL https://arxiv.org/abs/2408.11400.

[39] Andreas Winter. Energy-constrained diamond norm with applications to the uniform continuity of continuous variable channel capacities, 2017.

[40] Maxim E. Shirokov. On the energy-constrained diamond norm and its application in quantum information theory. *Probl. Inf. Transm.*, 54(1):20–33, 2018. ISSN 1608-3253. doi: 10.1134/S0032946018010027.

[41] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi. Fundamental limits of repeaterless quantum communications. *Nat. Commun.*, 8(1):15043, 2017. doi: 10.1038/ncomms15043.

[42] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. In *Proc. IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, pages 175–179, 1984. See also arXiv:2003.06557.

[43] S. Pirandola et al. Advances in quantum cryptography. *Advances in Optics and Photonics*, 12(4):1012–1236, 2020. doi: 10.1364/AOP.361502. URL http://opg.optica.org/aop/abstract.cfm?URI=aop-12-4-1012.

[44] F. Grosshans and P. Grangier. Continuous variable quantum cryptography using coherent states. *Phys. Rev. Lett.*, 88(5), 2002. ISSN 1079-7114. doi: 10.1103/physrevlett.88.057902. URL http://dx.doi.org/10.1103/PhysRevLett.88.057902.

[45] F. Laudenbach, C. Pacher, C.-H. F. Fung, A. Poppe, M. Peev, B. Schrenk, M. Hentschel, P. Walther, and H. Hübel. Continuous-variable quantum key distribution with gaussian modulation-the theory of practical implementations. *Advanced Quantum Technologies*, 1(1), 2018. ISSN 2511-9044. doi: 10.1002/qute.201800011. URL http://dx.doi.org/10.1002/qute.201800011.

[46] Y. Zhang, Z. Chen, S. Pirandola, X. Wang, C. Zhou, B. Chu, Y. Zhao, B. Xu, S. Yu, and H. Guo. Long-distance continuous-variable quantum key distribution over 202.81 km of fiber. *Phys. Rev. Lett.*, 125:010502, 2020. doi: 10.1103/PhysRevLett.125.010502. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.010502.

[47] A. A. E. Hajomer, I. Derkach, N. Jain, H.-M. Chin, U. L. Andersen, and T. Gehring. Long-distance continuous-variable quantum key distribution over 100-km fiber with local local oscillator. *Science Advances*, 10(1):eadi9474, 2024. doi: 10.1126/sciadv.adi9474. URL https://www.science.org/doi/abs/10.1126/sciadv.adi9474.

[48] Y. Zhang, Z. Li, Z. Chen, C. Weedbrook, Y. Zhao, X. Wang, Y. Huang, C. Xu, X. Zhang, Z. Wang, M. Li, X. Zhang, Z. Zheng, B. Chu, X. Gao, N. Meng, W. Cai, Z. Wang, G. Wang, S. Yu, and H. Guo. Continuous-variable QKD over 50 km commercial fiber. *Quantum Science and Technology*, 4(3):035006, 2019. ISSN 2058-9565. doi: 10.1088/2058-9565/ab19d1. URL http://dx.doi.org/10.1088/2058-9565/ab19d1.

[49] Y. Pi, H. Wang, Y. Pan, Y. Shao, Y. Li, J. Yang, Y. Zhang, W. Huang, and B. Xu. Sub-mbps key-rate continuous-variable quantum key distribution with local local oscillator over 100-km fiber. *Optics Letters*, 48(7):1766, 2023. ISSN 1539-4794. doi: 10.1364/ol.485913. URL http://dx.doi.org/10.1364/OL.485913.

[50] D. Huang, P. Huang, D. Lin, and G. Zeng. Long-distance continuous-variable quantum key distribution by controlling excess noise. *Scientific Reports*, 6:19201, 01 2016. doi: 10.1038/srep19201.

[51] S. Pirandola, S. L. Braunstein, R. Laurenza, C. Ottaviani, T. P. W. Cope, G. Spedalieri, and L. Banchi. Theory of channel simulation and bounds for private communication. *Quantum Sci. Technol.*, 3(3):035009, 2018. doi: 10.1088/2058-9565/aac394.

[52] P. Leviant, Q. Xu, L. Jiang, and S. Rosenblum. Quantum capacity and codes for the bosonic loss-dephasing channel. *Quantum*, 6:821, sep 2022. doi: 10.22331/q-2022-09-29-821. URL https://doi.org/10.22331/q-2022-09-29-821.

[53] F. Caruso, V. Giovannetti, C. Lupo, and S. Mancini. Quantum channels and memory effects. *Rev. Mod. Phys.*, 86:1203–1259, 2014. doi: 10.1103/RevModPhys.86.1203.

[54] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory*, 44(1):269–273, 1998. ISSN 1557-9654. doi: 10.1109/18.651037.

[55] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997. doi: 10.1103/PhysRevA.56.131.

[56] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, 55:1613–1622, 1997. doi: 10.1103/PhysRevA.55.1613.

[57] P. Shor. The quantum channel capacity and coherent information. Lecture notes, MSRI Workshop on Quantum Computation, 2002.

[58] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, 51(1):44–55, 2005. ISSN 1557-9654. doi: 10.1109/TIT.2004.839515.

[59] N. Cai, A. Winter, and R. W. Yeung. Quantum privacy and quantum wiretap channels. *Probl. Inf. Transm.*, 40(4):318–336, 2004. doi: 10.1007/s11122-005-0002-x. URL https://doi.org/10.1007/s11122-005-0002-x.

[60] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83(15):3081, 1999. doi: 10.1103/PhysRevLett.83.3081.

[61] 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. *IEEE Trans. Inf. Theory*, 48(10): 2637–2655, 2002. ISSN 0018-9448. doi: 10.1109/TIT.2002.802612.

[62] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, 1996. doi: 10.1103/PhysRevA.53.2046.

[63] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–725, 1996. doi: 10.1103/PhysRevLett.76.722.

[64] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, 1996. doi: 10.1103/PhysRevA.54.3824.

[65] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. Royal Soc. A*, 461(2053):207–235, 2005. doi: 10.1098/rspa.2004.1372.

[66] P. M. Hayden, M. Horodecki, and B. M. Terhal. The asymptotic entanglement cost of preparing a quantum state. *J. Phys. A*, 34(35):6891–6898, 2001. doi: 10.1088/0305-4470/34/35/314.

[67] H. Yamasaki, K. Kuroiwa, P. Hayden, and L. Lami. Entanglement cost for infinite-dimensional physical systems. *Preprint arXiv:2401.09554*, 2024. doi: 10.48550/arXiv.2401.09554.

[68] W. K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.*, 80:2245–2248, 1998. doi: 10.1103/PhysRevLett.80.2245. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.2245.

[69] K. G. H. Vollbrecht and R. F. Werner. Entanglement measures under symmetry. *Phys. Rev. A*, 64:062307, 2001. doi: 10.1103/PhysRevA.64.062307.

[70] P. W. Shor. Equivalence of additivity questions in quantum information theory. *Commun. Math. Phys.*, 246(3):473–473, 2004. doi: 10.1007/s00220-004-1071-1.

[71] P. Hayden and A. Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$. *Commun. Math. Phys.*, 284(1):263–280, 2008. doi: 10.1007/s00220-008-0624-0.

[72] T. Cubitt, A. W. Harrow, D. Leung, A. Montanaro, and A. Winter. Counterexamples to additivity of minimum output p-Rényi entropy for p close to 0. *Commun. Math. Phys.*, 284 (1):281–290, 2008. doi: 10.1007/s00220-008-0625-z.

[73] M. B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nat. Phys.*, 5(4):255–257, 2009. doi: 10.1038/nphys1224.

[74] G. Smith and J. Yard. Quantum communication with zero-capacity channels. *Science*, 321 (5897):1812–1815, 2008. doi: 10.1126/science.1162242.

[75] T. Cubitt, D. Elkouss, W. Matthews, M. Ozols, D. Pérez-García, and S. Strelchuk. Unbounded number of channel uses may be required to detect quantum capacity. *Nat. Commun.*, 6(1): 6739, 2015. doi: 10.1038/ncomms7739. URL https://doi.org/10.1038/ncomms7739.

[76] K. Audenaert, M. B. Plenio, and J. Eisert. Entanglement cost under positive-partial-transpose-preserving operations. *Phys. Rev. Lett.*, 90:027901, 2003. doi: 10.1103/PhysRevLett.90.027901.

[77] X. Wang and M. M. Wilde. Cost of quantum entanglement simplified. *Phys. Rev. Lett.*, 125: 040502, 2020. doi: 10.1103/PhysRevLett.125.040502.

[78] X. Wang and M. M. Wilde. Exact entanglement cost of quantum states and channels under positive-partial-transpose-preserving operations. *Phys. Rev. A*, 107:012429, 2023. doi: 10.1103/PhysRevA.107.012429.

[79] G. Gour and C. M. Scandolo. Dynamical entanglement. *Phys. Rev. Lett.*, 125:180505, 2020. doi: 10.1103/PhysRevLett.125.180505.

[80] G. Gour and C. M. Scandolo. Entanglement of a bipartite channel. *Phys. Rev. A*, 103:062422, 2021. doi: 10.1103/PhysRevA.103.062422.

Part I

Quantum learning theory with continuous-variable systems

Chapter 2

Learning quantum states of continuous-variable systems

Quantum state tomography, aimed at deriving a classical description of an unknown state from measurement data, is a fundamental task in quantum physics. In this Chapter, we analyse the ultimate achievable performance of tomography of continuous-variable systems, such as bosonic and quantum optical systems. We prove that tomography of these systems is extremely inefficient in terms of time resources, much more so than tomography of finite-dimensional systems: not only does the minimum number of state copies needed for tomography scale exponentially with the number of modes, but it also exhibits a dramatic scaling with the trace-distance error, even for low-energy states, in stark contrast with the finite-dimensional case. On a more positive note, we prove that tomography of Gaussian states is efficient. To accomplish this, we answer a fundamental question for the field of continuous-variable quantum information: if we know with a certain error the first and second moments of an unknown Gaussian state, what is the resulting trace-distance error that we make on the state? Lastly, we demonstrate that tomography of non-Gaussian states prepared through Gaussian unitaries and a few local non-Gaussian evolutions is efficient and experimentally feasible.

[This chapter is based on F. A. Mele, A. A. Mele, L. Bittel, J. Eisert, V. Giovannetti, L. Lami, L. Leone, S. F. E. Oliviero, Learning quantum states of continuous variable systems, Nature Physics 21, 2002–2008 (2025)]

2.1 Introduction

A fundamental task in quantum physics, known as *quantum state tomography* [1], is to derive a classical description of a quantum system from experimental data. It serves not only as a powerful method to investigate Nature, but also as a diagnostic tool for benchmarking and verifying quantum devices [1–4]. The concept of tomography dates back to the 1990s [5, 6], when it was first introduced for *continuous-variable* (CV) systems [7] and then for finite-dimensional systems. Over the years, tomography algorithms for CV systems have been extensively developed, becoming a bread-and-butter tool for quantum opticians [5, 6].

In the last decade, rapid advances in accurately preparing quantum states [8] have spurred the development of a new research field known as *quantum learning theory* [1], which investigates the question of how to learn properties of quantum systems as efficiently as possible. When the focus is on obtaining a full classical description of a quantum system, such investigation reduces to quantum state tomography, which is considered the cornerstone of quantum learning theory [1]. Despite tomography being a longstanding concept, many of its fundamental properties have only recently been unveiled due to advancements in this new field, albeit only for finite-dimensional systems and not for CV systems.

Let us introduce the concept of quantum state tomography from a quantum learning theory

perspective. Formally, the task of quantum state tomography involves two parameters: the number of state copies N and the error parameter ε. Given N copies of an unknown state, the goal is to output a classical description of a state which is guaranteed to be ε-close to the true unknown state with high probability (this is stated in all rigour in the Methods section). Among various notion of distance to measure the ε-closeness, the *trace distance* emerges as the most meaningful one due to its operational significance [9, 10]: given two states ε-close in trace distance, no quantum measurement can distinguish them by more than ε. Thus, the two states are in fact indistinguishable to any physical observer within a ε resolution. The optimal performance of tomography is quantified by the *sample complexity*, which is the minimum number of copies N needed to achieve tomography with error ε. When the sample complexity and the running time of the tomography procedure scale polynomially with the number of constituents (e.g., qubits, modes), tomography is deemed *efficient*.

With vast many-body systems naturally emerging in Nature on one side, and the rapid advancement in constructing large-scale quantum devices on the other, minimising required resources has become imperative. Consequently, determining the sample complexity of tomography of physically relevant classes of states is currently a pressing problem [1]. Notably, the sample complexity has been determined for n-qubit systems [1, 11–14]: tomography of mixed states requires $O(4^n/\varepsilon^2)$ copies, while for pure states the required number of copies reduces to $O(2^n/\varepsilon^2)$.

While quantum learning theory has been extensively developed for finite-dimensional systems, it remains almost unexplored for CV systems [15–18]. This is partly because the usual approaches to CV tomography are based on mere approximations of phase-space functions [5–7], which do not account for trace-distance error and thus lack operational significance. This is a significant gap in the literature, especially considering that in recent years photonic quantum devices have been at the forefront of attempts to demonstrate quantum advantage, particularly through boson sampling [19–21] and quantum simulation experiments [22]. Moreover, photonic platforms play a pivotal role in various quantum technologies, including quantum computation [23–27], communication [28–33], and sensing [34–37]. Therefore, determining the ultimate achievable performance of CV tomography has become a crucial, pressing problem — which we solve in this work.

2.2 Energy-constrained states

A CV system corresponds to n modes, each associated with an infinite-dimensional Hilbert space. From the outset, the infinitely-many degrees of freedom of a single bosonic mode make tomography a challenging concept: What does it mean for a tomography algorithm to output a classical description of an infinite-dimensional state? Does this imply the output would be an infinite-dimensional matrix, which no classical computer could store? After all, not even Nature itself can deal with infinitely many degrees of freedom, and, as a consequence, physical states are inevitably subjected to certain constraints. Indeed, in the real world, energy is finite. The energy budget available in quantum optics laboratories is limited, as is the energy emitted by the Sun. Thus, CV quantum states only gain physical meaning under energy constraints; this turns the task of tomography from unavoidably inconceivable to a precious tool for investigating bosonic quantum systems. As explained below, we can indeed devise tomography algorithms capable of achieving arbitrarily low trace-distance error, for infinite-dimensional states subjected to some energy constraint.

By definition, an n-mode state ρ satisfies the *energy constraint* E if the expectation value of the energy observable satisfies

$$\mathrm{Tr}[\hat{E}_n\rho] \leq nE\,. \tag{2.1}$$

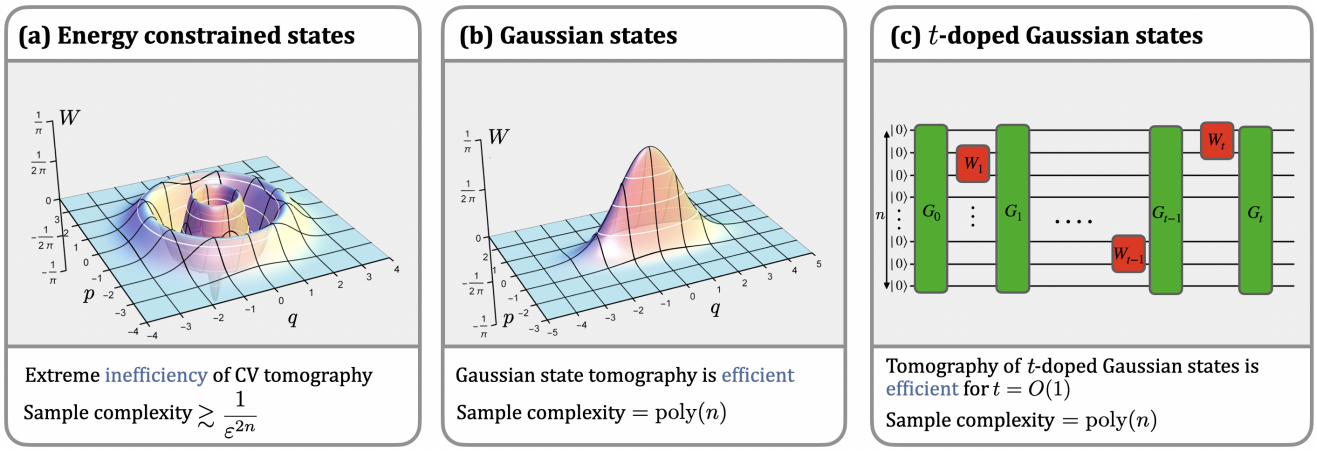


Figure 2.1: We identify strong limitations against **(a)** quantum state tomography of continuous-variable systems subject to energy constraints inherent in experimental platforms. Here, n is the number of modes, while ε is the trace-distance error. Our investigation reveals a new phenomenon dubbed 'extreme inefficiency' of continuous-variable quantum state tomography. Specifically, the number of copies required for tomography of n-mode energy-constrained states must scale at least as ε^{-2n}. This dramatic scaling is a unique feature of continuous-variable systems, standing in stark contrast to finite-dimensional systems where the required number of copies scales with the trace-distance error as ε^{-2}. Therefore, we ask whether there exist physically interesting classes of states for which tomography is efficient. We answer this in the affirmative by presenting **(b)** an efficient tomography algorithm for tomography of Gaussian states with provable guarantees in trace distance. Our analysis is based on novel technical tools of independent interest: specifically, we introduce simple bounds on the trace distance between two Gaussian states in terms of the norm distance between their first moments and covariance matrices. Finally, we demonstrate **(c)** that tomography of non-Gaussian states prepared by Gaussian unitaries and a few local non-quadratic Hamiltonian evolutions is still efficient. Remarkably, both of these efficient tomography algorithms are experimentally feasible to implement in quantum optics laboratories.

Here, the energy observable is defined as

$$\hat{E}_n := \frac{1}{2}\sum_{i=1}^{n}\left(\hat{p}_i^2 + \hat{x}_i^2\right), \tag{2.2}$$

where $\{\hat{x}_i\}_{i=1}^n$ and $\{\hat{p}_i\}_{i=1}^n$ denote the position and momentum operators of the n modes [7]. We normalise the right-hand-side of (2.1) with n because the energy is an extensive observable. Note that the energy constraint in (2.1) is equivalent to a constraint on the mean *total photon number* of the state (see (2.18) in the Methods).

The following theorem, proven in the Supplementary Material (SM), determines the sample complexity of tomography of energy-constrained pure states.

Theorem 1 (Tomography of energy-constrained pure states)**.** *The sample complexity of tomography of n-mode pure states is $O(E^n/\varepsilon^{2n})$. Here, ε is the trace-distance error and E is the energy constraint.*

This result establishes that *every* tomography algorithm — even standard algorithms based on homodyne and heterodyne detections (see Fig. 2.2) — must use *at least* $O(E^n/\varepsilon^{2n})$ copies to achieve error ε. Conversely, it also establishes that there exists a tomography algorithm that achieves an error ε given access to the above number of copies.

This finding reveals a striking phenomenon that we dub 'extreme inefficiency' of CV tomography: not only does the number of copies required for CV tomography scale exponentially with the number of modes n, as for finite-dimensional systems, but it also has a dramatic scaling with

respect to the error ε. Specifically, the scaling of $\sim \varepsilon^{-2n}$ is a unique feature of CV tomography, being in stark contrast with the finite-dimensional setting characterised by the scaling of $\sim \varepsilon^{-2}$. While in the finite-dimensional setting the error can be halved by increasing the number of copies by a factor 4, which is cheap, in the CV setting one needs an exponential factor 4^n, which is arresting.

To emphasise this remarkable behaviour, let us illustrate with an example the substantial difference between tomography of finite-dimensional systems and tomography of CV systems. Let us estimate the time required to achieve tomography with error $\varepsilon = 10\%$ in both settings. Assume that each copy of the state is produced and processed every 1 ns (typical for qubits and light pulses). Then, tomography of 10-qubit pure states only requires ~ 0.1 ms. However, as a consequence of Theorem 1, tomography of 10-mode pure states with energy constraint $E = 1$ requires ~ 3000 years, which shows that CV tomography becomes impractical even for a few modes. This highlights that tomography of CV systems is extremely inefficient, much more so than tomography of finite-dimensional systems.

In the forthcoming theorem, proven in the SM, we find bounds on the sample complexity of tomography of energy-constrained *mixed* states.

Theorem 2 (Tomography of energy-constrained mixed states)**.** *A number $O(E^{2n}/\varepsilon^{3n})$ of copies is sufficient to achieve tomography of n-mode mixed states. Conversely, tomography of such states requires at least $O(E^{2n}/\varepsilon^{2n})$ copies. Here, ε is the trace-distance error and E is the energy constraint.*

The core idea underpinning the attainable (yet arresting) performance guarantees of tomography presented in Theorem 2 is that energy-constrained CV states are effectively described by finite-dimensional systems. In particular, we prove that any (possibly mixed) energy-constrained state can be approximated, up to trace-distance error ε, by a D-dimensional state with rank r such that

$$\begin{aligned} D &= O(E^n/\varepsilon^{2n})\,, \\ r &= O(E^n/\varepsilon^n)\,. \end{aligned} \tag{2.3}$$

The first result of Theorem 2 simply follows from the observation that the sample complexity of tomography of D-dimensional states with rank r is $O(Dr)$ [1, 11–14]. Remarkably, Eq. (2.3) establishes that physical CV states — those with bounded energy — are characterised by a finite number of physically relevant parameters.

These conclusions bring us to shed light on the core question in CV tomography posed at the beginning of the section: while tomography of arbitrary CV states is indeed meaningless, considering that infinitely many parameters are physically inconceivable, with an energy constraint tomography gains real-world relevance. Indeed, in this latter case it is sufficient for the output to be a matrix of *finite* dimension, i.e. $O(E^n/\varepsilon^{2n})$.

The above findings can be generalised by considering constraints on any higher moments of the energy. By definition, similarly to (2.1), an n-mode state ρ satisfies the k-moment constraint E if

$$\left(\mathrm{Tr}[(\hat{E}_n)^k \rho]\right)^{\frac{1}{k}} \leq nE\,. \tag{2.4}$$

As proven in the SM, the sample complexity of tomography of pure states becomes $O(E^n/\varepsilon^{2n/k})$, while for mixed states the sample complexity is upper bounded by $O(E^{2n}/\varepsilon^{3n/k})$ and lower bounded by $O(E^{2n}/\varepsilon^{2n/k})$, thus recovering Theorems 1–2 for $k = 1$. Notably, the sample complexity decreases as k increases, and the dramatic scaling $\sim \varepsilon^{-2n/k}$ disappears for $k \gtrsim n$.

2.3 Gaussian states

We have identified strong limitations on tomography of CV systems, even under stringent energy constraints. This prompts a natural question: can tomography be efficiently performed for

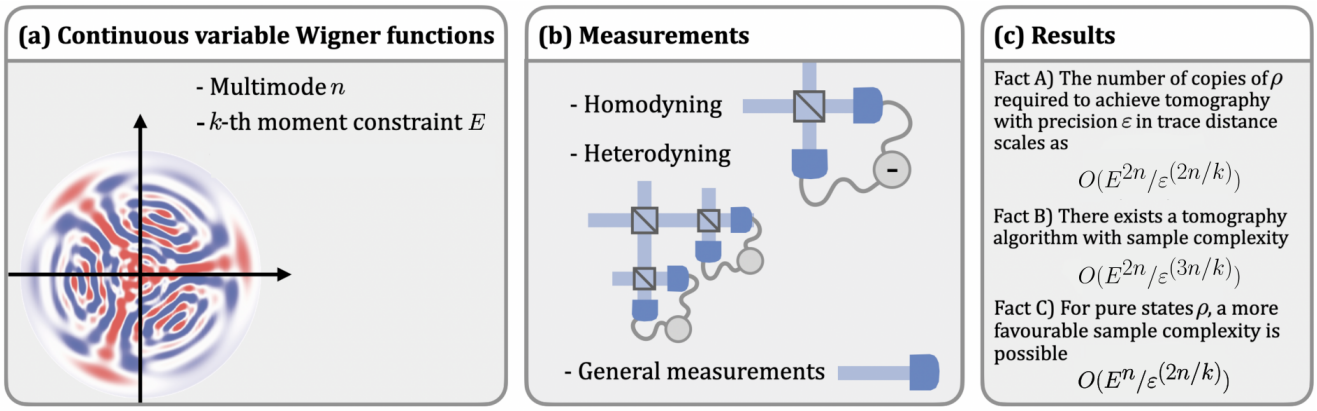


Figure 2.2: We establish fundamental bounds on the resources required for (a) quantum state tomography of continuous-variable k-th moment constrained quantum states, highlighting the pronounced inefficiency of any strategy aiming to solve this task. (b) Our results encompass *any* possible strategy, including those using only homodyne and heterodyne measurements, as well as other experimentally feasible operations in photonic platforms, and even general measurements. This means, independently from the techniques used, tomography of CV states is impractical. (c) We identify three key results, labelled Facts A-C. The implication is that the resources needed for tomography exhibit strong dependence on the desired accuracy, scaling as $\sim \varepsilon^{-2n/k}$.

more structured, yet physically interesting, classes of CV states? Here, we answer this question affirmatively by demonstrating that tomography of Gaussian states [7] is efficient. This finding is not only conceptually important, but also practically significant, as Gaussian states are the most commonly used states in quantum optics laboratories and have key applications in quantum sensing, communication, and computing [7].

We start by investigating the following problem, rather fundamental for the field of CV quantum information. It is well known that Gaussian states are in one-to-one correspondence with their first moments and covariance matrices [7]. However, since in practice one has access only to a finite number of copies of an unknown Gaussian state, it is impossible to determine its first moment and covariance matrix *exactly*. Instead, one can only obtain arbitrarily good approximations of them. Given the operational meaning of the trace distance [9, 10], it is thus a fundamental problem to answer the following question: 'if we know with an error ε the first moment and the covariance matrix of an unknown Gaussian state, what is the resulting trace-distance error that we make in approximating the underlying state?' The following theorem, proven in the Methods, answers this question.

Theorem 3 (Error propagation from moments to trace distance)**.** *If we know with an error ε the first moment and the covariance matrix of an unknown Gaussian state, the resulting trace-distance error is at most $O(\sqrt{\varepsilon})$ and at least $O(\varepsilon)$.*

To prove this result, in Theorem 10 and in Theorem 11 of the Methods we find stringent bounds on the trace distance between two Gaussian states in terms of the norm distance between their first moments and covariance matrices, which constitute technical tools of independent interest for the field of CV quantum information.

Notably, Theorem 3 is the key tool to prove the main result of this section: *tomography of Gaussian states is efficient*. Specifically, the forthcoming theorem, proven in the SM, shows that the sample complexity of tomography of Gaussian states scales polynomially with the number of modes.

Theorem 4 (Tomography of Gaussian states)**.** *There exists an algorithm that uses*

$$O\left(\frac{n^7 E^4}{\varepsilon^4}\right) = \mathrm{poly}(n) \tag{2.5}$$

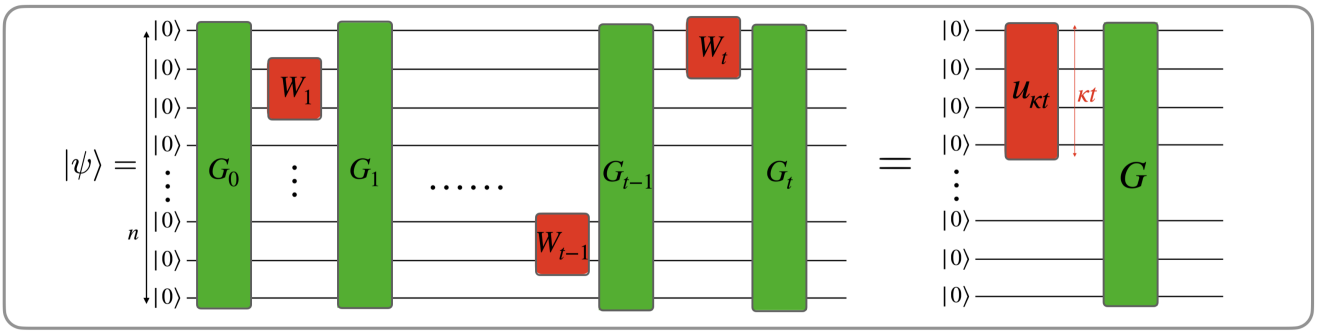


Figure 2.3: Pictorial representation of a t-doped Gaussian state. By definition, a t-doped Gaussian state vector $|\psi\rangle$ is a state prepared by applying Gaussian unitaries $G_0, \cdots, G_t$ (green boxes) and at most t non-Gaussian κ-local unitaries $W_1, \cdots, W_t$ (red boxes) to the n-mode vacuum. A unitary is said to be κ-local if it is generated by a Hamiltonian which is a polynomial in at most κ operators from the set of position operators $\{\hat{x}_i\}_{i=1}^n$ and momentum operators $\{\hat{p}_i\}_{i=1}^n$ of the n modes. The figure also shows the decomposition proven in Theorem 5, which establishes that all the non-Gaussianity in $|\psi\rangle$ can be compressed in a localised region consisting of κt modes by applying a Gaussian unitary $G^\dagger$ to $|\psi\rangle$.

copies to achieve tomography of an unknown n-mode Gaussian state. Here, ε is the trace-distance error and E is the energy constraint.

This algorithm simply consists in estimating the first moment and the covariance matrix of the unknown Gaussian state, both procedures routinely performed in quantum optics laboratories using homodyne detection [7]. Additionally, the algorithm runs in poly(n) time, and its output is efficient to store, as it consists only of the $O(n^2)$ parameters of the first moment and covariance matrix.

In summary, we have established the efficiency of tomography of Gaussian states. However, what if the state deviates slightly from being exactly Gaussian? This is a crucial question, especially considering imperfections during state preparation in experimental setups. In the SM, we demonstrate that our tomography algorithm is robust against little perturbations caused by non-Gaussian noise (e.g., dephasing noise). Even if the unknown state is a *slightly perturbed* Gaussian state — technically, a state with sufficiently small *relative entropy of non-Gaussianity* [38] — our algorithm remains effective and tomography remains efficient.

2.4 t-doped Gaussian states

Having demonstrated that tomography of arbitrary (non-Gaussian) states is extremely inefficient (Theorem 1) and that tomography of Gaussian states is efficient (Theorem 4), we now ask whether there are relevant classes of non-Gaussian states for which tomography remains efficient. This leads us to analyse 't-doped Gaussian states': states prepared by applying Gaussian unitaries and at most t non-Gaussian local unitaries on the vacuum state. The set of all t-doped Gaussian states coincides with that of all pure Gaussian states for $t = 0$, it grows as t increases, and it becomes the set of all pure states for $t \to \infty$ [39]. In this sense, analysing the performance of tomography as a function of t allows us to understand the trade-off between the efficiency of tomography and the degree of 'non-Gaussianity'.

Let us proceed with the mathematical definitions. A unitary U is a *t-doped Gaussian* unitary if it is a composition

$$U = G_t W_t \cdots G_1 W_1 G_0 \,, \tag{2.6}$$

of Gaussian unitaries $G_0, G_1, \ldots, G_t$ and at most t (non-Gaussian) κ-local unitaries $W_1, \ldots, W_t$, as depicted in Fig. 2.3. A unitary is κ-local if it is generated by a Hamiltonian which is a polynomial

in at most κ quadratures. An n-mode state vector is a *t-doped Gaussian state vector*

$$|\psi\rangle = U\,|0\rangle^{\otimes n} \tag{2.7}$$

if it can be prepared by applying a t-doped Gaussian unitary U to the vacuum. The forthcoming theorem, proven in the SM, provides a remarkable decomposition of t-doped unitaries and states.

Theorem 5 (Compression of non-Gaussianity)**.** *If $n \geq \kappa t$, any n-mode t-doped Gaussian unitary U can be decomposed as*

$$U = G(u_{\kappa t} \otimes \mathbb{1}_{n-\kappa t})G_{passive}\,, \tag{2.8}$$

for some suitable Gaussian unitary G, energy-preserving Gaussian unitary $G_{passive}$ [7], and κt-mode (non-Gaussian) unitary $u_{\kappa t}$. In particular, any n-mode t-doped Gaussian state vector can be decomposed as

$$|\psi\rangle = G\left(|\phi_{\kappa t}\rangle \otimes |0\rangle^{\otimes(n-\kappa t)}\right)\,, \tag{2.9}$$

for some suitable Gaussian unitary G and κt-mode (non-Gaussian) state vector $|\phi_{\kappa t}\rangle$.

Eq. (2.9) establishes that all the non-Gaussianity of a t-doped Gaussian state can be compressed into $O(t)$ modes via a suitable Gaussian unitary. This result can be seen as a bosonic counterpart of recent results within the stabiliser [40–45] and fermionic setting [46], which reveals a fascinating parallelism among the theories of stabilisers, fermions, and bosons.

Leveraging the decomposition in (2.9), we design a simple and experimentally feasible tomography algorithm for t-doped Gaussian states. Our algorithm involves: (i) estimating the first moment and the covariance matrix to construct an estimation of the Gaussian unitary G; (ii) applying its inverse to the state to compress the non-Gaussianity into the first κt modes; and (iii) performing full-state tomography on the first κt modes (see the SM for details). This algorithm is practical because it only requires tools commonly available in quantum optics laboratories, such as Gaussian evolutions and easily implementable Gaussian measurements like homodyne and heterodyne detection [7]. Specifically, step (iii) can be achieved using the CV classical shadow algorithm [17], which is experimentally feasible. Alternatively, in order to obtain a tighter upper bound on the sample complexity of tomography of t-doped Gaussian states, step (iii) can be performed using the optimal full-state tomography algorithm identified in Theorem 1.

The following Theorem 6, proven in the SM, analyses the performance of our tomography algorithm. On a technical note, deriving guarantees on the estimation of the covariance matrix in step (i) requires more than just an energy constraint; a second-moment constraint is necessary. Refer to (2.4) for the definition of second-moment constraint.

Theorem 6 (Tomography of t-doped Gaussian state)**.** *There exists an algorithm that exploits*

$$\mathrm{poly}(n) + O\Big((nE/\varepsilon)^{2\kappa t}\Big) \tag{2.10}$$

copies to achieve tomography of n-mode t-doped Gaussian states. Here, E is the second-moment constraint, ε is the trace-distance error, and κ is the locality of the non-Gaussian unitaries.

This theorem implies that the sample complexity of tomography of t-doped Gaussian states scales at most exponentially with κt, indicating that tomography becomes increasingly difficult as the degree of non-Gaussianity t increases. Additionally, both the time and memory complexity exhibit the same behaviour. Notably, this establishes the core finding of this section: *tomography of t-doped Gaussian states is efficient in the regime $\kappa t = O(1)$.*

The proof of Theorem 6 hinges on the fact that a t-doped state is *compressible*, meaning that it satisfies the decomposition in (2.9). Specifically, we can prove that tomography of compressible states is efficient *if and only if* $\kappa t = O(1)$. This contrasts with the stabiliser [40, 45] and fermionic setting [46], where tomography of compressible stabiliser/fermionic states is efficient if and only if $t = O(\log(n))$. The difference in our setting ultimately arises from the infinite-dimensional nature of CV states and the presence of energy constraints.

2.5 Discussion

Our work serves as bridge between the two fields of quantum learning theory and CV quantum information. We provided the first and at the same time exhaustive investigation of tomography of CV systems with guarantees on the trace-distance error.

First, we determined the sample complexity of tomography of energy-constrained pure states, which pinpoints the ultimate achievable performance of CV tomography. A new phenomenon, the 'extreme inefficiency' of CV tomography, emerged: *any* tomography algorithm for energy-constrained states must use a number of copies that dramatically scales at least as ε^{-2n}, where n is the number of modes and ε is the trace-distance error. This phenomenon, providing arresting fundamental limitations even for small n, is a unique feature of tomography of CV systems.

On a more positive note, we proved that tomography of Gaussian states is efficient. To establish this, we investigated a fundamental problem of CV quantum information: how the error in approximating the first moment and the covariance matrix of a Gaussian state propagates in the trace-distance error. Our solution introduces new tools of independent interest: simple, stringent bounds on the trace distance between two Gaussian states in terms of the norm distance between their first moments and covariance matrices.

Finally, we devised an experimentally-feasible algorithm that efficiently achieves tomography of t-doped Gaussian states for small t. This establishes that even if a few non-Gaussian local gates are applied to a Gaussian state, tomography of the resulting non-Gaussian state remains efficient. The main tool employed here is a new decomposition of t-doped Gaussian states, which shows that all the non-Gaussianity in the state can be compressed into only $O(t)$ modes via a Gaussian unitary.

2.6 Methods

Trace distance The *trace distance* between two quantum states ρ_1 and ρ_2 is defined as

$$d_{\mathrm{tr}}(\rho_1, \rho_2) \coloneqq \frac{1}{2}\|\rho_1 - \rho_2\|_1 \,, \tag{2.11}$$

where $\|A\|_1 \coloneqq \mathrm{Tr}\sqrt{A^\dagger A}$ denotes the trace norm. The trace distance is considered the most meaningful notion of distance between two states because of its operational meaning in terms of the optimal probability of discriminating between the two states having access to a single copy of the state (Holevo–Helstrom theorem [9, 10]). Consequently, in quantum information theory, the error in approximating a state is typically quantified using the trace distance.

Quantum state tomography In this section, we precisely formulate the problem of *quantum state tomography* [1], which forms the basis of our investigation. The basic setup is depicted in Fig. 2.4.

Problem 7 (Quantum state tomography)**.** *Let $\mathcal{S}$ be a set of quantum states. Consider $\varepsilon, \delta \in (0,1)$, and $N \in \mathbb{N}$. Let $\rho \in \mathcal{S}$ be an unknown quantum state. Given access to N copies of ρ, the goal is to provide a classical description of a quantum state $\tilde{\rho}$ such that*

$$\Pr[d_{\mathrm{tr}}(\tilde{\rho}, \rho) \leq \varepsilon] \geq 1 - \delta \,. \tag{2.12}$$

That is, with a probability $\geq 1-\delta$, the trace distance between $\tilde{\rho}$ and ρ is at most ε. Here, ε is called the trace-distance error, while δ is called the failure probability.

The *sample complexity*, the *time complexity*, and the *memory complexity* of tomography of states in $\mathcal{S}$ are defined as the minimum number of copies N, the minimum amount of classical and quantum computation time, and the minimum amount of classical memory, respectively, required

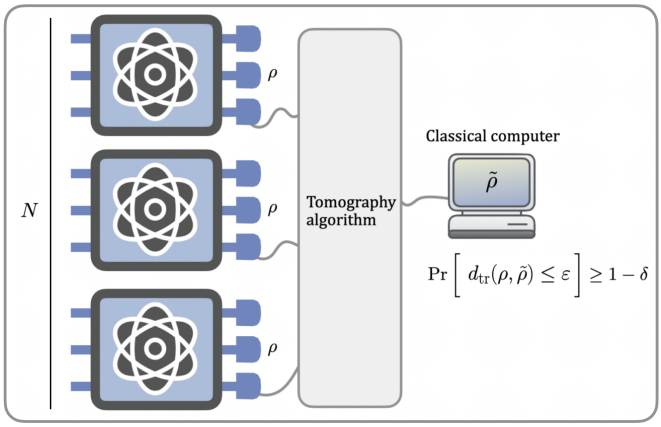


Figure 2.4: Pictorial representation of a quantum state tomography algorithm. Given access to N copies of an unknown state ρ, the goal of a tomography algorithm is to output a classical description of a state $\tilde{\rho}$ that serves as a 'good approximation' of the true unknown state ρ. Let us clarify what we mean by 'good approximation'. First, the error in this approximation is measured by the trace distance $d_{\mathrm{tr}}(\rho, \tilde{\rho})$, which is most meaningful notion of distance between states [9, 10]. Second, since quantum measurements yield probabilistic outcomes, the output $\tilde{\rho}$ is probabilistic rather than deterministic. Therefore, a 'good approximation' means that the probability of getting a small trace-distance error is high. In formula, this is expressed as $\Pr[d_{\mathrm{tr}}(\rho, \tilde{\rho}) \leq \varepsilon] \geq 1 - \delta$, where ε is the trace-distance error and δ is the failure probability. To measure the performance of tomography, one can define the *sample complexity*. For fixed ε and δ, the sample complexity is the minimum number of copies N required to achieve tomography with trace-distance error ε and failure probability δ. For example, the sample complexity of tomography of arbitrary n-qubit states is $O\left(\frac{4^n}{\varepsilon^2} \log(\frac{1}{\delta})\right)$ [1, 12–14]. In general, the sample complexity of any tomography task depends at most logarithmically on the failure probability δ [47], implying that this parameter has a minimal impact on the performance. In other words, the failure probability δ can be made very small with only a slight increase in the sample complexity. Therefore, throughout this work, we omit the dependence on δ.

to solve Problem 7 with trace distance error ε and failure probability δ. It is worth noting that the time complexity always upper bound the memory complexity, as well as the sample complexity.

One can think of $\mathcal{S}$ as a specific subset of the entire set of n-qubit states (e.g., pure states, r-rank states, stabiliser states) or n-mode states (e.g., energy-constrained states, moment-constrained states, Gaussian states, t-doped Gaussian states). By definition, tomography is deemed *efficient* if the sample, time, and memory complexity scales polynomial in n; otherwise, it is deemed *inefficient*. For example, in our work, we prove that tomography of energy-constrained states is (extremely) inefficient. In contrast, tomography of Gaussian states is efficient, while tomography of t-doped Gaussian states is efficient for small t and inefficient for large t.

Continuous-variable systems In this section, we provide a concise overview of quantum information with continuous variable (CV) systems [7]. A CV system is a quantum system associated with the Hilbert space $L^2(\mathbb{R}^n)$ of all square-integrable complex-valued functions over $\mathbb{R}^n$, which models n modes of electromagnetic radiation with definite frequency and polarisation. A quantum state on $L^2(\mathbb{R}^n)$ is called an *n-mode* state, and a unitary operator on $L^2(\mathbb{R}^n)$ is called an *n-mode* unitary. The *quadrature vector* is defined as

$$\hat{\mathbf{R}} \coloneqq (\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n)^{\intercal}\,, \tag{2.13}$$

where $\hat{x}_j$ and $\hat{p}_j$ are the well-known position and momentum operators of the j-th mode, collectively called *quadratures*. Let us proceed with the definitions of Gaussian unitary and Gaussian state.

Definition 8 (Gaussian unitary)**.** *An n-mode unitary is Gaussian if it is the composition of unitaries generated by quadratic Hamiltonians $\hat{H}$ in the quadrature vector:*

$$\hat{H} \coloneqq \frac{1}{2}(\hat{\mathbf{R}} - \mathbf{m})^{\intercal} h (\hat{\mathbf{R}} - \mathbf{m}) \,, \tag{2.14}$$

for some symmetric matrix $h \in \mathbb{R}^{2n,2n}$ and some vector $\mathbf{m} \in \mathbb{R}^{2n}$.

Definition 9 (Gaussian state)**.** *An n-mode state ρ is* Gaussian *if it can be written as a Gibbs state of a quadratic Hamiltonian $\hat{H}$ of the form in (2.14) with h being positive definite. The Gibbs states associated with the Hamiltonian $\hat{H}$ are given by*

$$\rho = \left(\frac{e^{-\beta \hat{H}}}{\mathrm{Tr}[e^{-\beta \hat{H}}]} \right)_{\beta \in [0,\infty]} , \tag{2.15}$$

where the parameter β is called the 'inverse temperature'.

This definition includes also the pathological cases where both β and certain terms of H diverge (e.g., this is the case for tensor products between pure Gaussian states and mixed Gaussian states). An example of Gaussian state is the *vacuum*, denoted as $|0\rangle^{\otimes n}$. Any pure n-mode Gaussian state vector

$$|\psi\rangle = G\,|0\rangle^{\otimes n} \tag{2.16}$$

can be written as a Gaussian unitary G applied to the vacuum. A Gaussian state ρ is uniquely identified by its *first moment* $\mathbf{m}(\rho)$ and *covariance matrix* $V(\rho)$. By definition, the first moment and the covariance matrix of an n-mode state ρ are given by

$$\begin{aligned} \mathbf{m}(\rho) &\coloneqq \mathrm{Tr}\left[\hat{\mathbf{R}}\,\rho\right] \,, \\ V(\rho) &\coloneqq \mathrm{Tr}\left[\left\{(\hat{\mathbf{R}} - \mathbf{m}(\rho)), (\hat{\mathbf{R}} - \mathbf{m}(\rho))^{\intercal}\right\}\rho\right] \,, \end{aligned} \tag{2.17}$$

where $\{\hat{A}, \hat{B}\} \coloneqq \hat{A}\hat{B} + \hat{B}\hat{A}$ is the anti-commutator.

By definition, the *energy* of an n-mode state ρ is given by the expectation value $\mathrm{Tr}[\hat{E}_n \rho]$ of the energy observable $\hat{E}_n \coloneqq \sum_{j=1}^{n} \left(\frac{\hat{x}_j^2}{2} + \frac{\hat{p}_j^2}{2} \right)$, where it is assumed that each mode has a frequency of one [7]. It is important to note that energy is an extensive quantity, because for any single-mode state σ the energy of $\sigma^{\otimes n}$ equals the energy of σ multiplied by n. Furthermore, the energy of an n-mode state is always greater than or equal to $\frac{n}{2}$, with the equality achieved only by the vacuum. The *total photon number* can be defined in terms of the energy observable as

$$\hat{N}_n \coloneqq \hat{E}_n - \frac{n}{2}\hat{\mathbb{1}} \tag{2.18}$$

Given $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{N}^n$, let us denote as

$$|\mathbf{k}\rangle = |k_1\rangle \otimes \ldots \otimes |k_n\rangle \tag{2.19}$$

the n-mode *Fock state* [7]. The total photon number is diagonal in the Fock basis:

$$\hat{N}_n = \sum_{\mathbf{k} \in \mathbb{N}^n} \|\mathbf{k}\|_1 \, |\mathbf{k}\rangle\langle\mathbf{k}| \,, \tag{2.20}$$

where $\|\mathbf{k}\|_1 \coloneqq \sum_{i=1}^{n} k_i$.

Effective dimension and rank of energy-constrained states In this section, we show that energy-constrained states can be well approximated by finite-dimensional states with low rank, as anticipated above in (2.3). Further technical details regarding the findings presented in this section can be found in the SM.

Let ρ be an n-mode state with total photon number satisfying the energy constraint

$$\mathrm{Tr}[\rho \hat{N}_n] \leq nE\,, \tag{2.21}$$

where $E \geq 0$. Given $M \in \mathbb{N}$, let $\mathcal{H}_M$ be the subspace spanned by all the n-mode Fock states with total photon number not exceeding M, and let Π_M the projector onto this space.

Let us begin by analysing the *effective dimension* of the set of energy-constrained states. The trace distance between the energy-constrained state ρ and its projection ρ_M onto $\mathcal{H}_M$, i.e.

$$\rho_M := \frac{\Pi_M \rho \Pi_M}{\mathrm{Tr}[\Pi_M \rho]}\,, \tag{2.22}$$

can be upper bounded as follows:

$$d_{\mathrm{tr}}(\rho, \rho_M) \overset{\text{(i)}}{\leq} \sqrt{\mathrm{Tr}[(\mathbb{1} - \Pi_M)\rho]} \overset{\text{(ii)}}{\leq} \sqrt{\frac{\mathrm{Tr}[\hat{N}_n \rho]}{M}} \leq \sqrt{\frac{nE}{M}}\,, \tag{2.23}$$

where in (i) we employed the Gentle measurement lemma [48] and in (ii) we used the simple operator inequality $\mathbb{1} - \Pi_M \leq \frac{\hat{N}_n}{M}$. Consequently, by setting $M_1 := \lceil nE/\varepsilon^2 \rceil$, it follows that the projection ρ_{M_1} is ε-close to ρ in trace distance. Moreover, the dimension of $\mathcal{H}_{M_1}$ can be upper bounded as

$$\dim \mathcal{H}_{M_1} = \binom{n+M_1}{n} \leq \left(\frac{e(n+M_1)}{n}\right)^n = O\left(\frac{(eE)^n}{\varepsilon^{2n}}\right), \tag{2.24}$$

where e denotes the Euler's number. Hence, we conclude that any energy-constrained state ρ can be approximated, up to trace-distance error ϵ, by its projection ρ_{M_1} onto the subspace $\mathcal{H}_{M_1}$, which has a *finite* dimension of $O((eE)^n/\varepsilon^{2n})$.

Now, let us analyse the *effective rank* of the energy-constrained state ρ. We say that ρ has effective rank r if it is ε-close to a state with rank r. Let us consider the spectral decomposition

$$\rho = \sum_{i=1}^{\infty} p_i^{\downarrow} \psi_i\,, \tag{2.25}$$

where the eigenvalues $(p_i^{\downarrow})_i$ are not increasing in i. In order to estimate the effective rank, let us choose an integer r such that

$$\sum_{i=r+1}^{\infty} p_i^{\downarrow} \leq \varepsilon\,, \tag{2.26}$$

which guarantees that the r-rank state $\rho^{(r)} \propto \sum_{i=1}^{r} p_i^{\downarrow} \psi_i$ is $O(\varepsilon)$-close to ρ. The *infinite-dimensional Schur Horn theorem* [49, Proposition 6.4] implies that for any r-rank projector Π it holds that

$$\sum_{i=r+1}^{\infty} p_i^{\downarrow} \leq \mathrm{Tr}[(\mathbb{1} - \Pi)\rho]\,. \tag{2.27}$$

Moreover, by setting $M_2 := \lceil nE/\varepsilon \rceil$, the projector Π_{M_2} is a $O((eE)^n/\varepsilon^n)$-rank projector satisfying

$$\mathrm{Tr}[(\mathbb{1} - \Pi_{M_2})\rho] \leq \frac{\mathrm{Tr}[\hat{N}_n \rho]}{M_2} \leq \frac{nE}{M} \leq \varepsilon\,, \tag{2.28}$$

where we employed the same inequalities used in (2.23) and (2.24). Hence, by setting $\Pi = \Pi_{M_2}$, we deduce that ρ is ε-close to a state $\rho^{(r)}$ having rank

$$r = O((eE)^n/\varepsilon^n)\,. \tag{2.29}$$

Finally, by exploiting Gentle measurement lemma [48] and triangle inequality, one can easily show that the projection of $\rho^{(r)}$ onto $\mathcal{H}_{M_1}$ is still $O(\varepsilon)$-close to ρ. Consequently, we conclude that any energy-constrained state can be approximated, up to trace-distance error ε, by a D-dimensional state with rank r such that

$$\begin{aligned} D &= O\left((eE)^n/\varepsilon^{2n}\right) , \\ r &= O\left((eE)^n/\varepsilon^n\right) . \end{aligned} \tag{2.30}$$

Based on these observations, we can devise a simple tomography algorithm for energy-constrained states. The first step involves performing the POVM $(\Pi_{M_1}, \mathbb{1} - \Pi_{M_1})$ and discarding the post-outcome state associated with $\mathbb{1} - \Pi_{M_1}$. This step transforms the unknown state ρ into the state ρ_{M_1} with high probability. The state ρ_{M_1} has two key properties: (i) it resides in the finite-dimensional subspace $\mathcal{H}_{M_1}$ of dimension $D = O\left((eE)^n/\varepsilon^{2n}\right)$, and (ii) it is $O(\varepsilon)$-close to a state residing in $\mathcal{H}_{M_1}$ with rank $r = O\left((eE)^n/\varepsilon^n\right)$. The second step involves performing the tomography algorithm of [50] designed for D-dimensional state with rank r, which has a sample complexity of $O(Dr)$. Importantly, this algorithm remains effective even if the unknown state, which is promised to reside in a given D-dimensional Hilbert space, has rank strictly larger than r, as long as it is $O(\varepsilon)$-close to a r-rank state within the *same* Hilbert space [50]. We thus conclude that the sample complexity of tomography of energy-constrained states is upper bounded by $O(Dr) = O\left((eE)^{2n}/\varepsilon^{3n}\right)$.

Analogously, by exploiting that the sample complexity of tomography of D-dimensional *pure* states is $O(D)$ [1, 11–14], we can show that the sample complexity of tomography of energy-constrained *pure* states is upper bounded by $O(D) = O\left((eE)^n/\varepsilon^{2n}\right)$.

For the sake of completeness, let us mention that the proof of the *lower* bounds on the sample complexity of tomography of energy-constrained states, as presented in Theorems 1-2, primarily relies on *epsilon-net* tools [51]. Detailed proofs of these results can be found in the SM.

Bounds on the trace distance between Gaussian states In this section, we address the question: 'if we know with a certain precision the first moment and the covariance matrix of an unknown Gaussian state, what is the resulting trace-distance error that we make on the state?'

Let us formalise the problem. Let us consider a Gaussian state ρ_1 and assume that we have an approximation of its first moment $\mathbf{m}(\rho_1)$ and an approximation of its covariance matrix $V(\rho_1)$. For example, these approximations may be retrieved through homodyne detection on many copies of ρ_1. We can then consider the Gaussian state ρ_2 with first moment and covariance matrix equal to such approximations: $\mathbf{m}(\rho_2)$ and $V(\rho_2)$ are thus the approximations of $\mathbf{m}(\rho_1)$ and $V(\rho_1)$, respectively. The errors incurred in these approximations are naturally measured by the norm distances $\|\mathbf{m}(\rho_1) - \mathbf{m}(\rho_2)\|$ and $\|V(\rho_1) - V(\rho_2)\|$, respectively, where $\|\cdot\|$ denotes some norm. Now, a natural question arises: 'given an error ε in the approximations of the first moment and covariance matrix, what is the error incurred in the approximation of ρ_1?' The most meaningful way to measure such an error is given by the trace distance $d_{\mathrm{tr}}(\rho_1, \rho_2)$ [9, 10]. Hence, the question becomes: 'if it holds that

$$\begin{aligned} \|\mathbf{m}(\rho_1) - \mathbf{m}(\rho_2)\| &= O(\varepsilon) , \\ \|V(\rho_1) - V(\rho_2)\| &= O(\varepsilon) , \end{aligned} \tag{2.31}$$

what can we say about the trace-distance $d_{\mathrm{tr}}(\rho_1, \rho_2)$?' Thanks to Theorems 10-11 below, we can answer this question: the trace distance $d_{\mathrm{tr}}(\rho_1, \rho_2)$ is at most $O(\sqrt{\varepsilon})$ and at least $O(\varepsilon)$.

This motivates the problem of finding upper and lower bounds on the trace distance between Gaussian states in terms of the norm distance of their first moments and covariance matrices. Now, we present our bounds, which constitutes technical tools of independent interest.

The forthcoming Theorem 10, proven in the SM, shows our upper bound on the trace distance between Gaussian states.

Theorem 10 (Upper bound on the distance between Gaussian states)**.** *Let ρ_1 and ρ_2 be n-mode Gaussian states satisfying the energy constraint* $\mathrm{Tr}[\hat{N}_n\rho_1], \mathrm{Tr}[\hat{N}_n\rho_2] \leq N$*. Then,*

$$d_{\mathrm{tr}}(\rho_1,\rho_2) \leq f(N)\Big(\|\boldsymbol{m}(\rho_1)-\boldsymbol{m}(\rho_2)\| + \sqrt{2}\sqrt{\|V(\rho_1)-V(\rho_2)\|_1}\Big)\,, \tag{2.32}$$

where $f(N) \coloneqq \frac{1}{\sqrt{2}}(\sqrt{N}+\sqrt{N+1})$*. Here,* $\|\boldsymbol{m}\| \coloneqq \sqrt{\boldsymbol{m}^\intercal\boldsymbol{m}}$ *and* $\|\cdot\|_1$ *denote the Euclidean norm and the trace norm, respectively.*

The above theorem turns out to be crucial to prove the upper bound on the sample complexity of tomography of Gaussian states provided in Theorem 4 in the Main text.

One might believe that proving Theorem 10 would be straightforward by bounding the trace distance using the closed formula for the fidelity between Gaussian states [52]. However, this approach turns out to be highly non-trivial due to the complexity of such fidelity formula [52], which makes it challenging to derive a bound based on the norm distance between the first moments and covariance matrices. Instead, our proof directly addresses the trace distance without relying on fidelity and involves a meticulous analysis based on the energy-constrained diamond norm [53].

The following theorem, proven in the SM, establishes our lower bound on the trace distance between Gaussian states.

Theorem 11 (Lower bound on the distance between Gaussian states)**.** *Let ρ_1 and ρ_2 be n-mode Gaussian states satisfying the energy constraint* $\mathrm{Tr}[\hat{E}_n\rho_1], \mathrm{Tr}[\hat{E}_n\rho_2] \leq E$*. Then,*

$$\begin{aligned} d_{\mathrm{tr}}(\rho_1,\rho_2) &\geq \frac{1}{200}\min\left\{1, \frac{\|\boldsymbol{m}(\rho_1)-\boldsymbol{m}(\rho_2)\|}{\sqrt{4E+1}}\right\}\,, \\ d_{\mathrm{tr}}(\rho_1,\rho_2) &\geq \frac{1}{200}\min\left\{1, \frac{\|V(\rho_2)-V(\rho_1)\|_2}{4E+1}\right\}\,, \end{aligned} \tag{2.33}$$

where $\|\boldsymbol{m}\| \coloneqq \sqrt{\boldsymbol{m}^\intercal\boldsymbol{m}}$ *and* $\|V\|_2 \coloneqq \sqrt{\mathrm{Tr}[V^\intercal V]}$ *denote the Euclidean norm and the Hilbert-Schmidt norm, respectively.*

The proof of this theorem heavily relies on state-of-the-art bounds recently established for Gaussian probability distributions [54].

Theorems 10-11 allow us to answer the question posed at the beginning of this section. Indeed, these theorems imply that, if we know with error ε the first moment and the covariance matrix of an unknown Gaussian state, the resulting trace-distance error that we make on the state is *at most* $O(\sqrt{\varepsilon})$ and *at least* $O(\varepsilon)$. In particular, this proves Theorem 3 in the Main text.

The trace-distance bound of Theorem 10 can be improved by assuming one of the Gaussian states to be pure, as we detail in the following theorem, proven in the SM.

Theorem 12 (Improved bound for pure states)**.** *Let ψ be a pure n-mode Gaussian state and let ρ be an n-mode (possibly non-Gaussian) state satisfying the energy constraints* $\mathrm{Tr}[\psi\hat{E}_n], \mathrm{Tr}[\rho\hat{E}_n] \leq E$*. Then*

$$d_{\mathrm{tr}}(\rho,\psi) \leq \sqrt{E}\sqrt{2\|\boldsymbol{m}(\rho)-\boldsymbol{m}(\psi)\|^2 + \|V(\rho)-V(\psi)\|_\infty}\,, \tag{2.34}$$

where $\|\boldsymbol{m}\| \coloneqq \sqrt{\boldsymbol{m}^\intercal\boldsymbol{m}}$ *and* $\|\cdot\|_\infty$ *denote the Euclidean norm and the operator norm, respectively.*

By exploiting this improved bound, in the SM we show that tomography of *pure* Gaussian states can be achieved using $O(n^5E^3/\varepsilon^4)$ copies of the state. This represents an improvement over the mixed-state scenario analysed in Theorem 4 in the Main text.

Moreover, the bound in Theorem 12 can be useful for *quantum state certification* [55], as we briefly detail now. Suppose one aims to prepare a pure Gaussian state ψ with known first and covariance matrix. In a noisy experimental setup, however, an unknown state ρ is effectively

prepared. By accurately estimating the first two moments of ρ (which can be done efficiently, as shown in the SM) one can estimate the right-hand side of (4.3), which provides an upper bound on the trace distance between the target state ψ and the noisy state ρ, thereby providing a measure of the quantum device's precision. Consequently, the device can be adjusted to minimise the error in the state preparation.

2.7 Supplementary Material

2.7.1 Preliminaries

Notation and basics

Let $\mathbb{N}$ denote the set of natural numbers, and for each $n \in \mathbb{N}$, define $[n] \coloneqq \{1, 2, \ldots, n\}$. We also define $\mathbb{N}_+ \coloneqq \mathbb{N} \setminus \{0\}$ and $\mathbb{R}_+$ as the set of positive real numbers. We introduce $\mathbb{R}^{n\times n}$ and $\mathbb{C}^{n\times n}$ as the sets of $n \times n$ real and complex matrices, respectively. The notation $\lceil x \rceil$ rounds $x \in \mathbb{R}$ up to the nearest integer, while $\lfloor x \rfloor$ rounds x down to the nearest integer. Given a vector $v \in \mathbb{C}^n$ and a scalar $p \in [1, \infty]$, the p-norm of v is denoted by $\|v\|_p$, defined as

$$\|v\|_p \coloneqq \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}. \tag{2.35}$$

For any matrix $A \in \mathbb{C}^{n\times n}$, its Schatten p-norm is given by $\|A\|_p \coloneqq \operatorname{Tr}\left((\sqrt{A^\dagger A})^p\right)^{\frac{1}{p}}$, which corresponds to the p-norm of the singular values of A. The trace norm and the Hilbert-Schmidt norm, instances of Schatten p-norms, are denoted $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively. The infinity norm, $\|\cdot\|_\infty$, represents the maximum singular value and equals the limit of the Schatten p-norms as $p \to \infty$. The Hölder inequality,

$$|\operatorname{Tr}(A^\dagger B)| \leq \|A\|_p \|B\|_q, \tag{2.36}$$

applies for $1 \leq p, q \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$. Moreover, for all matrices $A \in \mathbb{C}^{n\times n}$ and $1 \leq p \leq q$, it holds that $\|A\|_q \leq \|A\|_p$ and $\|A\|_p \leq \operatorname{rank}(A)^{\left(\frac{1}{p}-\frac{1}{q}\right)}\|A\|_q$. We use the bra-ket notation, where we denote a vector $v \in \mathbb{C}^d$ using the ket notation $|v\rangle$ and its adjoint using the bra notation $\langle v|$. We refer to a vector $|\psi\rangle \in \mathbb{C}^d$ as a (pure) state if $\| |\psi\rangle \|_2 = 1$.

Given a Hilbert space $\mathcal{H}$, we denote with $\mathcal{D}(\mathcal{H})$ the set of quantum states on $\mathcal{H}$, i.e., positive semi-definite operators with unit trace. The trace distance between two quantum states ρ and σ is defined by $\frac{1}{2}\|\rho - \sigma\|_1$. The von Neumann entropy of a quantum state ρ is given by $S(\rho) \coloneqq -\operatorname{Tr}[\rho \log_2 \rho]$.

We denote with $O(n)$ the group of $n \times n$ orthogonal matrices, and with $\mathrm{U}(n)$ the group of $n \times n$ unitary matrices. $\mathrm{Sp}(2n)$ denotes the group of symplectic matrices over the real field, defined as

$$\mathrm{Sp}(2n) \coloneqq \{S \in \mathbb{R}^{2n,2n} \,:\, S\Omega S^T = \Omega\}, \tag{2.37}$$

where

$$\Omega_n \coloneqq \bigoplus_{i=1}^{n} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.38}$$

We denote the Fourier transform $\mathscr{F}$ of a function $g : \mathbb{R}^{2n} \to \mathbb{R}$ as

$$\mathscr{F} g(\mathbf{r}) = \int_{\mathbf{r}' \in \mathbb{R}^{2n}} \mathrm{d}^{2n}\mathbf{r}'\, g(\mathbf{r}') e^{-i\mathbf{r}^\intercal \mathbf{r}'}. \tag{2.39}$$

Here, $\mathbf{r}^\intercal$ denotes the transpose of $\mathbf{r}$. Consequently, the inverse Fourier transform $\mathscr{F}^{-1}$ of a function $g : \mathbb{R}^{2n} \to \mathbb{R}$ is given by

$$\mathscr{F}^{-1} g(\mathbf{r}) = \frac{1}{(2\pi)^{2n}} \int_{\mathbf{r}' \in \mathbb{R}^{2n}} \mathrm{d}^{2n}\mathbf{r}'\, g(\mathbf{r}') e^{i\mathbf{r}'^\intercal \mathbf{r}}. \tag{2.40}$$

For any random variable Z and any convex function f, Jensen Inequality states that $f(\mathbb{E}[Z]) \leq \mathbb{E}[f(Z)]$. Furthermore, if f is concave, then the opposite inequality holds.
For any concave real function f, any Hermitian matrix X with eigenvalues pertaining to the domain of f, and any positive semi-definite matrix ρ, it holds that

$$\mathrm{Tr}(\rho f(X)) \leq f(\mathrm{Tr}(\rho X)), \tag{2.41}$$

which can be shown by expanding X in its eigendecomposition and using Jensen inequality.

We review some basic notions regarding the asymptotic notation:

- Big-O notation: For a function $f(n)$, if there exists a constant c and a specific input size n_0 such that $f(n) \leq c \cdot g(n)$ for all $n \geq n_0$, where $g(n)$ is a well-defined function, then we express it as $f(n) = O(g(n))$. This notation signifies the upper limit of how fast a function grows in relation to $g(n)$.
- Big-Omega notation: For a function $f(n)$, if there exists a constant c and a specific input size n_0 such that $f(n) \geq c \cdot g(n)$ for all $n \geq n_0$, where $g(n)$ is a well-defined function, then we express it as $f(n) = \Omega(g(n))$. This notation signifies the lower limit of how fast a function grows in relation to $g(n)$.
- Big-Theta notation: For a function $f(n)$, if $f(n) = O(g(n))$ and if $f(n) = \Omega(g(n))$, where $g(n)$ is a well-defined function, then we express it as $f(n) = \Theta(g(n))$.

A tilde over $\mathcal{O}(\cdot)$, i.e., $\tilde{\mathcal{O}}(\cdot)$, implies that we are neglecting n factors at the numerator or denominator (for us, n will always represent the number of modes or qubits). For example, $f(n) = \frac{2^n}{n}$ is $\tilde{T}heta\,(2^n)$. Analogously for the other asymptotic functions.

Note: Throughout the Supplementary Material, we will use the asymptotic notation described above. However, in the main text, we opted to simplify the notation by just using $O(\cdot)$ to informally denote the "relevant asymptotic scaling." Specifically, when in the Main text we wrote that $y = O(f(\varepsilon, n))$ for some function $f(\cdot)$ of ε and n, we informally meant that "y is of the order of $f(\varepsilon, n)$ when n is very large and ε is very small." Importantly, to ensure clarity and avoid ambiguity, we express every sample complexity bound using both asymptotic notation and explicit exact expressions.

Basics of statistical learning theory

We present here basic results of probability and statistical learning theory; for further details we refer to Ref. [51].

Lemma 13 (Union bound)**.** *Let $B_1, B_2, \ldots, B_M$ be events in a probability space. The probability of their union is bounded by the sum of their individual probabilities,*

$$\Pr\left[\bigcup_{i=1}^{M} B_i\right] \leq \sum_{i=1}^{M} \Pr[B_i]\,. \tag{2.42}$$

Lemma 14 (Markov inequality)**.** *Let X be a non-negative random variable and $a > 0$. Then, the probability that X is at least a is bounded by the expected value of X divided by a, i.e.*

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}. \tag{2.43}$$

Lemma 15 (Chernoff bound)**.** *Consider a set of independent and identically distributed binary random variables $\{Y_i\}_{i=1}^N$, taking values in $\{0, 1\}$. Define $Y := \sum_{i=1}^N Y_i$ and $\mu_Y := \mathbb{E}[Y]$. For any $\alpha \in (0, 1)$, the probability of Y being less than $(1 - \alpha)$ times its expected value is exponentially bounded as*

$$\Pr\left[Y \leq (1-\alpha)\mu_Y\right] \leq \exp\left(-\frac{\alpha^2 \mu_Y}{2}\right). \tag{2.44}$$

We are now going to define the median-of-means estimator [51, 56, 57]. Let $N', K \in \mathbb{N}$. Given $N = N'K$ samples $\{X_m\}_{m=1}^N$ of the random variable X, divide the samples into K disjoint bins $\{\mathcal{B}_l\}_{l=1}^K$. Specifically, for each $l \in [K]$, let $\mathcal{B}_l$ be a set containing the elements $\{X_{(l-1)N'+1}, \dots, X_{lN'}\}$, where X_i denotes the i-th ordered sample. For each bin $\mathcal{B}_l$, define $\tilde{x}_l$ as the arithmetic average, i.e.

$$\tilde{x}_l := \frac{1}{N'} \sum_{X_m \in \mathcal{B}_l} X_m. \tag{2.45}$$

The median-of-means estimator is then given by

$$\hat{\mu}(N', K) := \text{median}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_K). \tag{2.46}$$

We now present a Lemma which will be crucial in our subsequent analysis.

Lemma 16 (Medians of means [56, 57])**.** *Let X be a random variable with variance σ^2. Suppose K independent sample means of size $N' \geq \frac{34\sigma^2}{\varepsilon^2}$ suffice to construct a median-of-means estimator $\hat{\mu}(N, K)$ that satisfies*

$$\Pr[|\hat{\mu}(N, K) - \mathbb{E}[X]| \geq \varepsilon] \leq 2\mathrm{e}^{-K/2}, \quad \forall \varepsilon > 0\,. \tag{2.47}$$

As a consequence, $N \geq 68 \log\left(\frac{2}{\delta}\right) \frac{\sigma^2}{\varepsilon^2}$ samples of X suffice to construct a median-of-means estimator $\hat{\mu}$ which satisfies

$$\Pr[|\hat{\mu} - \mathbb{E}[X]| \geq \varepsilon] \leq \delta \quad \forall \varepsilon > 0\,. \tag{2.48}$$

We also mention the following (standard) fact that is useful in amplifying the probability of success of an algorithm.

Lemma 17 (Enhancing the probability of success)**.** *Let $\mathcal{A}$ be an algorithm with a success probability of at least $p_{\text{succ}} \in (0, 1]$. Let $\delta > 0$ and $N' \in \mathbb{N}$. If we execute $\mathcal{A}$ a total of m times, with*

$$m \geq \left\lceil \frac{3}{2p_{\text{succ}}} N' + \frac{18}{p_{\text{succ}}} \log\left(\frac{1}{\delta}\right) \right\rceil, \tag{2.49}$$

then $\mathcal{A}$ will achieve success at least N' times with a probability of at least $1 - \delta$.

Proof. Consider the binary random variables $\{X_i\}_{i=1}^m$ defined as

$$X_i := \begin{cases} 1 & \text{if } \mathcal{A} \text{ succeeds,} \\ 0 & \text{if } \mathcal{A} \text{ fails.} \end{cases} \tag{2.50}$$

Let $\hat{X} := \sum_{i=1}^m X_i$, and note that $\mathbb{E}[\hat{X}] = mp_{\text{succ}}$. Our goal is to upper bound the probability that $\mathcal{A}$ succeeds fewer than N' times. Define $\alpha := 1 - \frac{N'}{mp_{\text{succ}}}$. Using that $m \geq \frac{3}{2p_{\text{succ}}} N'$, we ensure that $\alpha \in \left(\frac{1}{3}, 1\right)$. Applying the Chernoff bound in Lemma 15, we get:

$$\begin{aligned} \Pr\left(\hat{X} \leq N'\right) &= \Pr\left(\hat{X} \leq (1-\alpha)\mathbb{E}[\hat{X}]\right) \\ &\leq \exp\left(-\frac{\alpha^2 \mathbb{E}[\hat{X}]}{2}\right) \\ &= \exp\left(-\frac{\alpha^2}{2} p_{\text{succ}} m\right) \\ &\leq \exp\left(-\frac{p_{\text{succ}}}{18} m\right) \\ &\leq \delta, \end{aligned} \tag{2.51}$$

where in the second inequality we have used that $\alpha \geq \frac{1}{3}$, while in the last inequality we have used

$$m \geq \frac{18}{p_{\text{succ}}} \log\left(\frac{1}{\delta}\right). \tag{2.52}$$

□

We present a lemma that enhances the probability of success of a learning algorithm. Although the proof follows standard steps similar to those in Ref. [47] (Proposition 2.4), we provide it here with precise constants, which were not explicitly stated in Ref. [47].

Lemma 18 (Enhancing the probability of success of an algorithm for learning objects in a metric space)**.** *Let $p_{\text{succ}} \in (\frac{1}{2}, 1]$ and $\varepsilon > 0$. Let $\mathcal{A}$ be an algorithm for learning unknown objects in a metric space with distance $d(\cdot,\cdot)$. Assume that for any (unknown) input object ρ the algorithm $\mathcal{A}$ outputs an object $\tilde{\rho}$ such that*

$$\Pr\left[d(\tilde{\rho}, \rho) \leq \varepsilon\right] \geq p_{\text{succ}}\,. \tag{2.53}$$

Then for any $\delta \in (0, 1]$ there exists an algorithm $\mathcal{A}'$, which executes $\mathcal{A}$ a total of m times with

$$m := \left\lceil \frac{2}{\left(1 - \frac{1}{2p_{\text{succ}}}\right)^2 p_{\text{succ}}} \log\left(\frac{1}{\delta}\right) \right\rceil , \tag{2.54}$$

such that for any (unknown) input object ρ it outputs an object $\tilde{\rho}$ such that

$$\Pr\left[d(\tilde{\rho}, \rho) \leq 3\varepsilon\right] \geq 1 - \delta\,. \tag{2.55}$$

Proof. The algorithm $\mathcal{A}'$, by executing $\mathcal{A}$ a total of m times, produces m random objects $\tilde{\rho}_1, \ldots, \tilde{\rho}_m$ such that each $\tilde{\rho}_i$ satisfies $\Pr\left[d(\tilde{\rho}_i, \rho) \leq \varepsilon\right] \geq p_{\text{succ}}$. Let us show that the probability $\bar{P}$ that there are more than $\frac{m}{2}$ objects in the set $\mathcal{S} := \{\tilde{\rho}_1, \ldots, \tilde{\rho}_m\}$ which are close at most ε to ρ is not smaller than $1 - \delta$. Consider the binary random variables $\{X_i\}_{i=1}^m$ defined as

$$X_i := \begin{cases} 1 & \text{if } d(\tilde{\rho}_i, \rho) \leq \varepsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{2.56}$$

Let $\hat{X} := \sum_{i=1}^m X_i$, and note that $\mathbb{E}[\hat{X}] = m p_{\text{succ}}$. Using that $p_{\text{succ}} \in (\frac{1}{2}, 1)$, we ensure that the quantity $\alpha := 1 - \frac{1}{2p_{\text{succ}}}$ satisfies $\alpha \in (0, 1)$. Then, by applying the Chernoff bound in Lemma 15, we have that

$$\begin{aligned} \bar{P} &= \Pr\left(\hat{X} > \frac{m}{2}\right) \\ &= 1 - \Pr\left(\hat{X} \leq \frac{m}{2}\right) \\ &= 1 - \Pr\left(\hat{X} \leq (1-\alpha)\mathbb{E}[\hat{X}]\right) \\ &\geq 1 - \exp\left(-\frac{\alpha^2 \mathbb{E}[\hat{X}]}{2}\right) \\ &= 1 - \exp\left(-\frac{\left(1 - \frac{1}{2p_{\text{succ}}}\right)^2 p_{\text{succ}}}{2} m\right) \\ &\geq 1 - \delta\,, \end{aligned} \tag{2.57}$$

where in the last inequality we have used that

$$m \geq \frac{2}{\left(1 - \frac{1}{2p_{\text{succ}}}\right)^2 p_{\text{succ}}} \log\left(\frac{1}{\delta}\right). \tag{2.58}$$

From now on, let us assume that there are more than $\frac{m}{2}$ objects in $\mathcal{S}$ which are close at most ε to ρ (which is an event that happens with probability at least $1 - \delta$, as proved above). Then, by the triangle inequality, there are more than $\frac{m}{2}$ objects in $\mathcal{S}$ which are close at most 2ε to each other.

The algorithm $\mathscr{A}'$ is as follows: First, it computes the distance between any two objects in $\mathcal{S}$ and, second, it outputs an object $\tilde{\rho} \in \mathcal{S}$ that satisfies the property of being close at most 2ε to more than $\frac{m}{2}$ objects in $\mathcal{S}$.

Let us show that the output $\tilde{\rho}$ satisfies $d(\tilde{\rho}, \rho) \leq 3\varepsilon$. Let $\mathcal{S}_{\tilde{\rho}}$ be the set of objects in $\mathcal{S}$ that are close at most 2ε to $\tilde{\rho}$. Since $|\mathcal{S}_{\tilde{\rho}}| > \frac{m}{2}$, $|\mathcal{S}| = m$, and $\mathcal{S}$ contains more than $\frac{m}{2}$ objects which are close at most ε to ρ, then there exists an object $\sigma \in \mathcal{S}_{\tilde{\rho}}$ with $d(\sigma, \rho) \leq \varepsilon$. Then, by the triangle inequality, we conclude that $d(\tilde{\rho}, \rho) \leq 3\varepsilon$. □

The right-hand side of (2.54) diverges as $p_{\text{succ}} \to \frac{1}{2}$. This is consistent with the simple fact that, in general, the probability of success can not be arbitrarily enhanced when $p_{\text{succ}} \leq \frac{1}{2}$. Indeed, in the latter case, we could consider an algorithm $\mathscr{A}$ such that for any unknown input ρ it outputs: the exact true object ρ with probability p_{succ}; an object σ_ρ, which is very distant from ρ, with the same probability p_{succ}; and a fixed object τ (independent of ρ) with probability $1 - 2p_{\text{succ}}$. Hence, since the objects ρ and σ_ρ are statistically indistinguishable, there is no way to arbitrarily enhance the probability of success of $\mathscr{A}$, even with infinite executions of it.

Preliminaries on continuous variable systems

In this section, we provide a concise overview of quantum information with *continuous variable* (CV) systems; for further details, we refer to Refs. [7, 58, 59]. We consider n modes of harmonic oscillators associated with the Hilbert space $L^2(\mathbb{R}^n)$, which comprises all square-integrable complex-valued functions over $\mathbb{R}^n$. Each mode represents a single-mode of electromagnetic radiation with definite frequency and polarisation. The set of n-mode states is denoted by $\mathcal{D}(L^2(\mathbb{R}^n))$. For each $j \in [n]$, the annihilation operator a_j of the j-th mode is defined by

$$a_j := \frac{\hat{x}_j + i\hat{p}_j}{\sqrt{2}}\,, \tag{2.59}$$

where $\hat{x}_j$ and $\hat{p}_j$ denote the well-known position and momentum operators of the j-th mode, which are Hermitian operator satisfying the canonical commutation relations $[\hat{x}_j, \hat{p}_k] = i\delta_{j,k}\hat{\mathbb{1}}$. Given a single mode with annihilation operator a, its m-th Fock state vector (corresponding to the quantum state vector with m photons) is defined as

$$|m\rangle := \frac{(a^\dagger)^m}{\sqrt{m!}}\,|0\rangle\,, \tag{2.60}$$

where $|0\rangle$ is the vacuum state vector. Crucially, the Fock states $(|m\rangle)_{n\in\mathbb{N}}$ form an orthonormal basis for the Hilbert space $L^2(\mathbb{R})$ of a single-mode system, meaning that such a system can be effectively viewed as an infinite-dimensional qudit. The operator $a^\dagger a$ is referred to as the photon number operator and it can be diagonalised in Fock basis as

$$a^\dagger a = \sum_{m=0}^{\infty} m\,|m\rangle\langle m|\,. \tag{2.61}$$

By introducing the quadrature vector

$$\hat{\mathbf{R}} := (\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n)^\intercal = (\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_{2n-1}, \hat{R}_{2n})^\intercal \tag{2.62}$$

the canonical commutation relations can be expressed as

$$[\hat{R}_k, \hat{R}_l] = i\,(\Omega_n)_{kl}\hat{\mathbb{1}} \qquad \forall\, k, l \in [2n]\,, \tag{2.63}$$

where

$$\Omega_n := \bigoplus_{i=1}^{n} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \mathbb{1}_n \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.64}$$

is the n-mode symplectic form, $\mathbb{1}_n$ is the $n \times n$ identity matrix, and $\hat{\mathbb{1}}$ is the identity operator over $L^2(\mathbb{R}^n)$. The relation in (2.63) is usually expressed in the continuous variable literature [7] in a compact form as

$$[\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal] = i\,\Omega_n \hat{\mathbb{1}}\,, \tag{2.65}$$

which we denote as 'vectorial notation'. The energy operator $\hat{E}_n$ is defined as

$$\hat{E}_n \coloneqq \frac{1}{2}\hat{\mathbf{R}}^\intercal\hat{\mathbf{R}} = \sum_{j=1}^{n}\left(\frac{\hat{x}_j^2}{2} + \frac{\hat{p}_j^2}{2}\right) = \sum_{j=1}^{n}\left(a_j^\dagger a_j + \frac{\hat{\mathbb{1}}}{2}\right) = \hat{N}_n + \frac{n}{2}\hat{\mathbb{1}}\,, \tag{2.66}$$

where $\hat{N}_n \coloneqq \sum_{j=1}^{n} a_j^\dagger a_j$ is the photon number operator. The characteristic function $\chi_\rho : \mathbb{R}^{2n} \to \mathbb{C}$ of an n-mode state $\rho \in \mathcal{D}(L^2(\mathbb{R}^n))$ is defined as $\chi_\rho(\mathbf{r}) \coloneqq \mathrm{Tr}[\rho\hat{D}_\mathbf{r}]$, where for all $\mathbf{r} \in \mathbb{R}^{2n}$ the displacement operator $\hat{D}_\mathbf{r}$ is given by $\hat{D}_\mathbf{r} \coloneqq e^{-i\mathbf{r}^\intercal\Omega_n\hat{\mathbf{R}}}$. Any state ρ can be written in terms of its characteristic function as [7]

$$\rho = \int_{\mathbb{R}^{2n}} \frac{\mathrm{d}^{2n}\mathbf{r}}{(2\pi)^n}\chi_\rho(\mathbf{r})\hat{D}_\mathbf{r}^\dagger\,, \tag{2.67}$$

and hence quantum states and characteristic functions are in one-to-one correspondence.

The Wigner function $W_\rho : \mathbb{R}^{2n} \to \mathbb{R}$ of an n-mode state ρ is defined as the inverse Fourier transform $\mathscr{F}^{-1}$ of the characteristic function χ_ρ, evaluated at $\Omega_n\mathbf{r}$

$$W_\rho(\mathbf{r}) = \mathscr{F}^{-1}\chi_\rho(\Omega_n\mathbf{r}) = \frac{1}{(2\pi)^{2n}}\int_{\mathbf{r}'\in\mathbb{R}^{2n}} \mathrm{d}^{2n}\mathbf{r}'\,\chi_\rho(\mathbf{r}')e^{i\mathbf{r}'^\intercal\Omega_n\mathbf{r}}. \tag{2.68}$$

Consequently, the characteristic function χ_ρ can be expressed as the Fourier transform

$$\chi_\rho(\mathbf{r}) = \mathscr{F}W_\rho(\Omega_n\mathbf{r}) = \int_{\mathbf{r}'\in\mathbb{R}^{2n}} \mathrm{d}^{2n}\mathbf{r}'\,W_\rho(\mathbf{r}')e^{-i\mathbf{r}^\intercal\Omega_n\mathbf{r}'}, \tag{2.69}$$

evaluated at $\Omega_n\mathbf{r}$, of the Wigner function W_ρ. Moreover, the Husimi function $Q_\rho(\mathbf{r}) : \mathbb{R}^{2n} \to \mathbb{R}$ of an n-mode state ρ is defined as

$$Q_\rho(\mathbf{r}) \coloneqq \frac{1}{(2\pi)^n}\left\langle\mathbf{r}\right|\rho\left|\mathbf{r}\right\rangle\,, \tag{2.70}$$

where $\left|\mathbf{r}\right\rangle \coloneqq \hat{D}_\mathbf{r}\left|0\right\rangle$ is a coherent state vector. It turns out that the Fourier transform of the Husimi function, evaluated at $\Omega_n\mathbf{r}$, is related to the characteristic function as [7]

$$\int_{\mathbf{r}'\in\mathbb{R}^{2n}} \mathrm{d}^{2n}\mathbf{r}'\,Q_\rho(\mathbf{r}')e^{-i\mathbf{r}^\intercal\Omega_n\mathbf{r}'} = e^{-\frac{1}{4}\mathbf{r}^\intercal\mathbf{r}}\chi_\rho(\mathbf{r})\,. \tag{2.71}$$

The Husimi function $Q_\rho(\mathbf{r})$ is a useful quantity since it is the probability distribution of the outcome $\mathbf{r} \in \mathbb{R}^{2n}$ of a (experimentally feasible) measurement — known as *heterodyne measurement* — performed on the state ρ [7]. The first moment of a quantum state ρ is defined as $\mathbf{m}(\rho) \coloneqq (m_1(\rho), \ldots, m_n(\rho))^\intercal$, where

$$m_k(\rho) \coloneqq \mathrm{Tr}\left[\hat{R}_k\,\rho\right]\,, \tag{2.72}$$

for each $k \in [n]$, or in its vectorial notation as $\mathbf{m}(\rho) = \mathrm{Tr}\left[\hat{\mathbf{R}}\,\rho\right]$. Additionally, the covariance matrix of ρ is defined by the matrix $V(\rho)$ with elements

$$[V(\rho)]_{k,l} \coloneqq \mathrm{Tr}\left[\left\{\hat{R}_k - m_k(\rho)\hat{\mathbb{1}}, \hat{R}_l - m_l(\rho)\hat{\mathbb{1}}\right\}\rho\right] = \mathrm{Tr}\left[\left\{\hat{R}_k, \hat{R}_l\right\}\rho\right] - 2m_k(\rho)m_l(\rho)\,, \tag{2.73}$$

for each $k, l \in [2n]$, where $\{\hat{A}, \hat{B}\} \coloneqq \hat{A}\hat{B} + \hat{B}\hat{A}$ is the anti-commutator. In its vectorial notation this reads as

$$V(\rho) = \mathrm{Tr}\left[\left\{(\hat{\mathbf{R}} - \mathbf{m}(\rho)), (\hat{\mathbf{R}} - \mathbf{m}(\rho))^\intercal\right\}\rho\right] = \mathrm{Tr}\left[\left\{\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal\right\}\rho\right] - 2\mathbf{m}(\rho)\mathbf{m}(\rho)^\intercal\,. \tag{2.74}$$

Any covariance matrix $V(\rho)$ satisfies the inequality

$$V(\rho) + i\Omega_n \geq 0\,, \tag{2.75}$$

known as *uncertainty relation*. As a consequence, since Ω_n is skew-symmetric, any covariance matrix $V(\rho)$ is positive semi-definite on $\mathbb{R}^{2n}$. Conversely, for any symmetric $W \in \mathbb{R}^{2n,2n}$ such that $W + i\Omega_n \geq 0$ there exists an n-mode (Gaussian) state ρ with covariance matrix $V(\rho) = W$ [7].

Definition 19 (Gaussian state)**.** *An n-mode state ρ is said to be a Gaussian state if it can be written as a Gibbs state of a quadratic Hamiltonian $\hat{H}$ in the quadrature operators $\{\hat{R}_i\}_{i\in[2n]}$, that is,*

$$\hat{H} \coloneqq \frac{1}{2}(\hat{\mathbf{R}} - \mathbf{m})^\intercal H(\hat{\mathbf{R}} - \mathbf{m}) \tag{2.76}$$

for some symmetric positive-definite matrix $H \in \mathbb{R}^{2n,2n}$ and some vector $\mathbf{m} \in \mathbb{R}^{2n}$. The Gibbs states associated with the Hamiltonian $\hat{H}$ are given by

$$\rho = \left(\frac{e^{-\beta\hat{H}}}{\mathrm{Tr}[e^{-\beta\hat{H}}]}\right)_{\beta\in[0,\infty]}\,, \tag{2.77}$$

where the parameter β is called the 'inverse temperature'.

Remark 20. *Definition 103 includes also the pathological cases where both β and certain terms of H diverge (e.g., this is the case for tensor products between pure Gaussian states and mixed Gaussian states). To formalise this mathematically, one can define the set of Gaussian states as the closure, with respect to the trace norm, of the set of Gibbs states of quadratic Hamiltonians [60].*

The characteristic function of a Gaussian state ρ is the Fourier transform of a Gaussian probability distribution, evaluated at $\Omega_n\mathbf{r}$, which can be written in terms of $\mathbf{m}(\rho)$ and $V(\rho)$ as [7]

$$\chi_\rho(\mathbf{r}) = \exp\!\left(-\frac{1}{4}(\Omega_n\mathbf{r})^\intercal V(\rho)\Omega_n\mathbf{r} + i(\Omega_n\mathbf{r})^\intercal\mathbf{m}(\rho)\right)\,. \tag{2.78}$$

Consequently, the Wigner function of a Gaussian state can be expressed as the following Gaussian probability distribution:

$$W_\rho(\mathbf{r}) = \frac{e^{-(\mathbf{r}-\mathbf{m}(\rho))^\intercal[V(\rho)]^{-1}(\mathbf{r}-\mathbf{m}(\rho))}}{\pi^n\sqrt{\det[V(\rho)]}}\,. \tag{2.79}$$

Note that a Gaussian probability distribution with first moment $\mathbf{m}$ and covariance matrix V is defined as

$$\mathcal{N}[\mathbf{m}, V](\mathbf{r}) \coloneqq \frac{e^{-\frac{1}{2}(\mathbf{r}-\mathbf{m})^\intercal V^{-1}(\mathbf{r}-\mathbf{m})}}{(2\pi)^n\sqrt{\det V}}\,. \tag{2.80}$$

Hence, using such notation, the Wigner function of a Gaussian state ρ is a Gaussian probability distribution with first moment $\mathbf{m}(\rho)$ and covariance matrix $\frac{V(\rho)}{2}$, that is,

$$W_\rho(\mathbf{r}) = \mathcal{N}\left[\mathbf{m}(\rho), \frac{V(\rho)}{2}\right](\mathbf{r})\,. \tag{2.81}$$

Hence, (2.71) and (2.78) imply that the Husimi function of a Gaussian state ρ is a Gaussian probability distribution with first moment $\mathbf{m}(\rho)$ and covariance matrix $\frac{V(\rho)+\mathbb{1}}{2}$, that is,

$$Q_\rho(\mathbf{r}) = \mathcal{N}\left[\mathbf{m}(\rho), \frac{V(\rho)+\mathbb{1}}{2}\right](\mathbf{r})\,. \tag{2.82}$$

Since any quantum state is uniquely identified by its characteristic function, it follows from (2.78) that any Gaussian state is uniquely identified by its first moment and covariance matrix. An example of Gaussian state is the single-mode thermal state

$$\tau_\nu := \frac{1}{\nu+1}\sum_{n=0}^{\infty}\left(\frac{\nu}{\nu+1}\right)^n |n\rangle\langle n| \,, \tag{2.83}$$

where $\nu \geq 0$. It holds that $\mathrm{Tr}[a^\dagger a\, \tau_\nu] = \nu$, thus ν is the mean photon number of τ_ν. The first moment and the covariance matrix of τ_ν satisfy

$$\begin{aligned} \mathbf{m}(\tau_\nu) &= (0,0)^\intercal \,, \\ V(\tau_\nu) &= (2\nu+1)\mathbb{1}_2 \,. \end{aligned} \tag{2.84}$$

It is worth noting that the thermal state with $\nu = 0$ is the vacuum state vector, $\tau_0 = |0\rangle\langle 0|$. Thermal states are important since they maximise the von Neumann entropy among all states with a fixed mean photon number, as established by Lemma 21 [61].

Lemma 21 (Extremality of thermal states)**.** *For all $\nu > 0$, the maximum von Neumann entropy among all n-mode states with a given mean photon number ν is achieved by a thermal state $\tau_{\nu/n}^{\otimes n}$, i.e.*

$$\max\left\{S(\rho) :\, \rho \in \mathcal{D}(L^2(\mathbb{R}^n)),\ \mathrm{Tr}\left[\rho\sum_{i=1}^{n} a_i^\dagger a_i\right] \leq \nu\right\} = S\left(\tau_{\nu/n}^{\otimes n}\right) = ng\left(\frac{\nu}{n}\right) \,, \tag{2.85}$$

where

$$g(\nu) := (\nu+1)\log_2(\nu+1) - \nu\log_2\nu \tag{2.86}$$

is a monotone increasing function called the bosonic entropy.

In the setting of continuous variable systems, symplectic matrices play a crucial role. Recall that a matrix $S \in \mathbb{R}^{2n,2n}$ is *symplectic* matrix if and only if $S\Omega_n S^\intercal = \Omega_n$ and that the group of symplectic matrices is denoted by $\mathrm{Sp}(2n)$. Fixed a symplectic matrix $S \in \mathrm{Sp}(2n)$, one can define a suitable n-mode unitary U_S — dubbed *symplectic unitary* or the *metaplectic representation* of S — such that, for each $k \in [2n]$ it holds

$$U_S^\dagger \hat{R}_k U_S = \sum_{k,l}^{2n} S_{k,l}\hat{R}_l. \tag{2.87}$$

In the vectorial notation, we can write this as

$$U_S^\dagger \hat{\mathbf{R}} U_S = S\hat{\mathbf{R}} \,. \tag{2.88}$$

More explicitly, the symplectic unitary U_S is defined in terms of the symplectic matrix S as follows [7]. Any symplectic matrix S can be written as $S = S_1 S_2$, where $S_1 := e^{\Omega_n H_1}$ and $S_2 := e^{\Omega_n H_2}$ for some symmetric matrices H_1, H_2 [7]. Then, U_S is defined as $U_S := U_{S_1}U_{S_2}$, where

$$\begin{aligned} U_{S_1} &:= e^{-\frac{i}{2}\hat{\mathbf{R}}^\intercal H_1 \hat{\mathbf{R}}} \,, \\ U_{S_2} &:= e^{-\frac{i}{2}\hat{\mathbf{R}}^\intercal H_2 \hat{\mathbf{R}}} \,. \end{aligned} \tag{2.89}$$

In particular, (2.88) implies that

$$\begin{aligned} \mathbf{m}(U_S \rho U_S^\dagger) &= S\mathbf{m}(\rho) \,, \\ V(U_S \rho U_S^\dagger) &= SV(\rho)S^\intercal \,. \end{aligned} \tag{2.90}$$

Given two symplectic matrices S_1 and S_2, it holds that $U_{S_1}U_{S_2} = U_{S_1 S_2}$. In addition, the displacement operator $\hat{D}_{\mathbf{r}}$ transforms the quadrature vector as

$$\hat{D}_{\mathbf{r}} \hat{\mathbf{R}} \hat{D}_{\mathbf{r}}^\dagger = \hat{\mathbf{R}} + \mathbf{r}\hat{\mathbb{1}}, \tag{2.91}$$

which implies that

$$\begin{aligned}\mathbf{m}\left(\hat{D}_{\mathbf{r}}\rho\hat{D}_{\mathbf{r}}^{\dagger}\right) &= \mathbf{m}(\rho) + \mathbf{r}\,,\\ V\left(\hat{D}_{\mathbf{r}}\rho\hat{D}_{\mathbf{r}}^{\dagger}\right) &= V(\rho)\,.\end{aligned} \tag{2.92}$$

Notably, any covariance matrix $V(\rho)$ of an n-mode state ρ satisfies the following decomposition, known as the Williamson decomposition [7]: there exists a symplectic matrix $S \in \mathrm{Sp}(2n)$ and n real numbers $d_1, d_2, \ldots, d_n \geq 1$, called the symplectic eigenvalues of $V(\rho)$, such that

$$V(\rho) = SDS^{\intercal}\,, \tag{2.93}$$

where

$$D = \bigoplus_{j=1}^{n}\left(d_j\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right) = \mathrm{diag}\left(d_1, d_1, d_2, d_2, \ldots, d_n, d_n\right)\,. \tag{2.94}$$

In addition, if ρ is Gaussian, it is possible to show that the above Williamson decomposition of the covariance matrix leads to the decomposition

$$\rho = \hat{D}_{\mathbf{m}(\rho)}U_S\left(\tau_{\nu_1}\otimes\tau_{\nu_2}\otimes\ldots\tau_{\nu_n}\right)U_S^{\dagger}\hat{D}_{\mathbf{m}(\rho)}^{\dagger} \tag{2.95}$$

for the state ρ, where $\{\tau_{\nu_k}\}_{k=1}^{n}$ are thermal states defined in Eq. (2.83), and the mean photon numbers $\nu_1, \nu_2, \ldots, \nu_n$ are defined in terms of the symplectic eigenvalues of $V(\rho)$ via the relation $\nu_i := \frac{d_i-1}{2}$ for all $i \in [n]$. In other words, any Gaussian state is unitarily equivalent — through displacement and symplectic unitaries — to a multi-mode thermal state. In particular, any pure Gaussian state vector $|\phi\rangle$ can be written as

$$|\phi\rangle = \hat{D}_{\mathbf{r}}U_S\,|0\rangle^{\otimes n}\,. \tag{2.96}$$

Moreover, Gaussian unitaries can be defined as follows.

Definition 22 (Gaussian unitary)**.** *An n-mode unitary is said to be Gaussian if it is the composition of unitaries generated by quadratic Hamiltonians. A Hamiltonian $\hat{H}$ is said to be quadratic if it can be written as $\hat{H} := \frac{1}{2}(\hat{\mathbf{R}}-\mathbf{m})^{\intercal}H(\hat{\mathbf{R}}-\mathbf{m})$, where $\mathbf{r} \in \mathbb{R}^{2n}$ and $H \in \mathbb{R}^{2n,2n}$ is a symmetric matrix.*

In particular, for any n-mode Gaussian unitary U there exists a symplectic matrix $S \in \mathrm{Sp}(2n)$ and a vector $\mathbf{r} \in \mathbb{R}^{2n}$ such that $U = \hat{D}_{\mathbf{r}}U_S$, where U_S is the symplectic unitary associated with S. Conversely, for any $S \in \mathrm{Sp}(2n)$ and $\mathbf{r} \in \mathbb{R}^{2n}$, the unitary $\hat{D}_{\mathbf{r}}U_S$ is a Gaussian unitary. In other words, Gaussian unitaries are the composition of displacement and symplectic unitaries.

Notably, any symplectic matrix $S \in \mathrm{Sp}(2n)$ can be written in the so-called Euler (or Bloch-Messiah) decomposition as follows: there exist symplectic orthogonal matrices $O_1, O_2 \in \mathrm{O}(2n) \cap \mathrm{Sp}(2n)$ and real numbers $z_1, z_1, \ldots, z_n \geq 1$ such that

$$S = O_1 Z O_2 \tag{2.97}$$

with

$$Z := \bigoplus_{j=1}^{n}\begin{pmatrix}z_j & 0\\ 0 & z_j^{-1}\end{pmatrix}\,. \tag{2.98}$$

Symplectic unitaries associated with symplectic matrices of the form of Z are said to be *squeezing* unitaries. Moreover, symplectic unitaries associated with symplectic orthogonal matrices are said to be *passive*. Importantly, passive unitaries preserve the energy operator, meaning that given $O \in \mathrm{O}(2n) \cap \mathrm{Sp}(2n)$ then

$$U_O^{\dagger}\hat{E}_n U_O = \hat{E}_n\,. \tag{2.99}$$

Moreover, note that the Euler decomposition in (2.97) implies that any Gaussian unitary can be written as a composition of passive unitaries and squeezing unitary, i.e., $U_S = U_{O_1}U_Z U_{O_2}$.

Finally, let us briefly mention two measurements that are presumably experimentally most feasible to implement in common quantum optics laboratories: *heterodyne* and *homodyne* measurements [7]. Heterodyne measurement is a POVM whose elements are given by all the coherent states, suitably normalised. Specifically, these POVM elements are $\{\frac{1}{(2\pi)^n}\mathrm{d}^{2n}\mathbf{r}\,|\mathbf{r}\rangle\langle\mathbf{r}|\}_{r\in\mathbb{R}^{2n}}$, which sum up to the identity thanks to the over-complete relation of coherent states [7] according to

$$\frac{1}{(2\pi)^n}\int_{\mathbb{R}^{2n}}\mathrm{d}^{2n}\mathbf{r}\,|\mathbf{r}\rangle\langle\mathbf{r}| = \mathbb{1}\,. \tag{2.100}$$

Moreover, a homodyne measurement entails measuring a quadrature observable, which is a component of either the quadrature operator vector $\hat{\mathbf{R}}$ or, more generally, a rotated version $U_S^\dagger\hat{\mathbf{R}}U_S = S\hat{\mathbf{R}}$, where U_S represents the symplectic unitary associated with the symplectic matrix S.

Preliminaries on quantum learning theory

In this section, we introduce the concept of *quantum state tomography*, which forms the basis of our investigation. We start by formulating the problem of quantum state tomography.

Problem 23 (Quantum state tomography)**.** *Let $n \in \mathbb{N}$ be the number of modes/qudits. Let $\mathcal{S} \subseteq \mathcal{D}(\mathcal{H}^{\otimes n})$ be a subset of the set of quantum states $\mathcal{D}(\mathcal{H}^{\otimes n})$. Consider $\varepsilon, \delta \in (0,1)$ and $N \in \mathbb{N}$. Let $\rho \in \mathcal{S}$ be an unknown quantum state. Given access to N copies of ρ, the goal is to provide a classical description of a quantum state $\tilde{\rho}$ such that*

$$\Pr\left(\frac{1}{2}\|\tilde{\rho}-\rho\|_1 \leq \varepsilon\right) \geq 1-\delta\,. \tag{2.101}$$

That is, with a probability $\geq 1-\delta$, the trace distance between $\tilde{\rho}$ and ρ is at most ε. Here, ε is called the accuracy, while δ is called the failure probability.

Definition 24 (Sample, time, and memory complexity of a tomography algorithm)**.** *A quantum state tomography algorithm is an algorithm that solves Problem 23. It is characterised by the sample complexity, i.e., the number of copies of the unknown state needed by the algorithm, represented by N in Problem 23. The time complexity of the tomography algorithm is the amount of classical and quantum computation time required to execute the algorithm. The memory complexity quantifies the amount of classical memory required by the algorithm.*

We say that a tomography algorithm is *efficient* if its sample, time, and memory complexity scale polynomially in the number of modes/qudits n. It is worth noting that the time complexity of an algorithm always upper bound its memory complexity, as well as its sample complexity. In Fig. 2.5 we provide a pictorial representation of a quantum state tomography algorithm.

In the literature, several tomography algorithms [11–13] are tailored to finite-dimensional rank-r quantum states $\rho \in \mathcal{D}(\mathbb{C}^D)$, where D is the dimension of the Hilbert space. In the case of n-qudits systems with local dimension d, then the dimension of the Hilbert space is $D = d^n$. If no assumptions on the states are known, then it is known that any tomography algorithm needs at least $\tilde{O}mega\left(rD/\varepsilon^2\right)$ copies of the unknown state to solve the tomography problem [13], where ε is the accuracy parameter. In this case, the tilde above the Big-O asymptotic functions means that we neglect logarithmic factors. There are algorithms that match these lower bounds, i.e., they require $\tilde{O}\left(rD/\varepsilon^2\right)$ copies of the unknown state to solve the tomography problem [12, 13]. Such algorithms that achieve optimal performances use entangled measurements between multiple copies of the unknown state. However, there are also algorithms, which might be more experimentally feasible, that use *unentangled* or *single-copies* measurements between the queried copies of the unknown state; in this case, a lower bound $\tilde{O}mega\left(r^2D/\varepsilon^2\right)$ is known to hold [62]. Also in this case, there are known algorithms which achieve optimal performances [63]. In the case of an

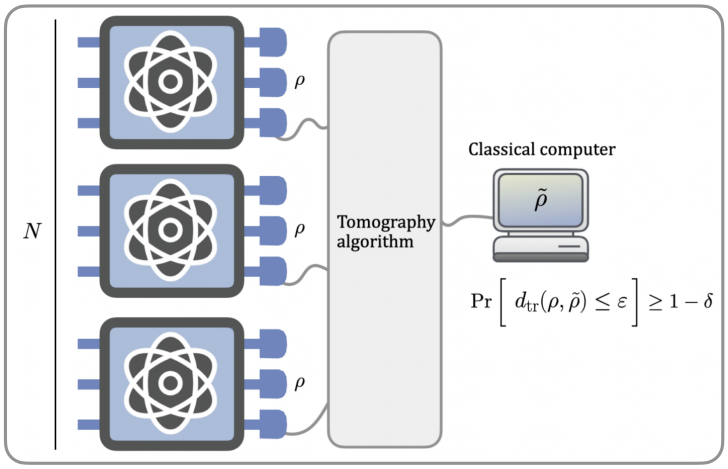


Figure 2.5: Pictorial representation of a quantum state tomography algorithm. Given access to N copies of an unknown state ρ, the goal of a tomography algorithm is to construct a classical description of a state $\tilde{\rho}$ that serves as a 'good approximation' of the true unknown state ρ. Mathematically, the error incurred in such an approximation is measured by the trace distance between ρ and $\tilde{\rho}$. This is the most meaningful way to measure the error incurred in a tomography algorithm, due to the operational meaning of the trace distance given by the Holevo–Helstrom theorem [9, 10]. Additionally, since quantum measurements inherently yield probabilistic outcomes, the output $\tilde{\rho}$ is probabilistic rather than deterministic. We thus require that the probability that 'the trace distance is small' is high. Mathematically, this translates to $\Pr\left[\frac{1}{2}\|\tilde{\rho} - \rho\|_1 \leq \varepsilon\right] \geq 1 - \delta$, where ε represents the trace distance error and δ denotes the failure probability. Fixed ε and δ, the minimum number of copies N required to achieve quantum state tomography with trace distance error ε and failure probability δ is called the sample complexity.

n-qudits system, all these algorithms necessarily scale exponentially with the number of qudits n. However, if further restrictions on the class of states are provided, then one might achieve a sample and time complexity scaling which is polynomial in n (or, equivalently, logarithmic in D). In the finite-dimensional case, only a few classes of states are currently known to be efficiently learnable, such as matrix product states [64], finitely-correlated states [65], states prepared by shallow quantum circuits [66], quantum phase states [67], stabiliser states [68], states prepared by Clifford circuits 'doped' with at most $O(\log(n))$ T-gates [40, 45, 69], fermionic Gaussian states [70], and states prepared by fermionic Gaussian circuits 'doped' with at most $O(\log(n))$ non-Gaussian gates [46]. For a detailed literature review of quantum learning theory, see Ref. [1].

Given a tomography algorithm, we define a shorthand for denoting its sample and time complexity that will be useful later.

Definition 25 (Sample and time complexity)**.** *Let $\rho \in \mathcal{D}(\mathbb{C}^D)$ be a D-dimensional quantum state where $D \in \mathbb{N}$. Let $\varepsilon, \delta \in (0, 1)$. Let $\mathcal{A}$ be a tomography algorithm. We denote as $N_{\mathrm{tom}}(\mathcal{A}, D, \varepsilon, \delta)$ and $T_{\mathrm{tom}}(\mathcal{A}, D, \varepsilon, \delta)$, respectively, the sample and time complexity of the algorithm $\mathcal{A}$ to solve the tomography problem of the state ρ with accuracy at most ε and a failure probability at most δ. That is, given access to $N_{\mathrm{tom}}(\mathcal{A}, D, \varepsilon, \delta)$ copies of ρ and using at most $T_{\mathrm{tom}}(\mathcal{A}, D, \varepsilon, \delta)$ computational time, the algorithm $\mathcal{A}$ outputs, with probability $\geq 1 - \delta$, a classical description of a state $\tilde{\rho}$ such that*

$$\frac{1}{2}\|\tilde{\rho} - \rho\|_1 \leq \varepsilon. \tag{2.102}$$

We now present a standard proof technique to provide lower bound on the sample complexity of a tomography algorithm (see, e.g., Ref. [14]), which relies on the use of Fano's inequality [71] and Holevo's bound [72].

Lemma 26 (Lower bound on the sample complexity)**.** *Let $M \in \mathbb{N}$ and $\varepsilon > 0$. Consider a distance metric $d(\cdot,\cdot)$ between quantum states (e.g., trace distance). Define the discrete set $\mathcal{M}_\varepsilon := \{\rho_1, \ldots, \rho_M\}$ as a set of quantum states such that for every $i \neq j \in [M]$, $d(\rho_i, \rho_j) > 2\varepsilon$. Let $\mathcal{S}$ be a subset of quantum states such that $\mathcal{M}_\varepsilon \subseteq \mathcal{S}$. Any tomography algorithm that learns states within the metric $d(\cdot,\cdot)$ from the set $\mathcal{S}$ with accuracy $\leq \varepsilon$ and failure probability $\leq \delta$ must use a number of state copies N satisfying:*

$$\chi \geq (1-\delta)\log_2(M) - H_2(\delta), \tag{2.103}$$

where $\chi := S\left(\frac{1}{M}\sum_{j=1}^M \rho_j^{\otimes N}\right) - \frac{1}{M}\sum_{j=1}^M S(\rho_j^{\otimes N})$ is the Holevo information, $S(\cdot)$ denotes the von Neumann entropy, and $H_2(x) := -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy. For instance, if the states in $\mathcal{M}_\varepsilon$ live in a 2^n dimensional Hilbert space, then this implies that the number of copies N satisfies

$$N \geq \Omega\left(\frac{\log_2(M)}{n}\right). \tag{2.104}$$

Proof. Consider a communication protocol between Alice and Bob. First, define a codebook that Alice will use to encode classical information in quantum states: each number $i \in [n]$ is associated with the quantum state $\rho_i \in \mathcal{M}_\varepsilon$. Alice samples uniformly a number $x \in [M]$. Let us denote the random variable associated to x with X. Then Alice prepares the state $\rho_x^{\otimes N}$ and sends it to Bob. Now Bob runs the tomography algorithms and learns the classical description of a matrix $\hat{\rho}$ such that $d(\rho_x, \hat{\rho}) \leq \varepsilon$ with failure probability $\leq \delta$. In case the tomography protocol succeeds, this would imply that for all $j \in [M]$ such that $j \neq x$, we have $d(\rho_j, \hat{\rho}) > \varepsilon$, because

$$d(\rho_j, \hat{\rho}) \geq d(\rho_j, \rho_x) - d(\rho_x, \hat{\rho}) > 2\varepsilon - \varepsilon = \varepsilon\,. \tag{2.105}$$

Thus, in such a case, there exists only one element in the set $\mathcal{M}_\varepsilon$ that is ε-close to $\hat{\rho}$ and this has to be ρ_x. Hence, by going through all the M states in $\mathcal{M}_\varepsilon$ and for each of them computing the trace distance with $\hat{\rho}$, Bob can find the number $x \in [M]$ corresponding to ρ_x (with probability $\geq 1-\delta$). Thus, at the ends of this procedure, Bob selects a number $y \in [M]$ and with probability $\geq 1-\delta$, it will coincide with x, i.e., the message sent by Alice. Therefore, the probability that Bob decodes the wrong message is $\leq \delta$. Call Y the random variable associated with y. By Fano's inequality (see Ref. [71]), we have

$$I(X:Y) \geq (1-\delta)\log_2(M) - H_2(\delta), \tag{2.106}$$

where $I(X:Y)$ is the mutual information (see, e.g. Ref. [71] for the standard definition) and $H_2(\cdot)$ is the binary entropy. By the data processing inequality [71], we have

$$I(X:Y) \leq I(X:Z), \tag{2.107}$$

where Z is the random variable associated with the POVM outcome associated with the tomography algorithm. By Holevo's theorem [72] (see also, e.g., Wikipedia), we have

$$I(X:Z) \leq S\left(\frac{1}{M}\sum_{j=1}^M \rho_j^{\otimes N}\right) - \frac{1}{M}\sum_{j=1}^M S(\rho_j^{\otimes N}) = \chi. \tag{2.108}$$

Hence, we get

$$\chi \geq (1-\delta)\log_2(M) - H_2(\delta). \tag{2.109}$$

If the states in $\mathcal{M}_\varepsilon$ live in a 2^n dimensional Hilbert space, then we have

$$\chi \leq S\left(\frac{1}{M}\sum_{j=1}^M \rho_j^{\otimes N}\right) \leq N\log(2^n). \tag{2.110}$$

This implies that

$$N \geq \frac{1}{\log_2(2^n)} \left((1-\delta)\log_2(M) - H_2(\delta) \right) = \Omega\left(\frac{\log_2(M)}{n} \right). \tag{2.111}$$

□

Preliminaries on ε-nets

In this section, we introduce key concepts related to ε-nets [51] before delving into our novel results on the topic. Let us begin by giving definitions relative to the notion of ε-covering net [51].

Definition 1 (ε-covering net/number)**.** Let $(T, \|\cdot\|)$ be a normed space with distance induced by the norm $\|\cdot\|$, $K \subseteq T$ be a subset, and let $\varepsilon > 0$. We define:

- **ε-covering net:** A subset $C \subseteq K$ is an ε-covering net of K if, for every $x \in K$, there exists $x_0 \in C$ such that $\|x - x_0\| \leq \varepsilon$.
- **ε-covering number:** The covering number of K, denoted as $\mathcal{C}(K, \|\cdot\|, \varepsilon)$, is the smallest cardinality of the ε-covering nets of K.
- **Optimal ε-covering:** $C \subseteq K$ is an optimal ε-covering of K if and only if $|C| = \mathcal{C}(K, \|\cdot\|, \varepsilon)$, where $|C|$ denotes the cardinality of C.

We now move to define definitions related to the notion of ε-packing net [51].

Definition 2 (ε-packing net/number)**.** Let $(T, \|\cdot\|)$ be a normed space with distance induced by the norm $\|\cdot\|$, $K \subseteq T$ be a subset, and let $\varepsilon > 0$. We define:

- **ε-packing net:** A subset $P \subseteq K$ is an *ε-packing net* of K if $\|x - y\| > \varepsilon$ for every $x, y \in P$.
- **ε-packing number:** The packing number of K, denoted as $\mathcal{P}(K, \|\cdot\|, \varepsilon)$, is the largest cardinality of the ε-packing nets of K.
- **Optimal ε-packing:** $P \subseteq K$ is is an optimal ε-covering of K if and only if $|P| = \mathcal{P}(K, \|\cdot\|, \varepsilon)$.

A result in the theory of ε-net [51] connects the ε-covering number $\mathcal{C}(K, \|\cdot\|, \varepsilon)$ to the ε-packing number $\mathcal{P}(K, \|\cdot\|, \varepsilon)$.

Lemma 27 (Ref. [51], Lemma 4.2.8) **.** *Let $(T, \|\cdot\|)$ be a normed space, $K \subseteq T$ be a subset, and let $\varepsilon > 0$. Then, it holds that*

$$\mathcal{P}(K, \|\cdot\|, \varepsilon) \geq \mathcal{C}(K, \|\cdot\|, \varepsilon) \geq \mathcal{P}(K, \|\cdot\|, 2\varepsilon)\,. \tag{2.112}$$

Our results on ε-nets

In this section we will present our results concerning ε-nets. We will begin by proving a lemma that relates the covering number of a $(d-1)$-dimensional sphere with unit radius to the covering number of the surface of a d-dimensional sphere with unit radius, where $d \in \mathbb{N}_+$.

Lemma 28 (Covering numbers)**.** *Consider the normed vector space $(\mathbb{R}^d, \|\cdot\|_2)$, where $d \in \mathbb{N}_+$. Let $B_1(d)$ be the d-dimensional unit ball and $\partial B_1(d)$ its boundary, defined as*

$$B_1(d) \coloneqq \{x \in \mathbb{R}^d \,:\, \|x\|_2 \leq 1\}, \tag{2.113}$$

$$\partial B_1(d) \coloneqq \{x \in \mathbb{R}^d \,:\, \|x\|_2 = 1\}. \tag{2.114}$$

For all $d \in \mathbb{N}_+$ and $\varepsilon > 0$, the covering numbers of $B_1(d-1)$ and $\partial B_1(d)$ are related as

$$\mathcal{C}(\partial B_1(d), \|\cdot\|_2, \varepsilon) \geq \mathcal{C}(B_1(d-1), \|\cdot\|_2, \varepsilon). \tag{2.115}$$

Proof. Consider an *optimal* ε-covering net of $\partial B_1(d)$, denote with $C^{\varepsilon,\mathrm{opt}}_{\partial B_1(d)}$. Note that for any $y \in B_1(d-1)$, the d-dimensional vector $v(y) \coloneqq \left(y, \sqrt{1-\|y\|_2^2}\right)$ satisfies $v(y) \in \partial B_1(d)$. Therefore, for any $y \in B_1(d-1)$ there exists $x \in C^{\varepsilon,\mathrm{opt}}_{\partial B_1(d)}$ such that

$$\|v(y) - (x_1, \ldots, x_d)\|_2 \leq \varepsilon. \tag{2.116}$$

Thus, we also have that

$$\begin{aligned} \|y - (x_1, \ldots, x_{d-1})\|_2 &\leq \sqrt{\|y - (x_1, \ldots, x_{d-1})\|_2^2 + (\sqrt{1-\|y\|_2^2} - x_d)^2} \\ &= \|v(y) - (x_1, \ldots, x_d)\|_2 \leq \varepsilon\,. \end{aligned} \tag{2.117}$$

This implies that the set

$$C^{\varepsilon}_{B_1(d-1)} \coloneqq \{(x_1, \ldots, x_{d-1}) \in \mathbb{R}^{d-1} \,|\, x \in C^{\varepsilon,\mathrm{opt}}_{\partial B_1(d)}\} \tag{2.118}$$

is an ε-covering net for $B_1(d-1)$. Thus, we have

$$\mathcal{C}(\partial B_1(d), \|\cdot\|_2, \varepsilon) = |C^{\varepsilon,\mathrm{opt}}_{\partial B_1(d)}| \geq |C^{\varepsilon}_{B_1(d-1)}| \geq \mathcal{C}(B_1(d-1), \|\cdot\|_2, \varepsilon) \tag{2.119}$$

where the first step follows from the definition of covering number, the second step follows from the fact that $C^{\varepsilon}_{B_1(d-1)}$ contains at most the same number of elements as $C^{\varepsilon,\mathrm{opt}}_{\partial B_1(d)}$, and the last inequality follows from the fact that the ε-covering number is a lower bound on the cardinality of every possible ε-covering net. This concludes the proof. □

We now present a lemma which relates the covering number of the set of pure states density matrices with respect to the (matrix) 1-norm and the covering number of the set of pure states vectors with respect to the (vector) 2-norm.

Lemma 29 (Covering numbers revisited)**.** *Let $d \in \mathbb{N}$ and $\varepsilon \in [0,1)$. We define the sets $\mathcal{S}_{|*\rangle\langle *|}$ and $\mathcal{S}_{|*\rangle}$, as*

$$\mathcal{S}_{|*\rangle\langle *|} \coloneqq \{|v\rangle\langle v| \,:\, |v\rangle \in \mathbb{C}^d \text{ such that } |\langle v|v\rangle| = 1\}, \tag{2.120}$$

$$\mathcal{S}_{|*\rangle} \coloneqq \{|v\rangle \,:\, |v\rangle \in \mathbb{C}^d \text{ such that } |\langle v|v\rangle| = 1\}. \tag{2.121}$$

It holds the following relation between their covering numbers

$$\mathcal{C}(\mathcal{S}_{|*\rangle\langle *|}, \|\cdot\|_1, \varepsilon) \geq \frac{\varepsilon}{8\pi}\mathcal{C}(\mathcal{S}_{|*\rangle}, \|\cdot\|_2, \varepsilon)\,. \tag{2.122}$$

Proof. Let $C^{\varepsilon,\mathrm{opt}}_{\mathcal{S}_{|*\rangle\langle *|}} = \{|\psi_1\rangle\langle\psi_1|, \ldots, |\psi_M\rangle\langle\psi_M|\}$ be an optimal ε-covering net of the set $\mathcal{S}_{|*\rangle\langle *|}$ with respect to the trace norm $\|\cdot\|_1$, where we defined $M \coloneqq \mathcal{C}(\mathcal{S}_{|*\rangle\langle *|}, \|\cdot\|_1, \varepsilon)$. For each $j \in [M]$, we take $|\phi_j\rangle$ to be a state such that $|\phi_j\rangle\langle\phi_j| = |\psi_j\rangle\langle\psi_j|$. Let $\delta > 0$ be a number that we will fix later in terms of ε. We define the set

$$C_\delta \coloneqq \left\{e^{i\delta m}\,|\phi_j\rangle \,:\, j \in [M],\, m \in \left\{0, 1, \ldots, \left\lfloor \frac{2\pi}{\delta} \right\rfloor\right\}\right\}. \tag{2.123}$$

Let us show that C_δ is an $(\frac{\varepsilon}{\sqrt{2}} + \delta)$-covering net of $\mathcal{S}_{|*\rangle}$ with respect to the 2-norm $\|\cdot\|_2$. To this end, fixed a state vector $|\psi\rangle \in \mathbb{C}^d$, it suffices to construct a state vector $|\tilde{\psi}\rangle \in C_\delta$ such that

$$\|\,|\psi\rangle - |\tilde{\psi}\rangle\,\|_2 \leq \frac{\varepsilon}{\sqrt{2}} + \delta\,. \tag{2.124}$$

By definition of $C^{\varepsilon,\mathrm{opt}}_{\mathcal{S}_{|*\rangle\langle *|}}$, there exists $j \in [M]$ such that $\| \, |\phi_j\rangle\langle\phi_j| - |\psi\rangle\langle\psi| \, \|_1 \leq \varepsilon$. Let $\bar{m} \coloneqq \lfloor \frac{2\pi - \arg(\langle\psi|\phi_j\rangle)}{\delta} \rfloor$, where $\arg(\langle\psi|\phi_j\rangle)$ is such that $\langle\psi|\phi_j\rangle = |\langle\psi|\phi_j\rangle| e^{i \arg(\langle\psi|\phi_j\rangle)}$, and let us define the state vector $|\tilde{\psi}\rangle \coloneqq e^{i\delta\bar{m}} |\phi_j\rangle \in C_\delta$. Let us now show that the so-defined $|\tilde{\psi}\rangle$ satisfies Eq. (2.124):

$$\begin{aligned}
\| \, |\psi\rangle - |\tilde{\psi}\rangle \, \|_2 &= \| \, |\psi\rangle - e^{i\delta\bar{m}} |\phi_j\rangle \, \|_2 \\
&\leq \| \, |\psi\rangle - e^{-i \arg(\langle\psi|\phi_j\rangle)} |\phi_j\rangle \, \|_2 + \| e^{-i \arg(\langle\psi|\phi_j\rangle)} |\phi_j\rangle - e^{i\delta\bar{m}} |\phi_j\rangle \, \|_2 \\
&\overset{\text{(i)}}{\leq} \frac{1}{\sqrt{2}} \| \, |\psi\rangle\langle\psi| - |\phi_j\rangle\langle\phi_j| \, \|_1 + \| e^{-i \arg(\langle\psi|\phi_j\rangle)} |\phi_j\rangle - e^{i\delta\bar{m}} |\phi_j\rangle \, \|_2 \\
&\leq \frac{\varepsilon}{\sqrt{2}} + |e^{i(2\pi - \arg(\langle\psi|\phi_j\rangle))} - e^{i\delta\bar{m}}| \\
&\overset{\text{(ii)}}{\leq} \frac{\varepsilon}{\sqrt{2}} + |2\pi - \arg(\langle\psi|\phi_j\rangle) - \delta\bar{m}| \\
&\leq \frac{\varepsilon}{\sqrt{2}} + \delta \, .
\end{aligned} \tag{2.125}$$

Here, (i) follows from the fact that

$$\begin{aligned}
\left\| |\psi\rangle - e^{-i \arg(\langle\psi|\phi_j\rangle)} |\phi_j\rangle \right\|_2 &= \sqrt{2 \left(1 - \mathrm{Re} \left(e^{-i \arg(\langle\psi|\phi_j\rangle)} \langle\psi|\phi_j\rangle \right) \right)} \\
&= \sqrt{2 \left(1 - |\langle\psi|\phi_j\rangle| \right)} \\
&\leq \sqrt{2 \left(1 - |\langle\psi|\phi_j\rangle|^2 \right)} \\
&= \frac{1}{\sqrt{2}} \left\| \, |\psi\rangle\langle\psi| - |\phi_j\rangle\langle\phi_j| \, \right\|_1 ,
\end{aligned} \tag{2.126}$$

where in the last step we have used the well-known formula for the trace distance between two pure states (see, e.g. Ref. [73]). Moreover, (ii) is a consequence of the fact that for any $x, y \in \mathbb{R}$ it holds that

$$|e^{ix} - e^{iy}| = \sqrt{2} \sqrt{1 - \cos(x - y)} \leq |x - y| \, . \tag{2.127}$$

Hence, we have proved that C_δ is a $(\frac{\varepsilon}{\sqrt{2}} + \delta)$-covering net of $\mathcal{S}_{|*\rangle}$ with respect to the 2-norm $\| \cdot \|_2$. Consequently, by setting $\delta \coloneqq \varepsilon - \frac{\varepsilon}{\sqrt{2}}$, we have that $C_{\varepsilon - \frac{\varepsilon}{\sqrt{2}}}$ is an ε-covering net of $\mathcal{S}_{|*\rangle}$ with respect to the 2-norm $\| \cdot \|_2$. In particular, its cardinality has to satisfy

$$|C_{\varepsilon - \frac{\varepsilon}{\sqrt{2}}}| \geq \mathcal{C}(\mathcal{S}_{|*\rangle}, \| \cdot \|_2, \varepsilon) \, . \tag{2.128}$$

Consequently, we can conclude that

$$\begin{aligned}
\mathcal{C}(\mathcal{S}_{|*\rangle\langle *|}, \| \cdot \|_1, \varepsilon) &= |C^{\varepsilon,\mathrm{opt}}_{\mathcal{S}_{|*\rangle\langle *|}}| \\
&\geq \left(\left\lfloor \frac{2\pi}{\varepsilon - \frac{\varepsilon}{\sqrt{2}}} \right\rfloor + 1 \right)^{-1} |C_{\varepsilon - \frac{\varepsilon}{\sqrt{2}}}| \\
&\geq \left(\left\lfloor \frac{2\pi}{\varepsilon - \frac{\varepsilon}{\sqrt{2}}} \right\rfloor + 1 \right)^{-1} \mathcal{C}(\mathcal{S}_{|*\rangle}, \| \cdot \|_2, \varepsilon) \\
&\geq \frac{\varepsilon}{8\pi} \mathcal{C}(\mathcal{S}_{|*\rangle}, \| \cdot \|_2, \varepsilon),
\end{aligned} \tag{2.129}$$

where the first inequality follows by a simple counting argument, based on Eq. (2.123), which shows that $|C_\delta| \leq |C^{\varepsilon,\mathrm{opt}}_{\mathcal{S}_{|*\rangle\langle *|}}| (\lfloor \frac{2\pi}{\delta} \rfloor + 1)$, and the last inequality follows by inspection. □

The following lemma establishes a lower bound on the ε-packing number of the set of pure state density matrices with respect to the trace norm.

Lemma 30 (ε-packings)**.** *Let $d \in \mathbb{N}$ and $\varepsilon \in [0,1)$. Define the set of pure state density matrices $\mathcal{S}_{|*\rangle\langle *|}$ as*

$$\mathcal{S}_{|*\rangle\langle *|} \coloneqq \{|v\rangle\langle v| \,:\, |v\rangle \in \mathbb{C}^d \text{ such that } |\langle v|v\rangle| = 1\}. \tag{2.130}$$

Then, we have

$$\mathcal{P}(\mathcal{S}_{|*\rangle\langle *|}, \|\cdot\|_1, \varepsilon) \geq \frac{1}{8\pi}\varepsilon^{-2(d-1)}\,. \tag{2.131}$$

Proof. Consider the set of pure state vectors $\mathcal{S}_{|*\rangle\langle *|}$ defined as

$$\mathcal{S}_{|*\rangle} \coloneqq \{|v\rangle \,:\, |v\rangle \in \mathbb{C}^d \text{ such that } |\langle v|v\rangle| = 1\}. \tag{2.132}$$

Let $B_1(k)$ be the k-dimensional unit ball and $\partial B_1(k)$ its boundary, with $k \in \mathbb{N}$ and $r \in \mathbb{R}_+$, defined as

$$B_r(k) \coloneqq \{x \in \mathbb{R}^k \,:\, \|x\|_2 \leq r\}, \tag{2.133}$$

$$\partial B_r(k) \coloneqq \{x \in \mathbb{R}^k \,:\, \|x\|_2 = r\}. \tag{2.134}$$

We can establish the following chain of inequalities:

$$\begin{aligned}
\mathcal{P}(\mathcal{S}_{|*\rangle\langle *|}, \|\cdot\|_1, \varepsilon) &\overset{\text{(i)}}{\geq} \mathcal{C}(\mathcal{S}_{|*\rangle\langle *|}, \|\cdot\|_1, \varepsilon) \\
&\overset{\text{(ii)}}{\geq} \frac{\varepsilon}{8\pi}\mathcal{C}(\mathcal{S}_{|*\rangle}, \|\cdot\|_2, \varepsilon) \\
&\overset{\text{(iii)}}{=} \frac{\varepsilon}{8\pi}\mathcal{C}(\partial B_1(2d), \|\cdot\|_2, \varepsilon) \\
&\overset{\text{(iv)}}{\geq} \frac{\varepsilon}{8\pi}\mathcal{C}(B_1(2d-1), \|\cdot\|_2, \varepsilon) \\
&\overset{\text{(v)}}{\geq} \frac{\varepsilon}{8\pi}\frac{\mathrm{Vol}(B_1(2d-1))}{\mathrm{Vol}(B_\varepsilon(2d-1))} \\
&\overset{\text{(vi)}}{=} \frac{\varepsilon^{-2(d-1)}}{8\pi}\,.
\end{aligned} \tag{2.135}$$

Here, (i) and (ii) follow from Lemma 27 and Lemma 29, respectively. Moreover, (iii) stems from viewing the set of pure states $\mathcal{S}_{|*\rangle}$ as the boundary of the $2d$-dimensional real ball $B_1(2d)$. In (iv), we apply Lemma 28. In (v), we denote $\mathrm{Vol}(B_1(2d-1))$ as the volume of the $(2d-1)$-dimensional ball with unit radius and $\mathrm{Vol}(B_\varepsilon(2d-1))$ as the volume of the $(2d-1)$-dimensional ball with radius ε. We then exploit a simple counting argument to get $\mathrm{Vol}(B_\varepsilon(2d-1))\mathcal{C}(B_1(2d-1), \|\cdot\|_2, \varepsilon) \geq \mathrm{Vol}(B_1(2d-1))$. Finally, in (vi), we use the known fact that the volume of an n-dimensional ball with radius r is $r^n\mathrm{Vol}(B_1(2d-1))$. □

We now present a lemma which relates the covering number of the set of quantum states on $\mathbb{C}^d$ with the covering number of the set of Hermitian matrices with trace norm smaller than one.

Lemma 31 (Covering number of the set of Hermitian matrices)**.** *Let $\varepsilon > 0$ and $d \in \mathbb{N}_+$. Let $\mathcal{D}\left(\mathbb{C}^d\right)$ and $\mathrm{Herm}_{\|*\|_1\leq 1}(d)$ be respectively the set of quantum states on $\mathbb{C}^d$ and the set of Hermitian matrices with trace norm smaller than one, i.e.*

$$\begin{aligned}
\mathcal{D}\left(\mathbb{C}^d\right) &\coloneqq \left\{\rho \in \mathbb{C}^{d,d} \,:\, \rho \geq 0,\, \mathrm{Tr}\,\rho = 1\right\}, \\
\mathrm{Herm}_{\|*\|_1\leq 1}(d) &\coloneqq \left\{X \in \mathbb{C}^{d,d} \,:\, X^\dagger = X,\, \|X\|_1 \leq 1\right\}.
\end{aligned} \tag{2.136}$$

Then, it holds the following relation between their covering numbers

$$\mathcal{C}(\mathcal{D}\left(\mathbb{C}^d\right), \|\cdot\|_1, \varepsilon) \geq \sqrt{\mathcal{C}(\mathrm{Herm}_{\|*\|_1\leq 1}(d-1), \|\cdot\|_1, 2\varepsilon)}\,. \tag{2.137}$$

Proof. The strategy of the proof is to construct a covering net for $\mathrm{Herm}_{\|*\|_1\leq 1}(d-1)$ from an optimal ε-covering net of set of quantum states $\mathcal{D}\big(\mathbb{C}^d\big)$. Let $C^{\varepsilon,\mathrm{opt}}_{\mathcal{D}\left(\mathbb{C}^d\right)}$ be an optimal ε-covering net for $\mathcal{D}\big(\mathbb{C}^d\big)$. Let us fix $X \in \mathrm{Herm}_{\|*\|_1\leq 1}(d-1)$, and let us write it in its Jordan–Hahn decomposition [74] as follows, $X = X_1 - X_2$, where $X_1, X_2 \geq 0$ and $X_1 X_2 = X_2 X_1 = 0$. Since

$$1 \geq \|X\|_1 = \operatorname{Tr} X_1 + \operatorname{Tr} X_2\,, \tag{2.138}$$

it follows that

$$\begin{aligned} 1 - \operatorname{Tr} X_1 &\geq 0\,,\\ 1 - \operatorname{Tr} X_2 &\geq 0\,. \end{aligned} \tag{2.139}$$

Let us consider the $d \times d$ block diagonal matrices

$$\begin{aligned} \bar{X}_1 &\coloneqq \begin{pmatrix} X_1 & \mathbf{0} \\ \mathbf{0}^\intercal & 1 - \operatorname{Tr} X_1 \end{pmatrix}\,,\\ \bar{X}_2 &\coloneqq \begin{pmatrix} X_2 & \mathbf{0} \\ \mathbf{0}^\intercal & 1 - \operatorname{Tr} X_2 \end{pmatrix}\,, \end{aligned} \tag{2.140}$$

where $\mathbf{0}$ is the $(d-1)$-dimensional zero vector. Hence, both $\bar{X}_1$ and $\bar{X}_2$ are density operators, i.e., $\bar{X}_1, \bar{X}_2 \in \mathcal{D}\big(\mathbb{C}^d\big)$. By definition of $C^{\varepsilon,\mathrm{opt}}_{\mathcal{D}\left(\mathbb{C}^d\right)}$, there exist $\rho_1, \rho_2 \in C^{\varepsilon,\mathrm{opt}}_{\mathcal{D}\left(\mathbb{C}^d\right)}$ such that $\|\bar{X}_1 - \rho_1\|_1 \leq \varepsilon$ and $\|\bar{X}_2 - \rho_2\|_1 \leq \varepsilon$. Given a matrix A, let us denote as $A|_n$ its $n \times n$ top-left block. Specifically, we denote as $\rho_1|_{d-1}$ and as $\rho_2|_{d-1}$ the $(d-1)\times(d-1)$ top-left blocks of ρ_1 and ρ_2, respectively. We have that

$$\begin{aligned} \|X - (\rho_1|_{d-1} - \rho_2|_{d-1})\|_1 &= \|X_1 - X_2 - (\rho_1|_{d-1} - \rho_2|_{d-1})\|_1\\ &\leq \|X_1 - \rho_1|_{d-1}\|_1 + \|X_2 - \rho_2|_{d-1}\|_1\\ &\overset{\text{(i)}}{\leq} \|\bar{X}_1 - \rho_1\|_1 + \|\bar{X}_2 - \rho_2\|_1\\ &\leq 2\varepsilon\,. \end{aligned} \tag{2.141}$$

Here, in (i), we exploited the general fact that the trace norm of any matrix A is an upper bound on the trace norm of its $n \times n$ top-left block $A|_n$, i.e., $\|A|_n\|_1 \leq \|A\|_1$. This latter fact can be easily proved by exploiting the *minimax principle for singular values* [75, Problem III.6.1]. Therefore, we have proved that the set

$$S \coloneqq \{\rho_1|_{d-1} - \rho_2|_{d-1} \,:\, \rho_1, \rho_2 \in C^{\varepsilon,\mathrm{opt}}_{\mathcal{D}\left(\mathbb{C}^d\right)}\} \tag{2.142}$$

constitutes a 2ε-covering net of $\mathrm{Herm}_{\|*\|_1\leq 1}(d-1)$, and thus $|S| \geq \mathcal{C}(\mathrm{Herm}_{\|*\|_1\leq 1}(d-1), \|\cdot\|_1, 2\varepsilon)$. Therefore, we conclude that

$$\mathcal{C}(\mathrm{Herm}_{\|*\|_1\leq 1}(d-1), \|\cdot\|_1, 2\varepsilon) \leq |S| \leq |C^{\varepsilon,\mathrm{opt}}_{\mathcal{D}\left(\mathbb{C}^d\right)}|^2 = \mathcal{C}(\mathcal{D}\big(\mathbb{C}^d\big), \|\cdot\|_1, \varepsilon)^2\,. \tag{2.143}$$

□

We now present a lemma which gives a lower bound on the covering number of the set of Hermitian matrices with trace norm smaller than one.

Lemma 32 (Lower bound on the covering number)**.** *Let $\varepsilon > 0$ and $d \in \mathbb{N}_+$. Let $\mathrm{Herm}_{\|*\|_1\leq 1}(d)$ be the set of Hermitian matrices with trace norm smaller than one, i.e.*

$$\mathrm{Herm}_{\|*\|_1\leq 1}(d) \coloneqq \left\{X \in \mathbb{C}^{d,d} \,:\, X^\dagger = X,\, \|X\|_1 \leq 1\right\}\,. \tag{2.144}$$

Then, its covering number can be lower bounded by

$$\mathcal{C}(\mathrm{Herm}_{\|*\|_1\leq 1}(d), \|\cdot\|_1, \varepsilon) \geq \varepsilon^{-d^2}\,. \tag{2.145}$$

Proof. Let us define the ball of $d \times d$ Hermitian matrices with radius r with respect the trace norm as

$$B_r(d) \coloneqq \left\{ X \in \mathbb{C}^{d,d} \,:\, X^\dagger = X,\, \|X\|_1 \leq r \right\} , \tag{2.146}$$

Note that the ball of radius one $B_1(d)$ is exactly equal to the set $\mathrm{Herm}_{\|*\|_1 \leq 1}(d)$. The volume of the ball with radius r can be expressed in terms of the volume of the ball with unit radius as

$$\mathrm{Vol}(B_r(d)) = r^{d^2} \mathrm{Vol}(B_1(d)) \ . \tag{2.147}$$

Let us prove this latter fact. First, note that there are d^2 independent real parameters $y_1, y_2, \ldots, y_{d^2}$ that define any $d \times d$ Hermitian matrix. Specifically, there are $d^2 - d$ independent real parameters for the off diagonal elements, and d independent real parameters for the diagonal ones. Given the d^2-dimensional vector $y = (y_1, y_2, \ldots, y_{d^2}) \in \mathbb{R}^{d^2}$ of such parameters, let $X(y)$ be the associated Hermitian matrix. Consequently, we have that

$$\mathrm{Vol}(B_r(d)) = \int\limits_{\substack{y \in \mathbb{R}^{d^2} \\ \|X(y)\|_1 \leq r}} 1 \, \mathrm{d}^{d^2} y \overset{\text{(i)}}{=} r^{d^2} \int\limits_{\substack{y \in \mathbb{R}^{d^2} \\ \|X(y)\|_1 \leq 1}} 1 \, \mathrm{d}^{d^2} y = \ r^{d^2} \mathrm{Vol}(B_1(d)) \ , \tag{2.148}$$

where in (i) we made the change of variable $y \mapsto ry$ and we exploited that $X(ry) = rX(y)$. Hence, we have proved (2.147). Finally, a simple counting argument shows that

$$\mathcal{C}(\mathrm{Herm}_{\|*\|_1 \leq 1}(d), \|\cdot\|_1, \varepsilon) \geq \frac{\mathrm{Vol}(B_1(d))}{\mathrm{Vol}(B_\varepsilon(d))} , \tag{2.149}$$

which, together with (2.147), concludes the proof. □

The following corollary establishes a lower bound on the covering number of the set of (mixed) quantum states.

Corollary 33. *Let $\varepsilon > 0$ and $d \in \mathbb{N}_+$. Let $\mathcal{D}\left(\mathbb{C}^d\right)$ be the set of quantum states on $\mathbb{C}^d$, i.e.*

$$\mathcal{D}\left(\mathbb{C}^d\right) \coloneqq \left\{ \rho \in \mathbb{C}^{d,d} \,:\, \rho \geq 0,\, \mathrm{Tr}\, \rho = 1 \right\} . \tag{2.150}$$

Then, its packing number can be lower bounded as follows

$$\mathcal{P}(\mathcal{D}\left(\mathbb{C}^d\right), \|\cdot\|_1, \varepsilon) \geq (2\varepsilon)^{-\frac{1}{2}(d-1)^2} \tag{2.151}$$

Proof. The result follows from the following chain of inequalities:

$$\begin{aligned} \mathcal{P}(\mathcal{D}\left(\mathbb{C}^d\right), \|\cdot\|_1, \varepsilon) &\overset{\text{(i)}}{\geq} \mathcal{C}(\mathcal{D}\left(\mathbb{C}^d\right), \|\cdot\|_1, \varepsilon) \\ &\overset{\text{(ii)}}{\geq} \sqrt{\mathcal{C}(\mathrm{Herm}_{\|*\|_1 \leq 1}(d-1), \|\cdot\|_1, 2\varepsilon)} \\ &\overset{\text{(iii)}}{\geq} (2\varepsilon)^{-\frac{1}{2}(d-1)^2} . \end{aligned} \tag{2.152}$$

Here, (i), (ii), and (iii) follow from Lemma 27, Lemma 31, and Lemma 32, respectively. □

2.7.2 Tomography of moment-constrained states

In this section we consider the problem of quantum state tomography of continuous-variable systems. Since continuous-variable states are associated with an infinite dimensional Hilbert space, a first trivial observation is that quantum state tomography is impossible if one has no any extra prior information about the unknown state. However, in a practical scenario, experimentalists often possess knowledge about the energy budget available of their light sources — and hence about the mean energy of the unknown quantum state. Thus, it is crucial to analyse the problem of tomography of n-mode states subject to an energy constraint. Such energy constraint can be expressed as the following upper bound on the mean photon number of the unknown n-mode state ρ:

$$\mathrm{Tr}[\hat{N}_n\rho] \leq nN_{\mathrm{phot}}\,, \tag{2.153}$$

where $\hat{N}_n$ is the n-mode photon number operator, and N_{phot} is a real number. In other words, one may have the extra prior information that the unknown quantum state is contained in the set

$$\Big\{\rho \in \mathcal{D}\Big(L^2(\mathbb{R}^n)\Big) : \quad \mathrm{Tr}\Big[\hat{N}_n\rho\Big] \leq nN_{\mathrm{phot}}\Big\} \tag{2.154}$$

of energy constrained states. Here, the number of modes n and the energy constraint N_{phot} are *known* parameters, in the sense that the tomography algorithm may explicitly depend on their values. Note that we normalise the right-hand-side of (2.153) with the factor n because the mean photon number $\mathrm{Tr}[\hat{N}_n\rho]$ is extensive, meaning that $\mathrm{Tr}[\hat{N}_n\sigma^{\otimes n}] = n\,\mathrm{Tr}[\hat{N}_1\sigma]$ for all single-mode states σ. In this section, we first consider the following question:

> Suppose there exists a tomography algorithm designed to learn, with ε-precision in trace distance and high probability, an unknown n-mode pure state ψ satisfying the energy constraint $\mathrm{Tr}[\hat{N}_n\psi] \leq nN_{\mathrm{phot}}$. What is the required sample complexity, represented by the number N of state copies?

In this section, we answer this question by showing that *any such tomography algorithm needs at least* $N = \Omega\Big(\frac{N_{phot}}{\varepsilon^2}\Big)^n$ *state copies*. This result imposes strong limitations on quantum state tomography of continuous-variable systems, due to the unfavourable scaling not only in the number of modes n but also in the trace-distance error ε. In contrast, in the finite-dimensional setting, quantum state tomography of an unknown n-qubit state needs at least $N = \tilde{T}heta\,(2^n/\varepsilon^2)$ copies of the state (indicating a more favorable dependence on ε). Conversely, we also show that there exists a tomography algorithm that achieves the same sample complexity performances. Thus, we have the following main result:

Theorem 34 (Tomography of energy-constrained pure states - informal version)**.** $N = \Theta\Big(\frac{N_{phot}}{\varepsilon^2}\Big)^n$ *state copies are necessary and sufficient for quantum state tomography of energy-constrained* n*-mode pure states with trace-distance error* ε*.*

In the general case of n-mode energy-constained (possibly mixed) states we show the following:

Theorem 35 (Tomography of energy-constrained mixed states - informal version)**.** $N = \Omega\Big(\frac{N_{phot}}{\varepsilon}\Big)^{2n}$ *state copies are necessary for quantum state tomography of energy-constrained* n*-mode (possibly mixed) states with trace-distance error* ε*. Conversely,* $N = O\Big(\frac{N_{phot}}{\varepsilon^{3/2}}\Big)^{2n}$ *state copies are sufficient for this task.*

In practical scenarios, experimentalists often have knowledge not only about the mean energy but also about higher moments (e.g., energy variance) of their light sources. Therefore, it is meaningful to analyse tomography of states satisfying constraints on higher moments of energy. Mathematically, the k-th moment constraint can be expressed as

$$\Big(\mathrm{Tr}[\hat{N}_n^k\rho]\Big)^{1/k} \leq nN_{\mathrm{phot}}\,. \tag{2.155}$$

Note that we normalise the right-hand-side of (2.155) with the factor n because the k-th moment $\left(\mathrm{Tr}[\hat{N}_n^k \rho]\right)^{1/k}$ is extensive when evaluated on Fock states $\rho = |m\rangle\langle m|^{\otimes n}$, meaning that

$$\left(\mathrm{Tr}[\hat{N}_n^k \, |m\rangle\langle m|^{\otimes n}]\right)^{1/k} = n \left(\mathrm{Tr}[\hat{N}_1^k \, |m\rangle\langle m|]\right)^{1/k}. \tag{2.156}$$

In summary, it is sometimes meaningful to possess the extra prior information that the unknown quantum state is contained in the following set of k-th moment-constrained states:

$$\mathcal{S}(n, N_{\mathrm{phot}}, k) \coloneqq \left\{ \rho \in \mathcal{D}\left(L^2(\mathbb{R}^n)\right) : \quad \mathrm{Tr}\left(\left[\hat{N}_n^k \rho\right]\right)^{1/k} \leq n N_{\mathrm{phot}} \right\}, \tag{2.157}$$

where we stress that n, N_{phot}, and k are *known* parameters. Note that the higher the value of k, the more prior extra information we have about the unknown quantum state. Indeed, this is established by the following simple result.

Lemma 36 (High moment-constraint implies low moment-constraint)**.** *For any $k, q \in \mathbb{N}_+$ with $k \geq q$, it holds that*

$$\mathcal{S}(n, N_{phot}, k) \subseteq \mathcal{S}(n, N_{phot}, q). \tag{2.158}$$

Proof. By Eq. 2.41 we have that, for any concave function f, any Hermitian matrix X, and any quantum state ρ, it holds that

$$\mathrm{Tr}(\rho f(X)) \leq f(\mathrm{Tr}(\rho X)). \tag{2.159}$$

Now, the claim follows from the fact that

$$\left(\mathrm{Tr}\left[\rho \, \hat{N}_n^q\right]\right)^{1/q} = \left(\mathrm{Tr}\left[\rho \left(\hat{N}_n^k\right)^{q/k}\right]\right)^{1/q} \leq \left(\mathrm{Tr}\left[\rho \, \hat{N}_n^k\right]\right)^{1/k}, \tag{2.160}$$

where in the first step we use the semi-definite positivity of $\hat{N}_n$, and the inequality follows from the concavity of the function $x \mapsto x^{q/k}$ for $x \geq 0$. □

In this section, we also analyse tomography of moment-constrained states and establish the following results.

Theorem 37 (Tomography of moment-constrained states - informal version)**.** *Let ρ be an unknown n-mode state satisfying the k-th moment constraint $(\mathrm{Tr}[\hat{N}_n^k \rho])^{1/k} \leq n N_{phot}$. The following facts hold:*

a. The number of copies of ρ required to achieve quantum state tomography with precision ε in trace distance scales at least as $\Omega\left(\frac{N_{phot}}{\varepsilon^{1/k}}\right)^{2n}$.

b. There exists a tomography algorithm with sample complexity scaling as $O\left(\frac{N_{phot}}{\varepsilon^{3/(2k)}}\right)^{2n}$.

c. If we assume the unknown state ρ to be pure, then $\Theta\left(\frac{N_{phot}}{\varepsilon^{2/k}}\right)^{n}$ copies of ρ are necessary and sufficient for tomography.

We note that the upper bound $O\left(\frac{N_{\mathrm{phot}}}{\varepsilon^{3/(2k)}}\right)^{2n}$ on the sample complexity of tomography of moment-constrained (mixed) states provided in Theorem 37 outperforms the one attainable through the CV classical shadow algorithm outlined in References [16, 17].

This section is organised as follows:

- In Subsection 2.7.2, we derive lower bounds on the sample complexity of tomography of moment-constrained states.

- In Subsection 2.7.2, we show upper bounds on the sample complexity by providing explicit tomography algorithms to learn moment-constrained states. Throughout this section, we also establish key properties of moment-constrained states, demonstrating that they can be effectively approximated by states with finite local dimension and rank.

Note: Throughout this section, we opted to simplify the notation in the use of parentheses in asymptotic expressions. For example, when we write $\Theta\left(\frac{N_{\text{phot}}}{\varepsilon^{2/k}}\right)^n$ we formally mean $\left[\Theta\left(\frac{N_{\text{phot}}}{\varepsilon^{2/k}}\right)\right]^n$. We stress that this notation omits only factors that are independent of the parameters of the problem, i.e., n, ε, and N_{phot}. For instance, with this notation, it is true that $\Theta\left(\frac{2N_{\text{phot}}}{\varepsilon^{2/k}}\right)^n = \Theta\left(\frac{N_{\text{phot}}}{\varepsilon^{2/k}}\right)^n$, but $\Theta\left(\frac{N_{\text{phot}}^2}{\varepsilon^{2/k}}\right)^n \neq \Theta\left(\frac{N_{\text{phot}}}{\varepsilon^{2/k}}\right)^n$. Importantly, to ensure clarity and avoid ambiguity, we express every sample complexity bound using both asymptotic notation and explicit exact expressions.

Sample complexity lower bounds

To establish lower bounds on the sample complexity of tomography in continuous variable systems, we rely on the concept of ε-nets (as discussed in Section 2.7.1). In the subsequent subsubsection 2.7.2, we delve into the analysis of sample complexity lower bounds for pure state tomography. Following that, we proceed to demonstrate lower bounds for mixed state tomography in subsubsection 2.7.2.

Lower bound for pure state tomography We now present and prove our main theorem concerning the number of copies necessary for learning moment-constrained pure states. The proof strategy hinges on the utilisation of Lemma 26, which relies on Fano's inequality and Holevo's bound, and on identifying a suitable 'packing' for the set of moment-constrained pure states.

Theorem 38 (Lower bound on sample complexity for moment-constrained pure states)**.** *Let us consider a tomography algorithm that learns, within a trace distance $\leq \varepsilon$ and failure probability $\leq \delta$, an arbitrary state belonging to the set*

$$\mathcal{S}_{pure}(n,k,N_{phot}) \coloneqq \left\{ |\psi\rangle \,:\, n\text{-mode pure state}, \left(\operatorname{Tr}\left[\left(\hat{N}_n\right)^k |\psi\rangle\langle\psi|\right]\right)^{1/k} \leq nN_{phot} \right\} \tag{2.161}$$

of n-mode pure states with bounded k-moment. Then, such a tomography algorithm must use a number of state copies N satisfying

$$\begin{aligned} N &\geq \frac{1}{ng(N_{phot})}\left[2(1-\delta)\left(\frac{N_{phot}}{(12\varepsilon)^{2/k}} - \frac{1}{n}\right)^n - (1-\delta)\log_2(32\pi) - H_2(\delta)\right] \\ &= \Theta\left(\frac{N_{phot}}{\varepsilon^{2/k}}\right)^n . \end{aligned} \tag{2.162}$$

Here, $g(x) \coloneqq (x+1)\log_2(x+1) - x\log_2 x$ is the bosonic entropy, and $H_2(x) \coloneqq -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy.

Proof. Let $M \in \mathbb{N}$ be such that there exist M states

$$\{|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_M\rangle\} \subseteq \mathcal{S}_{\text{pure}}(n,k,N_{\text{phot}}) \tag{2.163}$$

such that for every $i \neq j \in [M]$ it holds that

$$\frac{1}{2}\|\psi_i - \psi_j\|_1 > 2\varepsilon\,, \tag{2.164}$$

where $\psi_i \coloneqq |\psi_i\rangle\langle\psi_i|$ and $\psi_j \coloneqq |\psi_j\rangle\langle\psi_j|$ (as depicted in Fig. 2.7.2). Thanks to Lemma 26, any

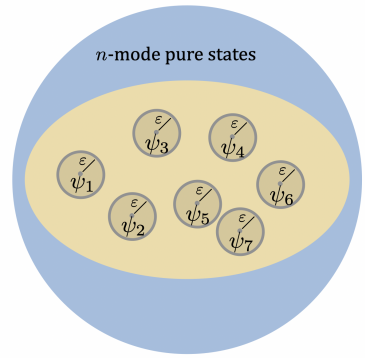


Figure 2.6: Pictorial representation of the set of all n-mode pure states (blue) and the set of of all n-mode pure states with bounded k-moment (yellow). We consider M moment-constrained pure states $\{\psi_1, \psi_2, \cdots, \psi_M\}$ such that they are at least 2ε-far from each other with respect to the trace distance.

tomography algorithm that learns states from the set $\mathcal{S}_{\text{pure}}(n, k, N_{\text{phot}})$ with accuracy $\leq \varepsilon$ in trace distance and failure probability $\leq \delta$ must use a number of state copies N satisfying:

$$\chi \geq (1-\delta)\log_2(M) - H_2(\delta)\,, \tag{2.165}$$

where $\chi := S\left(\frac{1}{M}\sum_{j=1}^{M}\psi_j^{\otimes N}\right) - \frac{1}{M}\sum_{j=1}^{M} S(\psi_j^{\otimes N})$ is the Holevo information, and $S(\cdot)$ is the von Neumann entropy. Note that the latter quantity can be upper bounded as

$$\chi = S\left(\frac{1}{M}\sum_{j=1}^{M}\psi_j^{\otimes N}\right) - \frac{1}{M}\sum_{j=1}^{M} S(\psi_j^{\otimes N}) \leq S\left(\frac{1}{M}\sum_{j=1}^{M}\psi_j^{\otimes N}\right) \leq nNg(N_{\text{phot}})\,, \tag{2.166}$$

where in the last inequality we exploited the fact that the von Neumann entropy of a state with a fixed mean photon number can be upper bounded in terms of the bosonic entropy function $g(\cdot)$ (see Lemma 21). Specifically, note that the mean total photon number of the nN-mode state $\frac{1}{M}\sum_{j=1}^{M}\psi_j^{\otimes N}$ can be upper bounded as

$$\text{Tr}\left[\hat{N}_{nN}\left(\frac{1}{M}\sum_{j=1}^{M}\psi_j^{\otimes N}\right)\right] = N\frac{1}{M}\sum_{j=1}^{M}\text{Tr}[\hat{N}_n\psi_j] \leq N\frac{1}{M}\sum_{j=1}^{M}\left(\text{Tr}[\hat{N}_n^k\psi_j]\right)^{1/k} \leq NnN_{\text{phot}}\,, \tag{2.167}$$

where in the first inequality we have used that $\text{Tr}[\hat{N}_n\psi_j] = \text{Tr}[(\hat{N}_n^k)^{\frac{1}{k}}\psi_j]$ and exploited the concavity of $x \mapsto x^{1/k}$ for $x > 0, k \in \mathbb{N}$ (see Eq. (2.41)), while in the last inequality we exploited that $\psi_j \in \mathcal{S}_{\text{pure}}(n, k, N_{\text{phot}})$. Hence, Lemma 21 establishes that

$$S\left(\frac{1}{M}\sum_{j=1}^{M}\psi_j^{\otimes N}\right) \leq nNg(N_{\text{phot}})\,. \tag{2.168}$$

By putting together (2.165) and (2.166), we deduce that, if such M exists, then the number of state copies N required for tomography of an unknown state belonging to $\mathcal{S}_{\text{pure}}(n, k, N_{\text{phot}})$ has to be at least

$$N \geq \frac{(1-\delta)\log_2(M) - H_2(\delta)}{ng(N_{\text{phot}})}\,. \tag{2.169}$$

Hence, in order to get a good lower bound on N, we need to find a large value of M such that there exist M states

$$\{\psi_1, \psi_2, \ldots, \psi_M\} \subseteq \mathcal{S}_{\text{pure}}(n, k, N_{\text{phot}}) \tag{2.170}$$

for which it holds that $\frac{1}{2}\|\psi_i - \psi_j\|_1 > 2\varepsilon$ for all $i \neq j \in [M]$. To this end, let us make a suitable construction of such M states.

Let $m \in \mathbb{N}$ such that $m \geq nN_{\text{phot}}$ which will be fixed later. Let us define the finite-dimensional Hilbert space $\mathcal{H}_m$ spanned by all the n-mode Fock states, apart from the n-mode vacuum state, with total number of photons less than m:

$$\mathcal{H}_m \coloneqq \left\{ |\phi\rangle \in \operatorname{span}\left(|\mathbf{k}\rangle : \ \mathbf{k} \in \mathbb{N}^n \setminus \{0\} \ , \sum_{i=1}^n k_i \leq m \right) : \ \langle \phi | \phi \rangle = 1 \right\} , \tag{2.171}$$

where $|\mathbf{k}\rangle$ is the n-mode Fock state defined by $|\mathbf{k}\rangle = |k_1\rangle \otimes |k_2\rangle \otimes \ldots \otimes |k_n\rangle$.

Given a state vector $|\phi\rangle \in \mathcal{H}_m$, we can define a corresponding state vector $|\psi_\phi\rangle \in \mathcal{S}_{\text{pure}}(n, k, N_{\text{phot}})$ as

$$|\psi_\phi\rangle \coloneqq \sqrt{1 - \left(\frac{nN_{\text{phot}}}{m}\right)^k} |0\rangle^{\otimes n} + \sqrt{\left(\frac{nN_{\text{phot}}}{m}\right)^k} |\phi\rangle \ . \tag{2.172}$$

Indeed, the k-moment of $|\psi_\phi\rangle$ can be upper bounded as

$$\operatorname{Tr}[\hat{N}_n^k |\psi_\phi\rangle\langle\psi_\phi|] = \frac{(nN_{\text{phot}})^k}{m^k} \operatorname{Tr}[\hat{N}_n^k |\phi\rangle\langle\phi|] \leq \frac{(nN_{\text{phot}})^k}{m^k} m^k = (nN_{\text{phot}})^k \ , \tag{2.173}$$

and thus $|\psi_\phi\rangle \in \mathcal{S}_{\text{pure}}(n, k, N_{\text{phot}})$. Moreover, as we show below in (2.182), the following inequality holds

$$\frac{1}{2} \left\| |\psi_\phi\rangle\langle\psi_\phi| - |\psi_{\tilde{\phi}}\rangle\langle\psi_{\tilde{\phi}}| \right\|_1 \geq \sqrt{\frac{(nN_{\text{phot}})^k}{8m^k}} \left\| |\phi\rangle\langle\phi| - |\tilde{\phi}\rangle\langle\tilde{\phi}| \right\|_1 \tag{2.174}$$

for all $|\phi\rangle, |\tilde{\phi}\rangle \in \mathcal{H}_m$. Hence, by choosing

$$m \coloneqq \left\lfloor \frac{nN_{\text{phot}}}{(12\varepsilon)^{2/k}} \right\rfloor \ , \tag{2.175}$$

one can check that for all $|\phi\rangle, |\tilde{\phi}\rangle \in \mathcal{H}_m$ such that

$$\left\| |\phi\rangle\langle\phi| - |\tilde{\phi}\rangle\langle\tilde{\phi}| \right\|_1 \geq \frac{1}{2} , \tag{2.176}$$

the inequality in (2.174) guarantees that

$$\frac{1}{2} \left\| |\psi_\phi\rangle\langle\psi_\phi| - |\psi_{\tilde{\phi}}\rangle\langle\psi_{\tilde{\phi}}| \right\|_1 \geq 2\varepsilon \, . \tag{2.177}$$

Therefore, let us find $M \in \mathbb{N}$ such that there exist M states $\{|\phi_1\rangle, |\phi_2\rangle, \ldots, |\phi_M\rangle\} \subseteq \mathcal{H}_m$ such that $\| |\phi_i\rangle\langle\phi_i| - |\phi_j\rangle\langle\phi_j| \|_1 \geq \frac{1}{2}$ for all $i \neq j \in [M]$. By definition of packing number (see Definition 2), such M can be chosen to be the $\frac{1}{2}$-packing number of $\mathcal{H}_m$ with respect to the trace norm, i.e., $M \coloneqq \mathcal{P}\left(\mathcal{H}_m, \|\cdot\|_1, \frac{1}{2}\right)$. We can find a simple lower bound on M as

$$M = \mathcal{P}\left(\mathcal{H}_m, \|\cdot\|_1, \frac{1}{2}\right) \overset{\text{(i)}}{\geq} \frac{1}{8\pi} 2^{2(\dim \mathcal{H}_m - 1)} \overset{\text{(ii)}}{\geq} \frac{1}{8\pi} 2^{2[\binom{n+m}{n} - 2]} \overset{\text{(iii)}}{\geq} \frac{1}{8\pi} 2^{2[(\frac{m}{n})^n - 1]} = \frac{1}{32\pi} 2^{2(\frac{m}{n})^n} \, . \tag{2.178}$$

Here, in (i), we exploit Lemma 30; in (ii), we observed that the dimension of $\mathcal{H}_m$ is given by

$$\dim \mathcal{H}_m = \left| \left\{ \mathbf{k} \in \mathbb{N}^n \setminus \{0\} \, , \quad \sum_{i=1}^n k_i \leq m \right\} \right| = \binom{m+n}{n} - 1 \, , \tag{2.179}$$

where the last equality follows from a standard combinatorial argument; in (iii), we just exploited that $\binom{m+n}{n} \geq (\frac{m+n}{n})^n \geq (\frac{m}{n})^n + 1$. Consequently, (2.169) implies that

$$N \geq \frac{(1-\delta) 2(\frac{m}{n})^n - (1-\delta)\log_2(32\pi) - H_2(\delta)}{n g(N_{\text{phot}})} \, . \tag{2.180}$$

Finally, by exploiting the definition of m in (2.175) we conclude that

$$N \geq \frac{2(1-\delta)\left(\frac{N_{\text{phot}}}{(12\varepsilon)^{2/k}} - \frac{1}{n}\right)^n - (1-\delta)\log_2(32\pi) - H_2(\delta)}{ng(N_{\text{phot}})} \,. \tag{2.181}$$

We are just left to show (2.174):

$$\begin{aligned}
\| \, |\psi_\phi\rangle\langle\psi_\phi| - |\psi_{\tilde{\phi}}\rangle\langle\psi_{\tilde{\phi}}| \, \|_1 &= 2\sqrt{1 - |\, \langle\psi_\phi|\psi_{\tilde{\phi}}\rangle \,|^2} \\
&= 2\sqrt{1 - \left|\left(1 - \frac{\left(nN_{\text{phot}}\right)^k}{m^k}\right) + \frac{\left(nN_{\text{phot}}\right)^k}{m^k}\langle\phi|\tilde{\phi}\rangle\right|^2} \\
&\geq 2\sqrt{1 - \left(1 - \frac{\left(nN_{\text{phot}}\right)^k}{m^k} + \frac{\left(nN_{\text{phot}}\right)^k}{m^k}\left|\langle\phi|\tilde{\phi}\rangle\right|\right)^2} \\
&= 2\sqrt{2\frac{\left(nN_{\text{phot}}\right)^k}{m^k}\left(1 - \left|\langle\phi|\tilde{\phi}\rangle\right|\right) - \left(\frac{\left(nN_{\text{phot}}\right)^k}{m^k}\left(1 - \left|\langle\phi|\tilde{\phi}\rangle\right|\right)\right)^2} \\
&\geq 2\sqrt{\frac{\left(nN_{\text{phot}}\right)^k}{m^k}\left(1 - \left|\langle\phi|\tilde{\phi}\rangle\right|\right)} \\
&\geq \sqrt{2}\sqrt{\frac{\left(nN_{\text{phot}}\right)^k}{m^k}}\sqrt{\left(1 - \left|\langle\phi|\tilde{\phi}\rangle\right|\right)\left(1 + \left|\langle\phi|\tilde{\phi}\rangle\right|\right)} \\
&= \sqrt{2}\sqrt{\frac{\left(nN_{\text{phot}}\right)^k}{m^k}}\sqrt{\left(1 - \left|\langle\phi|\tilde{\phi}\rangle\right|^2\right)} \\
&= \sqrt{\frac{\left(nN_{\text{phot}}\right)^k}{2m^k}}\left\| |\phi\rangle\langle\phi| - |\tilde{\phi}\rangle\langle\tilde{\phi}| \right\|_1 ,
\end{aligned} \tag{2.182}$$

which concludes the proof. □

In particular, in the single-mode case, i.e., $n = 1$, the number of state copies N has to satisfy

$$\begin{aligned}
N &\geq \frac{1}{g(N_{\text{phot}})}\left[2(1-\delta)\left(\frac{N_{\text{phot}}}{(12\varepsilon)^{2/k}} - 1\right) - (1-\delta)\log_2(32\pi) - H_2(\delta)\right] \\
&= \Theta\left(\frac{N_{\text{phot}}}{\varepsilon^{2/k}}\right) .
\end{aligned} \tag{2.183}$$

Lower bound for mixed state tomography In this subsection, we establish a theorem concerning the number of copies necessary for learning moment-constrained mixed states. The proof strategy follows a similar logic to the one used previously for the pure state case.

Theorem 39 (Lower bound on sample complexity for moment-constrained mixed states)**.** *Let us consider a tomography algorithm that learns, within a trace distance $\leq \varepsilon$ and failure probability $\leq \delta$, an arbitrary state belonging to the set*

$$\mathcal{S}(n, k, N_{phot}) := \left\{\rho \in \mathcal{D}(L^2(\mathbb{R}^n)), \quad \left(\text{Tr}\left[\left(\hat{N}_n\right)^k \rho\right]\right)^{1/k} \leq nN_{phot}\right\} \tag{2.184}$$

of n-mode (possibly mixed) states with bounded k-moment. Then, such a tomography algorithm must use a number of state copies N satisfying

$$\begin{aligned}
N &\geq \frac{1}{2ng(N_{phot})}\left[(1-\delta)\left(\frac{N_{phot}}{(16\varepsilon)^{1/k}} - \frac{1}{n}\right)^{2n} - \frac{1}{2}(1-\delta) - 2H_2(\delta)\right] \\
&= \Theta\left(\frac{N_{phot}}{\varepsilon^{1/k}}\right)^{2n} .
\end{aligned} \tag{2.185}$$

Here, $g(x) \coloneqq (x+1)\log_2(x+1) - x\log_2 x$ is the bosonic entropy, and $H_2(x) \coloneqq -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy.

Proof. Let $M \in \mathbb{N}$ be such that there exist M states

$$\{\rho_1, \rho_2, \dots, \rho_M\} \subseteq \mathcal{S}(n, k, N_{\text{phot}}) \tag{2.186}$$

such that for every $i \neq j \in [M]$ it holds that

$$\frac{1}{2}\|\rho_i - \rho_j\|_1 > 2\varepsilon\,. \tag{2.187}$$

By applying Lemma 26 in the exact same way as we did in the proof of Theorem 38, we deduce that the number of state copies N required for tomography of an unknown state belonging to $\mathcal{S}(n, k, N_{\text{phot}})$ has to be at least

$$N \geq \frac{(1-\delta)\log_2(M) - H_2(\delta)}{n g(N_{\text{phot}})}\,. \tag{2.188}$$

Let $m \in \mathbb{N}$ such that $m \geq nN_{\text{phot}}$. Let $\mathcal{H}_m$ be the Hilbert space, defined in (2.171), spanned by all the n-mode Fock states, apart from the n-mode vacuum state, with total number of photons less than m. Moreover, let $\mathcal{D}(\mathcal{H}_m)$ be the set of density operators on $\mathcal{H}_m$. Given a state $\rho \in \mathcal{D}(\mathcal{H}_m)$, we can define a corresponding state $\sigma_\rho \in \mathcal{S}(n, k, N_{\text{phot}})$ as

$$\sigma_\rho \coloneqq \left[1 - \left(\frac{nN_{\text{phot}}}{m}\right)^k\right] |0\rangle\langle 0|^{\otimes n} + \left(\frac{nN_{\text{phot}}}{m}\right)^k \rho\,. \tag{2.189}$$

Indeed, the k-moment of $|\psi_\phi\rangle$ can be upper bounded as

$$\operatorname{Tr}[\hat{N}_n^k \sigma_\rho] = \left(\frac{nN_{\text{phot}}}{m}\right)^k \operatorname{Tr}[\hat{N}_n^k \rho] \leq \frac{(nN_{\text{phot}})^k}{m^k} m^k = (nN_{\text{phot}})^k\,, \tag{2.190}$$

and thus $\sigma_\rho \in \mathcal{S}(n, k, N_{\text{phot}})$. Moreover, it holds that

$$\|\sigma_\rho - \sigma_{\rho'}\|_1 = \left(\frac{nN_{\text{phot}}}{m}\right)^k \|\rho - \rho'\|_1 \tag{2.191}$$

for all $\rho, \rho' \in \mathcal{D}(\mathcal{H}_m)$. Hence, by choosing

$$m \coloneqq \left\lfloor \frac{nN_{\text{phot}}}{(16\varepsilon)^{1/k}} \right\rfloor\,, \tag{2.192}$$

one can check that for all $\rho, \rho' \in \mathcal{D}(\mathcal{H}_m)$ such that

$$\|\rho - \rho'\|_1 \geq \frac{1}{4}\,, \tag{2.193}$$

the relation in (2.191) guarantees that

$$\frac{1}{2}\|\sigma_\rho - \sigma_{\rho'}\|_1 \geq 2\varepsilon\,. \tag{2.194}$$

Therefore, let us find $M \in \mathbb{N}$ such that there exist M states $\{\rho_1, \rho_2, \dots, \rho_M\} \subseteq \mathcal{D}(\mathcal{H}_m)$ such that $\|\rho_i - \rho_j\|_1 \geq \frac{1}{4}$ for all $i \neq j \in [M]$. By definition of packing number (see Definition 2), such M can be chosen to be the $\frac{1}{4}$-packing number of $\mathcal{D}(\mathcal{H}_m)$ with respect to the trace norm:

$$M \coloneqq \mathcal{P}\left(\mathcal{D}(\mathcal{H}_m), \|\cdot\|_1, \frac{1}{4}\right)\,. \tag{2.195}$$

We can find a simple lower bound on M as

$$\begin{aligned} M &= \mathcal{P}\Big(\mathcal{D}(\mathcal{H}_m), \|\cdot\|_1, \frac{1}{4}\Big) \\ &\overset{\text{(i)}}{\geq} 2^{\frac{(\dim \mathcal{H}_m - 1)^2}{2}} \\ &\overset{\text{(ii)}}{\geq} 2^{\frac{\left(\binom{n+m}{n}-2\right)^2}{2}} \\ &\overset{\text{(iii)}}{\geq} 2^{\frac{1}{2}((\frac{m}{n})^{2n}-1)} \\ &= \frac{1}{\sqrt{2}} 2^{\frac{1}{2}(\frac{m}{n})^{2n}} . \end{aligned} \tag{2.196}$$

Here, in (i), we exploit Corollary 33; in (ii), we exploited (2.179); in (iii), we have used that $\binom{m+n}{n} \geq (\frac{m+n}{n})^n \geq (\frac{m}{n})^n + 1$. Consequently, (2.188) implies that

$$N \geq \frac{\frac{1-\delta}{2}(\frac{m}{n})^{2n} - (1-\delta)\log_2(\sqrt{2}) - H_2(\delta)}{n g(N_{\text{phot}})} . \tag{2.197}$$

Finally, by exploiting the definition of m in (2.175) we conclude that

$$N \geq \frac{1}{2n g(N_{\text{phot}})} \left[(1-\delta) \left(\frac{N_{\text{phot}}}{(16\varepsilon)^{1/k}} - \frac{1}{n} \right)^{2n} - \frac{1}{2}(1-\delta) - 2H_2(\delta) \right] . \tag{2.198}$$

□

In particular, in the single-mode case, i.e., $n = 1$, the number of state copies N has to satisfy

$$\begin{aligned} N &\geq \frac{1}{2g(N_{\text{phot}})} \left[(1-\delta) \left(\frac{N_{\text{phot}}}{(16\varepsilon)^{1/k}} - 1 \right)^{2} - \frac{1}{2}(1-\delta) - 2H_2(\delta) \right] \\ &= \Theta\left(\frac{N_{\text{phot}}^2}{\varepsilon^{2/k}} \right) . \end{aligned} \tag{2.199}$$

Sample complexity upper bounds

In this subsection, we present a tomography algorithm aimed at learning a classical description of an unknown n-mode k-th moment-constrained state. We analyse both pure and mixed scenarios. The tomography algorithm involves two main steps: first, projecting onto a finite-dimensional subspace, and second, applying a known tomography algorithm specifically designed for finite-dimensional systems.

The subsection is organised as follows. In Subsubsection 2.7.2, we establish useful properties of moment-constrained states. Specifically, we demonstrate how to approximate a moment-constrained state using a finite-dimensional state with bounded effective rank. In Subsubsection 2.7.2, we compare our approximation with those of other works, highlighting the improvements achieved. Subsubsection 2.7.2 outlines the precise performances of an optimal tomography algorithm applicable to finite-dimensional systems. Finally, in Subsubsection 2.7.2, we present the tomography algorithms for moment-constrained pure and mixed states and establish their correctness. We denote the Euler's Constant as e throughout this subsection.

Properties of moment-constrained states: approximation with finite-dimensional and finite-rank states The following lemma establishes that k-th moment-constrained states can be effectively approximated by finite-dimensional states.

Lemma 40 (Effective dimension of moment-constrained states)**.** *Let $\varepsilon \in (0,1)$ and let ρ be an n-mode state with k-moment bounded as $(\mathrm{Tr}[\hat{N}_n^k \rho])^{1/k} \leq n N_{phot}$. Then, ρ is ε-close in trace distance to a state supported on a space of dimension $\Theta\Big(\frac{eN_{phot}}{\varepsilon^{2/k}}\Big)^n$.*
More specifically, let $\mathcal{H}_m$ be the Hilbert space spanned by all the n-mode Fock states with total photon number less than m:

$$\mathcal{H}_m \coloneqq \mathrm{Span}\left\{ |k_1\rangle \otimes |k_2\rangle \otimes \ldots \otimes |k_n\rangle : \ k_1, k_2, \ldots, k_n \in \mathbb{N}, \ \sum_{i=1}^n k_i \leq m \right\}. \tag{2.200}$$

Let $m \coloneqq \left\lceil \frac{nN_{phot}}{\varepsilon^{2/k}} \right\rceil$. Define the state $\rho_{d_{\mathit{eff}}} \coloneqq \Pi_m \rho \, \Pi_m / \mathrm{Tr}[\Pi_m \rho \, \Pi_m]$ as the projection of ρ onto $\mathcal{H}_m$, where Π_m is the projector onto $\mathcal{H}_m$, whose dimension satisfies

$$d_{\mathit{eff}} \coloneqq \dim \mathcal{H}_m = \binom{m+n}{n} \leq \left(\frac{eN_{phot}}{\varepsilon^{2/k}} + 2e \right)^n = \Theta\Big(\frac{eN_{phot}}{\varepsilon^{2/k}} \Big)^n. \tag{2.201}$$

Then, $\rho_{d_{\mathit{eff}}}$ is ε-close in trace distance to ρ in that

$$\frac{1}{2} \left\| \rho - \rho_{d_{\mathit{eff}}} \right\|_1 \leq \varepsilon \, . \tag{2.202}$$

Proof. First, observe that the dimension $d_{\mathrm{eff}} = \dim \mathcal{H}_m$ is given by

$$\begin{aligned} \dim \mathcal{H}_m &= \left| \left\{ k_1, k_2, \ldots, k_n \in \mathbb{N} \, , \quad \sum_{i=1}^n k_i \leq m \right\} \right| \\ &\overset{\text{(i)}}{=} \binom{m+n}{n} \\ &\overset{\text{(ii)}}{\leq} \left(\frac{e(m+n)}{n} \right)^n \\ &\overset{\text{(iii)}}{\leq} \left(\frac{eN_{\mathrm{phot}}}{\varepsilon^{2/k}} + 2e \right)^n . \end{aligned} \tag{2.203}$$

Here, in (i), we exploited a standard combinatorial argument; in (ii), we have used the fact that $\binom{a}{b} \leq (\frac{ea}{b})^b$ for all $a, b \in \mathbb{N}$; in (iii), we have used the definition of $m \coloneqq \left\lceil \frac{nN_{\mathrm{phot}}}{\varepsilon^{2/k}} \right\rceil$. Moreover, let us note that

$$\begin{aligned} \mathrm{Tr}[(\mathbb{1} - \Pi_m)\rho] &\overset{\text{(iv)}}{\leq} \frac{1}{m^k} \mathrm{Tr}\Big[\hat{N}_n^k \rho \Big] \\ &\overset{\text{(v)}}{\leq} \frac{(nN_{\mathrm{phot}})^k}{m^k} \\ &\overset{\text{(vi)}}{\leq} \varepsilon^2 \, , \end{aligned} \tag{2.204}$$

where: (iv) follows by the operator inequality $m^k(\mathbb{1} - \Pi_m) \leq \hat{N}_n^k$; (v) follows by the hypothesis on the k-th moment of ρ; in (vi) we have used again the definition of m. Finally, the trace distance between ρ and the projected state $\rho_{d_{\mathrm{eff}}}$ satisfies

$$\begin{aligned} \frac{1}{2} \left\| \rho_{d_{\mathrm{eff}}} - \rho \right\|_1 &= \frac{1}{2} \left\| \frac{\Pi_m \rho \, \Pi_m}{\mathrm{Tr}[\Pi_m \rho \, \Pi_m]} - \rho \right\|_1 \\ &\overset{\text{(vii)}}{\leq} \sqrt{\mathrm{Tr}[(\mathbb{1} - \Pi_m)\rho]} \\ &\leq \varepsilon \, , \end{aligned} \tag{2.205}$$

where in (vii) we have used the *gentle measurement lemma* [48, Lemma 6.15]. □

We now mention a lemma which will be useful consequently.

Lemma 41 (Infinite-dimensional Schur-Horn theorem ([49], Proposition 6.4))**.** *Let $\mathcal{H}$ be a Hilbert space. Let $N \in \mathbb{N}$ and $M \geq 0$ be a semi-definite operator on $\mathcal{H}$. For any set of orthonormal vectors $\{|v_n\rangle \in \mathcal{H} : n \in \{0, 1, \dots, N-1\}\}$, it holds that*

$$\sum_{n=0}^{N-1} \lambda_n \geq \sum_{n=0}^{N-1} \langle v_n | M | v_n \rangle, \tag{2.206}$$

where $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N-1}$ are the N largest eigenvalues of M.

In the following lemma, we establish that any moment-constrained state is close in trace distance to a state with bounded rank.

Lemma 42 (Effective rank of moment-constrained states)**.** *Let $\varepsilon \in (0, 1)$. Consider an n-mode state ρ satisfying the k-th moment constraint $(\mathrm{Tr}[\rho \hat{N}_n^k])^{1/k} \leq n N_{phot}$. Then, ρ is ε-close in trace distance to a state $\rho_{r_{eff}}$ with rank $r_{eff} = \Theta\Big(\frac{eN_{phot}}{\varepsilon^{1/k}}\Big)^n$.*
More specifically, there exists a state $\rho_{r_{eff}}$ with rank r_{eff}, where

$$r_{eff} := \mathrm{rank}(\rho_{r_{eff}}) \leq \left(\frac{eN_{phot}}{\varepsilon^{1/k}} + 2e\right)^n = \Theta\left(\frac{eN_{phot}}{\varepsilon^{1/k}}\right)^n, \tag{2.207}$$

such that $\frac{1}{2}\|\rho - \rho_{r_{eff}}\|_1 \leq \varepsilon$.

Proof. Let $\rho = \sum_{j=1}^{\infty} \lambda_j^{\downarrow} \psi_j$ be the the spectral decomposition of ρ, with $(\lambda_j^{\downarrow})_{j \in \mathbb{N}_+}$ being its eigenvalues ordered in decreasing order with respect to j. For each $r_{\mathrm{eff}} \in \mathbb{N}$, let us define the operator

$$\Theta_{r_{\mathrm{eff}}} := \sum_{j=1}^{r_{\mathrm{eff}}} \lambda_j^{\downarrow} \psi_j \tag{2.208}$$

which has rank equal to r_{eff}. Note that the *infinite-dimensional Schur Horn theorem* (Lemma 41) implies that for any r_{eff}-rank projector $\Pi^{(r_{\mathrm{eff}})}$ it holds that

$$\|\rho - \Theta_{r_{\mathrm{eff}}}\|_1 = \sum_{j=r_{\mathrm{eff}}+1}^{\infty} \lambda_j^{\downarrow} \leq \mathrm{Tr}[(\mathbb{1} - \Pi^{(r_{\mathrm{eff}})})\rho]\,. \tag{2.209}$$

Now, let us choose $\Pi^{(r_{\mathrm{eff}})}$ to be the projector $\Pi_{m'}$ onto the subspace $\mathcal{H}_{m'}$, defined in (2.200), spanned by the n-mode Fock states with total photon number at most m', with $m' := \left\lceil \frac{nN_{\mathrm{phot}}}{\varepsilon^{1/k}} \right\rceil$. Hence, we have that

$$r_{\mathrm{eff}} := \mathrm{Tr}\,\Pi_{m'} = \binom{m'+n}{n} \leq \left(\frac{e(m'+n)}{n}\right)^n \leq \left(\frac{eN_{\mathrm{phot}}}{\varepsilon^{1/k}} + 2e\right)^n. \tag{2.210}$$

Thanks to (2.204), we have that

$$\mathrm{Tr}[(\mathbb{1} - \Pi_{m'})\rho] \leq \left(\frac{nN_{\mathrm{phot}}}{m'}\right)^k \leq \varepsilon\,, \tag{2.211}$$

and thus it holds that $\mathrm{Tr}[(\mathbb{1} - \Pi^{(r_{\mathrm{eff}})})\rho] \leq \varepsilon$. Consequently, we deduce that $\|\rho - \Theta_{r_{\mathrm{eff}}}\|_1 \leq \varepsilon$. Therefore, by setting

$$\rho_{r_{\mathrm{eff}}} := \frac{\Theta_{r_{\mathrm{eff}}}}{\mathrm{Tr}\,\Theta_{r_{\mathrm{eff}}}} \tag{2.212}$$

we have that

$$\begin{aligned}\|\rho-\rho_{r_{\text{eff}}}\|_1 &= \left\|\rho-\frac{\Theta_{r_{\text{eff}}}}{\operatorname{Tr}\Theta_{r_{\text{eff}}}}\right\|_1\\ &\leq \|\rho-\Theta_{r_{\text{eff}}}\|_1+\left\|\Theta_{r_{\text{eff}}}-\frac{\Theta_{r_{\text{eff}}}}{\operatorname{Tr}\Theta_{r_{\text{eff}}}}\right\|_1\\ &= \|\rho-\Theta_{r_{\text{eff}}}\|_1+\left|1-\frac{1}{\operatorname{Tr}\Theta_{r_{\text{eff}}}}\right|\operatorname{Tr}\Theta_{r_{\text{eff}}}\\ &\leq \varepsilon+(1-\operatorname{Tr}\Theta_{r_{\text{eff}}})\\ &\leq \varepsilon+\sum_{j=r_{\text{eff}}+1}^{\infty}\lambda_j^{\downarrow}\\ &\leq 2\varepsilon\,.\end{aligned} \tag{2.213}$$

□

We have demonstrated that an n-mode k-th moment-constrained state is ε-close in trace distance to a qudit state residing in a known Hilbert space of dimension $d_{\text{eff}} \leq \left(\frac{eN_{\text{phot}}}{\varepsilon^{2/k}}+2e\right)^n$. Additionally, we have shown that this moment-constrained state is also ε-close in trace distance to a state with rank $r_{\text{eff}} \leq \left(\frac{eN_{\text{phot}}}{\varepsilon^{1/k}}+2e\right)^n$.
Now, we combine the previous two lemmas to show that any moment-constrained state is close to a finite-dimensional state which is close to a state with smaller rank.

Lemma 43 (Low rank and finite-dimensional approximation of moment-constrained states)**.** *Let $\varepsilon \in (0,1)$. Let ρ be an n-mode state with k-moment bounded as $\left(\operatorname{Tr}[\hat{N}_n^k\rho]\right)^{1/k} \leq nN_{phot}$. Let $\mathcal{H}_m$ be the Hilbert space, defined in (2.200), spanned by all the n-mode Fock states with total photon number less than m with $m \coloneqq \left\lceil \frac{nN_{phot}}{\varepsilon^{2/k}} \right\rceil$, with dimension*

$$\dim \mathcal{H}_m \leq \left(\frac{eN_{phot}}{\varepsilon^{2/k}}+2e\right)^n = \Theta\left(\frac{N_{phot}}{\varepsilon^{2/k}}\right)^n\,. \tag{2.214}$$

Let $\rho_{d_{eff}}$ be the projected state of ρ onto $\mathcal{H}_m$. Then, $\rho_{d_{eff}}$ is $\eta(\varepsilon)$-close in trace distance to a state on $\mathcal{H}_m$ with rank r_{eff} satisfying

$$r_{eff} \leq \left(\frac{N_{phot}}{\varepsilon^{1/k}}+2e\right)^n = \Theta\left(\frac{N_{phot}}{\varepsilon^{1/k}}\right)^n\,, \tag{2.215}$$

where we have defined $\eta(\varepsilon) \coloneqq \left(2+\frac{1}{\sqrt{1-\varepsilon}}\right)\varepsilon$.

Proof. Thanks to Lemma 42, there exists an n-mode state $\rho_{r_{\text{eff}}}$, with rank satisfying

$$\operatorname{rank}(\rho_{r_{\text{eff}}}) \leq \left(\frac{eN_{\text{phot}}}{\varepsilon^{1/k}}+2e\right)^n = \Theta\left(\frac{eN_{\text{phot}}}{\varepsilon^{1/k}}\right)^n\,, \tag{2.216}$$

which is ε-close in trace distance to ρ. Let ρ' be the projected state of $\rho_{r_{\text{eff}}}$ onto $\mathcal{H}_m$, that is,

$$\rho' \coloneqq \frac{\Pi_m\rho_{r_{\text{eff}}}\Pi_m}{\operatorname{Tr}[\Pi_m\rho_{r_{\text{eff}}}\Pi_m]}, \tag{2.217}$$

where Π_m is the projector onto $\mathcal{H}_m$. Clearly, ρ' is a state supported on $\mathcal{H}_m$. The following chain

of inequality shows that ρ' is $\eta(\varepsilon)$-close in trace distance to $\rho_{d_{\text{eff}}}$:

$$\begin{aligned}\frac{1}{2}\|\rho_{d_{\text{eff}}}-\rho'\|_1 &\overset{\text{(i)}}{\leq} \frac{1}{2}\|\rho_{d_{\text{eff}}}-\rho\|_1+\frac{1}{2}\|\rho-\rho_{r_{\text{eff}}}\|_1+\frac{1}{2}\|\rho_{r_{\text{eff}}}-\rho'\|_1\\ &\overset{\text{(ii)}}{\leq} 2\varepsilon+\frac{1}{2}\|\rho_{r_{\text{eff}}}-\rho'\|_1\\ &\overset{\text{(iii)}}{\leq} 2\varepsilon+\sqrt{\operatorname{Tr}[(\mathbb{1}-\Pi_m)\rho_{r_{\text{eff}}}]}\\ &\overset{\text{(iv)}}{\leq} 2\varepsilon+\frac{\varepsilon}{\sqrt{1-\varepsilon}}\\ &= \eta(\varepsilon)\,.\end{aligned} \tag{2.218}$$

Here, in (i), we have used triangle inequality. In (ii), we exploited Lemma 40 and Lemma 42 to ensure that $\frac{1}{2}\|\rho_{d_{\text{eff}}}-\rho\|_1\leq\varepsilon$ and $\frac{1}{2}\|\rho-\rho_{r_{\text{eff}}}\|_1$, respectively. In (iii), we have used the *gentle measurement lemma* [48, Lemma 6.15]. In (iv), we have used that

$$\begin{aligned}\operatorname{Tr}[(\mathbb{1}-\Pi_m)\rho_{r_{\text{eff}}}] &\overset{\text{(v)}}{=} \frac{1}{\operatorname{Tr}\Theta_{r_{\text{eff}}}}\operatorname{Tr}\left[(\mathbb{1}-\Pi_m)\Theta_{r_{\text{eff}}}\right]\\ &\overset{\text{(vi)}}{\leq} \frac{1}{1-\varepsilon}\operatorname{Tr}\left[(\mathbb{1}-\Pi_m)\Theta_{r_{\text{eff}}}\right]\\ &\overset{\text{(vii)}}{\leq} \frac{1}{1-\varepsilon}\operatorname{Tr}\left[(\mathbb{1}-\Pi_m)\rho\right]\\ &\overset{\text{(viii)}}{\leq} \frac{\varepsilon^2}{1-\varepsilon}\,.\end{aligned} \tag{2.219}$$

Here, in (v), we have used the definition $\rho_{r_{\text{eff}}}\coloneqq\frac{\Theta_{r_{\text{eff}}}}{\operatorname{Tr}\Theta_{r_{\text{eff}}}}$ in (2.212), where we recall that $\Theta_{r_{\text{eff}}}$ is defined in (2.208) as $\Theta_{r_{\text{eff}}}\coloneqq\sum_{j=1}^{r_{\text{eff}}}\lambda_j^{\downarrow}\psi_j$. In (vi), we have used that $\operatorname{Tr}\Theta_{r_{\text{eff}}}\geq 1-\varepsilon$, as it follows from the third and fifth line of (2.213). In (vii), we exploited that

$$\Theta_{r_{\text{eff}}}=\sum_{j=1}^{r_{\text{eff}}}\lambda_j^{\downarrow}\psi_j\leq\sum_{j=1}^{\infty}\lambda_j^{\downarrow}\psi_j=\rho\,. \tag{2.220}$$

Finally, in (viii) we applied that (2.204). □

Related works on approximating moment-constrained states We demonstrated that an n-mode state ρ with k-moment bounded by N_{phot} can be approximated within trace distance precision ε by a projected state living in a subspace of *effective dimension* $d_{\text{eff}}=\lceil(eN_{\text{phot}}/\varepsilon^{2/k}+2e)^n\rceil$ (Lemma 40). Naturally, the lower the effective dimension, the better the efficiency of the subsequent tomography algorithm for learning the projected state.

Previous works such as [17, 30] have established similar results for approximating n-mode moment-constrained states with a projected finite-dimensional state. However, our Lemma 40 improves upon these findings. Specifically, our effective dimension scales as $d_{\text{eff}}=\Theta(eN_{\text{phot}}/\varepsilon^{2/k})^n$, is in contrast to the scaling $d_{\text{eff}}=\Theta(nN_{\text{phot}}/\varepsilon^{2/k})^n$ obtained in [17, 30]. Unlike [17, Proposition 8], which involves approximating an n-mode state with an n-qudit state, our projection approximates an n-mode state with a single-qudit state. Furthermore, Ref. [30, Supplementary Note 1] exclusively focuses on the case $k=1$ (energy-constrained states). While they use a projection similar to ours, they end up with an unfavourable scaling $d_{\text{eff}}=\Theta(nN_{\text{phot}}/\varepsilon)^n$ due to an overly crude upper bound on the effective dimension. In conclusion, our bound on the effective dimension of moment-constrained states represents to our knowledge the first achieved bound that is non-super-exponential in the number of modes.

Known optimal tomography algorithm for qudits In this section, we review the guarantees of the optimal tomography algorithm known as the *truncated version of Keyl's algorithm*, which is detailed in Wright's PhD thesis [50, Theorem 1.4.13].

Lemma 44 (Truncated version of Keyl's algorithm [50])**.** *Let ρ be an unknown qudit state of dimension d, and let $\lambda_1^{\downarrow} \geq \lambda_2^{\downarrow} \geq \ldots \geq \lambda_d^{\downarrow}$ be its eigenvalues. For any $r \in [d]$, N copies of ρ are sufficient to build an estimator $\tilde{\rho}$ with rank r such that*

$$\mathbb{E}\|\rho - \tilde{\rho}\|_1 \leq \sum_{j=r+1}^{d} \lambda_j^{\downarrow} + 6\sqrt{\frac{rd}{N}}. \tag{2.221}$$

We now introduce a lemma that later will be useful for expressing the sample complexity of the tomography algorithm.

Lemma 45. *Let ρ be a qudit state of dimension d (possibly $d = \infty$), and let $\lambda_1^{\downarrow} \geq \lambda_2^{\downarrow} \geq \ldots \geq \lambda_d^{\downarrow}$ be its eigenvalues ordered in decreasing order. For any state ρ_r with rank r, we have*

$$\sum_{j=r+1}^{d} \lambda_j^{\downarrow} \leq \frac{1}{2}\|\rho - \rho_r\|_1. \tag{2.222}$$

Proof. Let Π_r be the projector onto the support of ρ_r. Note that

$$\begin{aligned} \sum_{j=r+1}^{d} \lambda_j^{\downarrow} &\overset{\text{(i)}}{\leq} \operatorname{Tr}[(\mathbb{1} - \Pi_r)\rho] \\ &\overset{\text{(ii)}}{=} \operatorname{Tr}[(\mathbb{1} - \Pi_r)(\rho - \rho_r)] \\ &\leq \max_{0 \leq E \leq \mathbb{1}} \operatorname{Tr}[E(\rho - \rho_r)] \\ &\overset{\text{(iii)}}{=} \frac{1}{2}\|\rho - \rho_r\|_1 \,. \end{aligned} \tag{2.223}$$

Here, in (i), we have used the *(infinite) dimensional Schur Horn theorem* (Lemma 41); in (ii), we just exploited the fact that $\Pi_r \rho_r = \rho_r \Pi_r = \rho_r$, which holds true since Π_r is the projector onto the support of ρ_r; in (iii), we just used the well-known variational characterisation of the trace distance in terms of POVM [76]. □

We are now ready to provide the precise performance guarantees of the algorithm [50].

Lemma 46 (Sample complexity of the optimal tomography algorithms for qudit systems)**.** *Let $\varepsilon, \delta \in (0, 1)$. Let ρ be an (unknown) state of dimension d such that it is $\frac{\varepsilon}{3}$-close in trace distance to a state with rank r. Then, there exists a tomography algorithm such that given*

$$N \geq 2^{18} \frac{rd}{\varepsilon^2} \log\left(\frac{2}{\delta}\right) \tag{2.224}$$

copies of ρ can build (a classical description of) an r-rank state estimator $\tilde{\rho}$ satisfying

$$\Pr\left[\frac{1}{2}\|\rho - \tilde{\rho}\|_1 \leq \varepsilon\right] \geq 1 - \delta \,. \tag{2.225}$$

Proof. Thanks to Lemma 44, N' copies of ρ are sufficient in order to build an estimator $\tilde{\rho}$ (with rank r) such that

$$\mathbb{E}\|\rho - \tilde{\rho}\|_1 \leq \sum_{j=r+1}^{d} \lambda_j^{\downarrow} + 6\sqrt{\frac{rd}{N'}} \,, \tag{2.226}$$

where $\lambda_1^{\downarrow} \geq \lambda_2^{\downarrow} \geq \ldots \geq \lambda_d^{\downarrow}$ are the eigenvalues of ρ. By assumption, there exists a state ρ_r with rank r such that $\frac{1}{2}\|\rho - \rho_r\|_1 \leq \frac{\varepsilon}{3}$. Consequently, Lemma 45 implies that $\sum_{j=r+1}^{d} \lambda_j^{\downarrow} \leq \frac{\varepsilon}{3}$. Hence, by choosing $N' := \left\lceil \frac{(18)^2 rd}{\varepsilon^2} \right\rceil$ and by applying Markov's inequality, we have that

$$\begin{aligned} \Pr\left[\frac{1}{2}\|\rho - \tilde{\rho}\|_1 \geq \varepsilon\right] &\leq \frac{\mathbb{E}\|\rho - \tilde{\rho}\|_1}{2\varepsilon} \\ &\leq \frac{\sum_{j=r+1}^{d} \lambda_j^{\downarrow} + 6\sqrt{\frac{rd}{N'}}}{2\varepsilon} \\ &\leq \frac{1}{6} + \frac{3}{\varepsilon}\sqrt{\frac{rd}{N'}} \\ &\leq \frac{1}{3}. \end{aligned} \tag{2.227}$$

Consequently, we have proved that $N' = \left\lceil 324\frac{rd}{\varepsilon^2} \right\rceil$ copies of ρ are sufficient in order to build an estimator $\tilde{\rho}$ such that $\Pr\left[\frac{1}{2}\|\rho - \tilde{\rho}\|_1 \leq \varepsilon\right] \geq \frac{2}{3}$. Now, we can apply the standard argument presented in Lemma 18 with $P_{\text{succ}} = \frac{2}{3}$ in order to enhance the probability of success from $\frac{2}{3}$ to $1 - \frac{\delta}{2}$. From such a lemma, by defining

$$m := \left\lceil \frac{2}{\left(1 - \frac{1}{2p_{\text{succ}}}\right)^2 p_{\text{succ}}} \log\left(\frac{2}{\delta}\right) \right\rceil = \left\lceil 48 \log\left(\frac{2}{\delta}\right) \right\rceil , \tag{2.228}$$

it follows that a total of

$$mN' = \left\lceil 48 \log\left(\frac{2}{\delta}\right) \right\rceil \left\lceil 324\frac{rd}{\varepsilon^2} \right\rceil \tag{2.229}$$

copies of ρ suffices in order to build an estimator $\tilde{\rho}$ such that $\Pr\left[\frac{1}{2}\|\rho - \tilde{\rho}\|_1 \leq 3\varepsilon\right] \geq 1 - \frac{\delta}{2} \geq 1 - \delta$. Finally, by redefining $\varepsilon \mapsto \frac{\varepsilon}{3}$, we have that

$$\left\lceil 48 \log\left(\frac{2}{\delta}\right) \right\rceil \left\lceil 324\frac{rd}{(\varepsilon/3)^2} \right\rceil \leq 2^{18}\frac{rd}{\varepsilon^2} \log\left(\frac{2}{\delta}\right) , \tag{2.230}$$

and thus we conclude that $2^{18}\frac{rd}{\varepsilon^2}\log\left(\frac{2}{\delta}\right)$ copies of ρ are sufficient in order to build an estimator $\tilde{\rho}$ such that $\Pr\left[\frac{1}{2}\|\rho - \tilde{\rho}\|_1 \leq \varepsilon\right] \geq 1 - \delta$. □

One can prove that $\tilde{T}heta\left(\frac{rd}{\varepsilon^2}\right)$ copies are not only sufficient (as established by the lemma above) but also necessary to perform quantum state tomography of a finite-dimensional quantum state [1, 14].

Tomography algorithm for moment-constrained mixed and pure states In this subsection, we present a tomography algorithm to learn a classical description of an unknown n-mode moment-constrained state (pure and mixed), together with its sample complexity analysis. The tomography algorithm involves two main steps: first, projecting onto a finite-dimensional subspace (specifically, the subspace defined in Lemma 40), and second, applying a known tomography algorithm designed for finite-dimensional states (specifically, the algorithm described in Lemma 46). Table 2.1 presents the steps of our tomography algorithm. We begin with Theorem 47, which analyses the sample complexity of our tomography algorithm for moment-constrained mixed states. The basic idea of Theorem 47 is the following: the unknown moment-constrained mixed state is effectively a qudit state of dimension $d_{\text{eff}} = O\left(\frac{eN_{\text{phot}}}{\varepsilon^{2/k}}\right)^n$ and rank $r_{\text{eff}} = O\left(\frac{eN_{\text{phot}}}{\varepsilon^{1/k}}\right)^n$ (as established by Lemma 43), and thus the sample complexity of tomography is $O(r_{\text{eff}}\, d_{\text{eff}}) = O\left(\frac{N_{\text{phot}}}{\varepsilon^{3/(2k)}}\right)^{2n}$ (thanks to Lemma 46).

Table 2.1: Tomography algorithm for n-mode k-th moment constrained states (pure and mixed). The algorithm for pure states and the one for mixed states are the same, apart from two distinctions: the number N of copies of the unknown state, and the specific input parameters provided to the subroutine called at Line 10 of the algorithm. For the mixed case the details of the algorithm are provided in the proof of Theorem 47; while for the pure case the details are reported in the proof of Theorem 48.

Input: Accuracy ε, failure probability δ, N copies of the unknown n-mode moment-constrained state ρ (as defined in Theorem 47 and Theorem 48 for mixed and pure states, respectively).
Output: A classical description of $\tilde{\rho}$, such that $\frac{1}{2}\|\tilde{\rho}-\rho\|_1 \leq \varepsilon$ with probability at least $1-\delta$.

1: **for** $i \leftarrow 1$ **to** N **do**
2: Query a copy of ρ.
3: Perform the POVM $\{\Pi_m, \mathbb{1}-\Pi_m\}$, where Π_m is the projector onto the subspace $\mathcal{H}_m$, defined in Eq. (2.200).
4: **if** the POVM outcome corresponds to $\mathbb{1}-\Pi_m$ **then**
5: Discard the post-measurement state.
6: **else**
7: Keep the post-measurement state.
8: **end if**
9: **end for**
10: Perform the full state tomography algorithm described in Lemma 46 on the kept copies of the post-measurement states, obtaining as output the classical description of a state $\tilde{\rho}$.
11: **return** $\tilde{\rho}$.

Theorem 47 (Learning moment-constrained mixed states)**.** *Let $\varepsilon, \delta \in (0,1)$. Let ρ be an n-mode state with k-moment bounded as $\left(\mathrm{Tr}[\hat{N}_n^k \rho]\right)^{1/k} \leq n N_{phot}$. There exists a quantum algorithm that, utilising*

$$N = \left\lceil 2^{21} \frac{r_{\mathit{eff}}\, d_{\mathit{eff}}}{\varepsilon^2} \log\left(\frac{4}{\delta}\right) \right\rceil = O\left(\frac{N_{phot}}{\varepsilon^{3/(2k)}}\right)^{2n} \tag{2.231}$$

copies of ρ, generates a classical representation of a d_{eff}-dimensional state $\tilde{\rho}$ with rank r_{eff} such that

$$\Pr\left[\frac{1}{2}\|\rho-\tilde{\rho}\|_1 \leq \varepsilon\right] \geq 1-\delta\,, \tag{2.232}$$

where $d_{\mathit{eff}} \leq \left(\frac{eN_{phot}}{\varepsilon^{2/k}} + 2e\right)^n$ and $r_{\mathit{eff}} \leq \left(\frac{eN_{phot}}{(\frac{\varepsilon}{20})^{1/k}} + 2e\right)^n$.

Proof. We aim to establish the correctness of the algorithm presented in Table 2.1. The algorithm queries $\left\lceil 2N' + 24\log\left(\frac{2}{\delta}\right)\right\rceil$ copies of ρ, where we will fix N' later. Let $m := \left\lceil \frac{nN_{\mathrm{phot}}}{(\varepsilon/2)^{2/k}} \right\rceil$. On each copy, it executes the POVM $\{\Pi_m, \mathbb{1}-\Pi_m\}$, where $\Pi_m := \sum_{\mathbf{m}:\sum_{i=1}^n m_i \leq m} |\mathbf{m}\rangle\langle\mathbf{m}|$ is the projector onto the subspace $\mathcal{H}_m$, defined in Eq. (2.200), spanned by the n-mode Fock states with total photon number at most m, with dimension (see Lemma 40)

$$d_{\mathrm{eff}} := \dim \mathcal{H}_m \leq \left(\frac{eN_{\mathrm{phot}}}{\varepsilon^{2/k}} + 2e\right)^n = \Theta\left(\frac{eN_{\mathrm{phot}}}{\varepsilon^{2/k}}\right)^n. \tag{2.233}$$

The probability of obtaining the first POVM outcome is

$$\begin{aligned}\operatorname{Tr}\left[\Pi_m \rho\right] &= 1 - \operatorname{Tr}[(\mathbb{1} - \Pi_m)\rho] \\ &\geq 1 - \frac{1}{m^k} \operatorname{Tr}\left[\hat{N}_n^k \rho\right] \\ &\geq 1 - \frac{(n N_{\text{phot}})^k}{m^k} \\ &\geq 1 - \frac{\varepsilon^2}{4} \\ &\geq \frac{3}{4},\end{aligned} \tag{2.234}$$

where the second step follows from the operator inequality $m^k(\mathbb{1} - \Pi_m) \leq \hat{N}_n^k$ and the third step follows from the hypothesis on the k-th moment of ρ.
Therefore, applying Lemma 17, we can assert that the algorithm, which uses a number of copies $\left\lceil 2N' + 24 \log\left(\frac{2}{\delta}\right)\right\rceil$, with a probability $\geq 1 - \frac{\delta}{2}$ obtains at least N' copies of the post-measurement state $\rho_{d_{\text{eff}}} := \Pi_m \rho \Pi_m / \operatorname{Tr}\left[\Pi_m \rho\right]$. Due to Lemma 40, the post-measurement state satisfies

$$\frac{1}{2} \left\| \rho_{d_{\text{eff}}} - \rho \right\|_1 \leq \frac{\varepsilon}{2}. \tag{2.235}$$

Moreover, Lemma 43 implies that $\rho_{d_{\text{eff}}}$ is $\eta(\varepsilon/20)$-close in trace distance to a state supported on $\mathcal{H}_m$ with rank r_{eff} satisfying

$$r_{\text{eff}} \leq \left(\frac{e N_{\text{phot}}}{(\frac{\varepsilon}{20})^{1/k}} + 2e \right)^n = \Theta\left(\frac{N_{\text{phot}}}{\varepsilon^{1/k}} \right)^n, \tag{2.236}$$

where $\eta(\frac{\varepsilon}{20}) := \left(2 + \frac{1}{\sqrt{1 - \frac{\varepsilon}{20}}}\right) \frac{\varepsilon}{20} \leq \frac{\varepsilon}{6}$.
Thus, from Lemma 46, it follows that there exists a quantum algorithm that, utilising only

$$N' \geq 2^{18} \frac{r_{\text{eff}}\, d_{\text{eff}}}{(\varepsilon/2)^2} \log\left(\frac{4}{\delta}\right) \tag{2.237}$$

copies of the post-measurement state $\rho_{d_{\text{eff}}}$, builds a classical description of an r_{eff}-rank state estimator $\tilde{\rho}$ such that, with a probability $\geq 1 - \frac{\delta}{2}$, it holds that

$$\frac{1}{2} \| \rho_{d_{\text{eff}}} - \tilde{\rho} \|_1 \leq \frac{\varepsilon}{2}. \tag{2.238}$$

By the triangle inequality, we then conclude that $\frac{1}{2} \left\| \rho - \tilde{\rho} \right\|_1 \leq \varepsilon$. The total failure probability of the algorithm is $\leq \delta$ by applying a union bound. □

We now proceed with Theorem 48, which shows an upper bound on the sample complexity of tomography of moment-constrained pure states. The basic idea of Theorem 48 is the following: the unknown moment-constrained pure state is effectively a pure qudit state of dimension $d_{\text{eff}} = O\left(\frac{e N_{\text{phot}}}{\varepsilon^{2/k}}\right)^n$ (as established by Lemma 40), and thus the sample complexity of tomography is $O(d_{\text{eff}}) = O\left(\frac{N_{\text{phot}}}{\varepsilon^{2/k}}\right)^n$ (thanks to Lemma 46).

Theorem 48 (Learning moment-constrained pure states)**.** *Let $\varepsilon, \delta \in (0, 1)$. Consider an n-mode pure state ψ with its k-th moment bounded by $\left(\operatorname{Tr}[\hat{N}_n^k \psi]\right)^{1/k} \leq n N_{phot}$. There exists a quantum algorithm that, utilising*

$$N = \left\lceil 2^{21} \frac{d_{eff}}{\varepsilon^2} \log\left(\frac{4}{\delta}\right) \right\rceil = O\left(\frac{N_{phot}}{\varepsilon^{2/k}}\right)^n \tag{2.239}$$

copies of ψ, generates a classical representation of a pure state $\tilde{\psi}$ such that

$$\Pr\left[\frac{1}{2}\|\psi-\tilde{\psi}\|_1 \leq \varepsilon\right] \geq 1-\delta, \tag{2.240}$$

where $d_{\mathit{eff}} \leq \left(\frac{eN_{phot}}{\varepsilon^{2/k}}+2e\right)^n$.

Proof. We adopt the same notation used in Theorem 47. The algorithm is presented in Table 2.1, and the proof of its correctness follows similar steps to the previous mixed state case (Theorem 47). Specifically, the algorithm queries $\left\lceil 2N' + 24\log\left(\frac{2}{\delta}\right)\right\rceil$ copies of ψ and on each of them performs the POVM $\{\Pi_m, \mathbb{1}-\Pi_m\}$, where Π_m is the projector onto the subspace $\mathcal{H}_m$ (defined in Eq. (2.200)). As in the proof of Theorem 47, with probability $\geq 1-\frac{\delta}{2}$, on at least N' copies the post-measurement state will be $\psi_{d_{\text{eff}}} := \Pi_m\psi\,\Pi_m/\operatorname{Tr}[\Pi_m\psi]$, satisfying $\frac{1}{2}\|\psi_{d_{\text{eff}}}-\psi\|_1 \leq \frac{\varepsilon}{2}$.
By applying Lemma 46 with $r=1$, there exists a quantum algorithm that, utilising

$$N' \geq 2^{18}\frac{d_{\text{eff}}}{(\varepsilon/2)^2}\log\left(\frac{4}{\delta}\right), \tag{2.241}$$

copies of the post-measurement state $\psi_{d_{\text{eff}}}$, builds a classical description of a pure state $\tilde{\psi}$ such that, with a probability $\geq 1-\frac{\delta}{2}$, it holds that

$$\frac{1}{2}\|\psi_{d_{\text{eff}}}-\tilde{\psi}\|_1 \leq \frac{\varepsilon}{2}. \tag{2.242}$$

By the triangle inequality, we conclude that $\frac{1}{2}\left\|\psi-\tilde{\psi}\right\|_1 \leq \varepsilon$. The total failure probability of the algorithm is $\leq \delta$ by applying a union bound. □

Remarkably, as a consequence of the sample-complexity lower bound proved in Theorem 38 and the sample-complexity upper bound proved in Theorem 48, we have the following.

Theorem 49 (Optimal sample complexity of tomography of moment-constrained pure states)**.** *Let ψ be an unknown pure n-mode state vector satisfying the k-th moment constraint* $(\operatorname{Tr}[\hat{N}_n^k\rho])^{1/k} \leq nN_{phot}$. *Then,* $\Theta\left(\frac{N_{phot}}{\varepsilon^{2/k}}\right)^n$ *copies of ψ are necessary and sufficient to perform quantum state tomography with precision ε in trace distance.*

2.7.3 Tomography of bosonic Gaussian states

Gaussian states play a crucial role in applications of quantum optics, such as quantum sensing, quantum communication, and optical quantum computing [59, 77], and they form a small subset of the entire infinite-dimensional Hilbert space of continuous variable quantum states. In contrast to an arbitrary continuous variable quantum state, which is defined in terms of an infinite number of parameters, a Gaussian state is uniquely characterised by only a few parameters — specifically those present in its first moment and its covariance matrix. Indeed, it is well established that: 'In order to *know* a Gaussian state it is sufficient to *know* its first moment and its covariance matrix.' However, in practice, we never know the first moment and the covariance matrix *exactly*, but we can only have estimates of them, meaning that we can only approximately know the Gaussian state. Crucially, the *trace distance* between the exact quantum state and its approximation is the most meaningful figure of merit to measure of the error incurred in the approximation, due to the operational meaning of the trace distance given by the Holevo–Helstrom theorem [9, 10]. It is thus a fundamental problem — yet never tackled before — of Gaussian quantum information to determine what is the error incurred in *trace distance* when estimating the first moment and covariance matrix of an unknown Gaussian state up to a precision ε. In this section, we address this fundamental problem, by finding upper and lower bounds on the trace distance between two arbitrary Gaussian states, determined by the norm distance of their covariance matrices and first moments. We present such bounds in the forthcoming Theorem 50.

One might be inclined to believe that there is a simple approach to solving this problem, involving first bounding the trace distance in terms of the fidelity and then employing the known formula for fidelity between Gaussian states [52]. Although this approach may seem promising at first glance, it is actually highly non-trivial because the expressions involved in the fidelity formula appear to be too complicated to allow the derivation of a bound in terms of the norm distance between the first moments and the covariance matrices.

Theorem 50 (Bounds on the trace distance between Gaussian states)**.** *Let ρ_1 and ρ_2 be n-mode Gaussian states satisfying the energy constraint $\mathrm{Tr}[\rho_1 \hat{N}_n] \leq N$ and $\mathrm{Tr}[\rho_2 \hat{N}_n] \leq N$. Let $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ be the first moments and let V_1 and V_2 be the covariance matrices of ρ_1 and ρ_2, respectively. The trace distance between ρ_1 and ρ_2 can be upper bounded as*

$$\frac{1}{2}\|\rho_1 - \rho_2\|_1 \leq f(N)\left(\|\boldsymbol{m}_1 - \boldsymbol{m}_2\|_2 + \sqrt{2}\sqrt{\|V_1 - V_2\|_1}\right), \tag{2.243}$$

where $f(N) \coloneqq \frac{1}{\sqrt{2}}\left(\sqrt{N} + \sqrt{N+1}\right)$. Moreover, it can be lower bounded as

$$\frac{1}{2}\|\rho_1 - \rho_2\|_1 \geq \frac{1}{200}\max\left(\min\left(1, \frac{\|\boldsymbol{m}_1 - \boldsymbol{m}_2\|_2}{\sqrt{h(N,n)}}\right), \min\left(1, \frac{\|V_1 - V_2\|_2}{h(N,n)}\right)\right), \tag{2.244}$$

where $h(N,n) \coloneqq 4N + 2n + 1$.

Theorem 50 is not only a technical result of independent interest for the field of Gaussian quantum information, but also it answers the fundamental question: *if we approximate the first moment and covariance matrix of an unknown Gaussian state with precision ε, what is the resulting trace distance error on the state?* Theorem 50 establishes that if we make an error $O(\varepsilon)$ in approximating the first moment and the covariance matrix, then the trace-distance error that we make in approximating the unknown Gaussian state is *at most* $O(\sqrt{\varepsilon})$ and *at least* $O(\varepsilon)$.

The upper bound presented in Theorem 50 allows us to analyse the sample complexity of tomography of Gaussian states. In particular, we prove that tomography of (energy-constrained) Gaussian states is efficient, as there exists a tomography algorithm whose sample and time complexity scales polynomially in the number of modes. Notably, our result demonstrates that Gaussian states can be *efficiently* learned by estimating the first moment and the covariance matrix, a result that has been previously assumed but never rigorously proved in the literature.

Theorem 51 (Upper bound on the sample complexity of tomography of Gaussian states - informal version)**.** *A number* $O\left(\frac{n^7 N_{phot}^4}{\varepsilon^4}\right)$ *of state copies are sufficient for quantum state tomography of* n*-mode Gaussian states* ρ *satisfying the energy constraint* $\mathrm{Tr}[\rho \hat{N}_n] \leq nN_{phot}$*, where* ε *is the accuracy in trace distance.*

Notably, we can improve the trace distance upper bound in Theorem 50 if we assume one of the states, say ρ_2, to be a pure Gaussian state (interestingly, such an improved bound holds even if ρ_1 is not Gaussian).

Theorem 52 (Upper bound on the trace distance between a pure Gaussian state and a possibly non-Gaussian mixed state)**.** *Let* $\psi := |\psi\rangle\langle\psi|$ *be a pure* n*-mode Gaussian state with first moment* $\boldsymbol{m}(\psi)$ *and second moment* $V(\psi)$*. Let* ρ *be an* n*-mode (possibly non-Gaussian) state with first moment* $\boldsymbol{m}(\rho)$ *and second moment* $V(\rho)$*. Assume that* ρ *and* ψ *satisfy the energy constraint* $\mathrm{Tr}[\rho \hat{N}_n] \leq N$*. Then*

$$\frac{1}{2} \| \rho - |\psi\rangle\langle\psi| \|_1 \leq \sqrt{N + \frac{n}{2}} \sqrt{2\|\boldsymbol{m}(\rho) - \boldsymbol{m}(\psi)\|_2^2 + \|V(\rho) - V(\psi)\|_\infty}\,. \tag{2.245}$$

The improved bound presented in Theorem 52 allows us to analyse the sample complexity of tomography of pure Gaussian states.

Theorem 53 (Upper bound on the sample complexity of tomography of pure Gaussian states - informal version)**.** *A number* $O\left(\frac{n^5 N_{phot}^4}{\varepsilon^4}\right)$ *of state copies are sufficient for quantum state tomography of* n*-mode pure Gaussian states* ψ *satisfying the energy constraint* $\mathrm{Tr}[\psi \hat{N}_n] \leq nN_{phot}$*, where* ε *is the accuracy in trace distance.*

We have seen that learning unknown Gaussian states is efficient, with sample complexity scaling polynomially in the number of modes. But what if the unknown state is not exactly Gaussian? Is our tomography procedure robust against slight perturbations of the set of Gaussian states? These questions are crucial conceptually, given that experimental imperfections may transform Gaussian states into slightly-perturbed Gaussian states. In this section, we will also prove that *quantum state tomography of slightly-perturbed Gaussian states is efficient*. Technically, by 'slightly-perturbed Gaussian state' we mean that the minimum quantum relative entropy between the state and any Gaussian state is sufficiently small. The latter is a meaningful measure of 'non-Gaussianity', thanks to results from [38, 78–80].

Our tomography algorithm for Gaussian states is feasible to realise in practice, as it only requires the estimation of the first moment and the covariance matrix of the unknown Gaussian state, tasks routinely accomplished in quantum optics laboratories through homodyne detection [7, 55]. We stress that the non-trivial aspect of our tomography algorithm is that it comes with rigorous performance guarantees. That is, if the number of state copies is larger than a critical value (which scales polynomially in the number of modes), then the algorithm outputs a classical description of a Gaussian state which is guaranteed to be a 'good' approximation of the true state in trace distance with high probability.

This section is organised as follows:

- In Subsection 2.7.3, we prove upper and lower bounds on the trace distance between Gaussian states in terms of the norm distance between their first moments and their covariance matrices.
- In Subsection 2.7.3, we analyse the sample complexity of learning the first moment and the covariance matrix of unknown (possibly non-Gaussian) states.
- In Subsection 2.7.3, we put together all the above-mentioned results to analyse the sample complexity of tomography of (energy-constrained) Gaussian states both in the mixed-state

and in the pure-state setting. We also how the second moment of the energy of a Gaussian state can be expressed in terms of its first moment and its covariance matrix, a result that turns out to be useful to obtain rigorous performance guarantees on tomography of energy-constrained Gaussian states. Finally, we show the our tomography procedure is robust under little perturbations from the set of Gaussian states, meaning that learning slightly-perturbed Gaussian states is efficient.

Note: Throughout this section, we sometimes represent energy constraints using the energy operator $\hat{E}_n$, such as $\mathrm{Tr}[\rho\hat{E}_n] \leq nE$, rather than the total photon number operator $\hat{N}_n$, like $\mathrm{Tr}[\rho\hat{N}_n] \leq nN_{\mathrm{phot}}$. This simplifies the analysis without altering the sample complexity scaling, because of the identity $\hat{E}_n = \hat{N}_n + \frac{n}{2}\mathbb{1}$ that allows us to identify $E = N_{\mathrm{phot}} + \frac{1}{2}$. For example, the upper bound on the sample-complexity of Gaussian-state tomography in Theorem 51, given by $O\left(\frac{n^7 N_{\mathrm{phot}}^4}{\varepsilon^4}\right)$, can equivalently be expressed as $O\left(\frac{n^7 E^4}{\varepsilon^4}\right)$.

Bounds on the trace distance between Gaussian states

Possessing knowledge of the first moment and covariance matrix of a Gaussian state is sufficient to determine the state itself. However, when dealing with a finite number of copies of an unknown Gaussian state ρ, it is possible to obtain only estimates $\tilde{\mathbf{m}}$ and $\tilde{V}$ of its first moment $\mathbf{m}(\rho)$ and of its covariance matrix $V(\rho)$. Consequently, the resulting Gaussian state $\tilde{\rho}$ with first moment $\tilde{\mathbf{m}}$ and covariance matrix $\tilde{V}$ constitutes only an approximation of the true, unknown Gaussian state ρ. A natural question arises: What is the error incurred in the approximation $\tilde{\rho} \approx \rho$ in terms of the errors incurred in the approximations $\tilde{\mathbf{m}} \approx \mathbf{m}(\rho)$ and $V(\rho) \approx \tilde{V}$? The existing literature lacked such an error estimate, despite it being a natural question in Gaussian quantum information theory. In this section, we address this gap. To quantify the error incurred in the approximation $\tilde{\rho} \approx \rho$, we employ the trace distance $\frac{1}{2}\|\tilde{\rho} - \rho\|_1$ as it is the most meaningful notion of distance between quantum states, given its operational meaning [9, 10]. Specifically:

- In Subsubsection 2.7.3, we derive an upper bound on the trace distance between two arbitrary (possibly mixed) Gaussian state in terms of the norm distance between their first moment and their covariance matrices. More explicitly, by using the notation introduced above, we find an upper bound on $\frac{1}{2}\|\tilde{\rho} - \rho\|_1$ in terms of $\|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2$ and $\|\tilde{V} - V(\rho)\|_1$.
- In Subsubsection 2.7.3, we find an improved upper bound on the trace distance between a pure Gaussian state and a possibly-mixed possibly-non-Gaussian state.
- In Subsubsection 2.7.3, we find a lower bound on the trace distance between two arbitrary (possibly mixed) Gaussian state. One may apply this lower bound, together with ε-net techniques, to obtain a lower bound on the sample complexity on tomography of Gaussian states. However, we do not conduct such an analysis in this work since it leads to a too weak lower bound on the sample complexity.

Upper bound in the general (possibly mixed-state) setting This subsubsection is devoted to the proof of the following theorem, which is one of our main technical result.

Theorem 54. *Let ρ, σ be n-mode Gaussian states with mean photon number bounded by N, that is,*

$$\begin{aligned} \mathrm{Tr}\left[\rho\,\hat{N}_n\right] &\leq N\,, \\ \mathrm{Tr}\left[\sigma\,\hat{N}_n\right] &\leq N\,. \end{aligned} \tag{2.246}$$

Then, the trace distance between ρ and σ is upper bounded by

$$\frac{1}{2}\,\|\rho - \sigma\|_1 \leq f(N)\left(\|\boldsymbol{m}(\rho) - \boldsymbol{m}(\sigma)\|_2 + \sqrt{2}\,\sqrt{\|V(\rho) - V(\sigma)\|_1}\right), \tag{2.247}$$

where $f(x) \coloneqq \frac{1}{\sqrt{2}}\left(\sqrt{x}+\sqrt{x+1}\right)$. *Here,* $\boldsymbol{m}(\rho), V(\rho)$ *denote the first moment and covariance matrix of* ρ*, while* $\boldsymbol{m}(\sigma), V(\sigma)$ *denote the first moment and covariance matrix of* σ*.*

Remark 55. *One might be tempted to believe that a straightforward approach exists for establishing an upper bound akin to the one presented in Theorem 54. Specifically, the Fuchs-van de Graaf inequality [81] implies that the trace distance between the Gaussian states* ρ *and* σ *can be upper bounded in terms of the fidelity* $F(\rho,\sigma)$ *as*

$$\frac{1}{2}\|\rho-\sigma\|_1 \leq \sqrt{1-F(\rho,\sigma)^2}. \tag{2.248}$$

Additionally, a closed formula for the fidelity between Gaussian states is available [52]. However, it is important to note that this closed formula is exceedingly intricate, involving determinants, square roots, and inverses of functions of the covariance matrices $V(\rho), V(\sigma)$*. Consequently, attempting to derive such an upper bound through this method appears too hard, and, indeed, our attempts in this direction were unsuccessful.*

Before proving Theorem 54, let us first provide some useful preliminary results.

For any positive semi-definite matrix $K \in \mathbb{R}^{2n,2n}$, the *Gaussian noise channel* $\mathcal{N}_K$ is the convex combination of displacement transformations $(\cdot) \mapsto \hat{D}_u(\cdot)\hat{D}_u^\dagger$, where $u \in \mathbb{R}^{2n}$ is distributed according to a Gaussian probability distribution with vanishing mean value and covariance matrix equal to K [7, Chapter 5]. Mathematically, $\mathcal{N}_K$ can be written as

$$\mathcal{N}_K(\Theta) \coloneqq \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u)\, \hat{D}_u \Theta \hat{D}_u^\dagger \,, \tag{2.249}$$

where $P_K(u)$ is a Gaussian probability distribution with vanishing mean and covariance matrix equal to K, and $\hat{D}_u \coloneqq e^{iu^\intercal \Omega_n \hat{\mathbf{R}}}$ is the displacement operator (see the preliminaries subsection 3.8.1). Of course, in the case when K is strictly positive, $P_K(u)$ can be written as

$$P_K(u) \coloneqq \frac{e^{-\frac{1}{2}u^\intercal K^{-1}u}}{(2\pi)^n\sqrt{\det K}} \,. \tag{2.250}$$

Instead, when K has some zero eigenvalues, the definition of $P_K(u)$ involves Dirac deltas as follows. Let $K = \sum_{i=1}^{2n} \lambda_i v_i v_i^\intercal$ be a spectral decomposition of K, where $(v_i)_{i=1,\dots,2n}$ are orthonormal eigenvectors, while $\lambda_1,\dots,\lambda_r > 0$ and $\lambda_{r+1},\dots,\lambda_{2n} = 0$ are its eigenvalues, with r being the rank of K. Then, $P_K(u)$ can be written as

$$P_K(u) \coloneqq \frac{e^{-\frac{1}{2}u^\intercal\left(\sum_{i=1}^r \lambda_i^{-1} v_i v_i^\intercal\right)u}}{(2\pi)^{(2n-r)/2}\sqrt{\lambda_1\dots\lambda_r}} \prod_{i=r+1}^{2n} \delta(v_i^\intercal u) \,, \tag{2.251}$$

where $\delta(\cdot)$ denotes the Dirac delta distribution.

It turns out that $\mathcal{N}_K$ leaves the first moments unchanged, and it acts on the covariance matrix of the input state by adding K [7, Chapter 5]. For the sake of completeness, we prove the latter fact in the following lemma.

Lemma 56. ([7, Chapter 5]) *Let* $K \in \mathbb{R}^{2n,2n}$ *be a positive semi-definite matrix. The Gaussian noise channel* $\mathcal{N}_K$*, defined in (3.34), acts on the first moments and covariance matrices as*

$$\boldsymbol{m}(\mathcal{N}_K(\rho)) = \boldsymbol{m}(\rho) \,, \tag{2.252}$$
$$V(\mathcal{N}_K(\rho)) = V(\rho) + K \,, \tag{2.253}$$

for all input states $\rho \in \mathcal{D}(L^2(\mathbb{R}^n))$*.*

Proof. It holds that

$$\begin{aligned}\mathbf{m}(\mathcal{N}_K(\rho)) &= \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u)\, \mathbf{m}\Big(\hat{D}_u \rho \hat{D}_u^\dagger\Big) \\ &= \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u) \; (\mathbf{m}(\rho) + u) \\ &= \mathbf{m}(\rho) + \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u) u \\ &= \mathbf{m}(\rho)\,,\end{aligned} \tag{2.254}$$

where in the second step we have used Eq. (3.28). Moreover, we have that

$$\begin{aligned}V(\mathcal{N}_K(\rho)) &= \mathrm{Tr}[\{\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal\}\mathcal{N}_K(\rho)] - 2\mathbf{m}(\mathcal{N}_K(\rho))\mathbf{m}(\mathcal{N}_K(\rho))^\intercal \\ &= \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u)\, \mathrm{Tr}\left[\rho\{\hat{\mathbf{R}} + u\mathbb{1}, \hat{\mathbf{R}}^\intercal + u^\intercal \mathbb{1}\}\right] - 2\mathbf{m}(\rho)\mathbf{m}(\rho)^\intercal \\ &= V(\rho) + \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u)\, u\mathbf{m}(\rho)^\intercal + \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u)\, \mathbf{m}(\rho)u^\intercal + \int_{\mathbb{R}^{2n}} \mathrm{d}^{2n}u \; P_K(u)\, uu^\intercal \\ &= V(\rho) + K\,,\end{aligned} \tag{2.255}$$

where in the second step we have used Eq. (3.29) and Eq. (2.254), and in the last step we solved a Gaussian integral. □

Lemma 57. ([7, Chapter 5]) *For any $K \geq 0$, the Gaussian noise channel $\mathcal{N}_K$ is actually a Gaussian channel. That is, $\mathcal{N}_K(\rho)$ is a Gaussian state for any Gaussian state ρ.*

Let us recall the definition of *diamond norm*, as it will be useful for the following. The diamond norm of a superoperator $\Delta : \mathcal{D}(\mathcal{H}) \mapsto \mathcal{D}(\mathcal{H})$ is defined as

$$\|\Delta\|_\diamond := \sup_{\rho \in \mathcal{D}(\mathcal{H}\otimes\mathcal{H}_C)} \|\Delta \otimes \mathrm{Id}_C(\rho)\|_1 \;, \tag{2.256}$$

where the sup is taken also over the choice of the ancilla system $\mathcal{H}_C$. In the infinite-dimensional scenario the topology induced by the diamond norm is often too strong (e.g., see Ref. [82, Proposition 1]) and thus it is customary to define the *energy-constrained diamond norm*. Given $N > 0$, given an n-mode system $\mathcal{H}_S = L^2(\mathbb{R}^n)$, the energy-constrained diamond norm of a superoperator $\Delta : \mathcal{D}(\mathcal{H}_S) \mapsto \mathcal{D}(\mathcal{H}_S)$ is defined as [30, 82, 83]

$$\|\Delta\|_{\diamond N} := \sup_{\rho \in \mathcal{D}(\mathcal{H}_S\otimes\mathcal{H}_C) \,:\, \mathrm{Tr}\left[\rho\, \hat{N}_n \otimes \mathbb{1}_C\right] \leq N} \|(\Delta \otimes I_C)\rho\|_1 \;, \tag{2.257}$$

where the sup is taken also over the choice of the ancilla system $\mathcal{H}_C$, and $\hat{N}_n$ is the n-mode photon number operator on $\mathcal{H}_S$. The supremum in the definition of diamond norm in (2.256), as well as the one in (2.257), is achieved by taking the ancilla system $\mathcal{H}_C$ to be equal to the input system $\mathcal{H}$.

Lemma 58. ([53, Eq. (3)]) *Let $u \in \mathbb{R}^{2n}$ and let $\mathcal{D}_u(\cdot) := \hat{D}_u(\cdot)\hat{D}_u^\dagger$ be the displacement channel. Then, for all $N > 0$ the energy-constrained diamond norm of the difference between $\mathcal{D}_u$ and the identity map* Id *can be upper bounded as*

$$\frac{1}{2}\|\mathcal{D}_u - \mathrm{Id}\,\|_{\diamond N} \leq \sin\left(\min\left\{\|u\|_2 f(N),\, \frac{\pi}{2}\right\}\right) \;, \tag{2.258}$$

where $f(N) := \frac{1}{\sqrt{2}}\left(\sqrt{N} + \sqrt{N+1}\right)$.

Lemma 59 (Diamond norm bound)**.** *Let $K \in \mathbb{R}^{2n,2n}$ be a positive semi-definite matrix. Then, for all $N \geq 0$ the energy-constrained diamond norm of the difference between the Gaussian noise channel $\mathcal{N}_K$ and the identity map* Id *can be upper bounded as*

$$\frac{1}{2}\,\|\mathcal{N}_K - \mathrm{Id}\|_{\diamond N} \leq f(N)\sqrt{\mathrm{Tr}\, K}\,, \tag{2.259}$$

where $f(N) := \frac{1}{\sqrt{2}}\left(\sqrt{N} + \sqrt{N+1}\right)$.

Proof. It holds that

$$\begin{aligned}\frac{1}{2}\left\|\mathcal{N}_K-\mathrm{Id}\right\|_{\diamond N} &\overset{\text{(i)}}{\leq} \int \mathrm{d}^{2n}u\; P_K(u)\; \frac{1}{2}\left\|\mathcal{D}_u-\mathrm{Id}\right\|_{\diamond N}\\ &\overset{\text{(ii)}}{\leq} \int \mathrm{d}^{2n}u\; P_K(u)\; \sin\left(\min\left\{\|u\|_2 f(N),\, \frac{\pi}{2}\right\}\right)\\ &\leq f(N)\int \mathrm{d}^{2n}u\; P_K(u)\; \sqrt{u^\intercal u}\\ &\overset{\text{(iii)}}{\leq} f(N)\left(\int \mathrm{d}^{2n}u\; P_K(u)\; u^\intercal u\right)^{1/2}\\ &= f(N)\left(\mathrm{Tr}\, K\right)^{1/2}.\end{aligned} \tag{2.260}$$

Here, in (i), we have used the triangle inequality of the energy-constrained diamond norm and we have introduced the displacement channel $\mathcal{D}_u(\cdot) \coloneqq \hat{D}_u(\cdot)\hat{D}_u^\dagger$. In (ii), we have exploited Lemma 58. In (iii), we have used the concavity of the square root and Jensen inequality. □

Lemma 60 (Expectation value of energy operator)**.** *Let ρ be an n-mode state. The expectation value of the energy operator $\hat{E}_n \coloneqq \frac{\hat{\mathbf{R}}^\intercal\hat{\mathbf{R}}}{2}$ can be expressed in terms of the covariance matrix and the first moment as*

$$\mathrm{Tr}\left[\rho\hat{E}_n\right] = \frac{\mathrm{Tr}V(\rho)}{4} + \frac{\|\boldsymbol{m}(\rho)\|_2^2}{2}. \tag{2.261}$$

In particular, in terms of the photon number operator $\hat{N}_n \coloneqq \sum_{i=1}^n a_i^\dagger a_i$, it holds that

$$\mathrm{Tr}\left[\rho\,\hat{N}_n\right] = \frac{\mathrm{Tr}[V(\rho)-\mathbb{1}]}{4} + \frac{\|\boldsymbol{m}(\rho)\|_2^2}{2}. \tag{2.262}$$

Proof. By definition of covariance matrix, it holds that

$$\begin{aligned}\mathrm{Tr}V(\rho) &= \sum_{i=1}^{2n} V_{i,i}(\rho)\\ &= 2\sum_{i=1}^{2n}\left[\mathrm{Tr}[\hat{R}_i^2\rho] - m_i(\rho)^2\right]\\ &= 4\,\mathrm{Tr}[\rho\hat{E}_n] - 2\|\mathbf{m}(\rho)\|_2^2.\end{aligned} \tag{2.263}$$

Moreover, (2.262) follows by exploiting $\hat{E}_n = \hat{N}_n + \frac{n}{2}\hat{\mathbb{1}}$. □

We are now ready to prove Theorem 54.

Proof of Theorem 54. Let us first prove the upper bound in (2.247) under the assumption that the Gaussian states have the same first moments. Without loss of generality, we can clearly assume that their first moments vanish, as the trace distance is invariant under (displacement) unitary transformations. Given two Gaussian states ρ and σ with zero first moment, let V be the covariance matrix of ρ, and let W be the covariance matrix of σ. In addition, set

$$T \coloneqq \frac{1}{2}\left(V+W+|V-W|\right), \tag{2.264}$$

where we recall that $|V-W| \coloneqq \sqrt{(V-W)^\dagger(V-W)}$. Since $T-V \geq 0$ and $T-W \geq 0$, we can consider the Gaussian noise channels $\mathcal{N}_{T-V}$ and $\mathcal{N}_{T-W}$, as defined in (3.34). Since ρ and σ are Gaussian states and since $\mathcal{N}_{T-V}$ and $\mathcal{N}_{T-W}$ are Gaussian channels (thanks to Lemma 57), it follows that $\mathcal{N}_{T-V}(\rho)$ and $\mathcal{N}_{T-W}(\sigma)$ are Gaussian states. Moreover, by exploiting Lemma 56, both the covariance matrix of $\mathcal{N}_{T-V}(\rho)$ and that of $\mathcal{N}_{T-W}(\sigma)$ are equal to T. Additionally, both the first moment of $\mathcal{N}_{T-V}(\rho)$ and that of $\mathcal{N}_{T-W}(\sigma)$ are equal to the zero vector. Hence, since

the Gaussian states $\mathcal{N}_{T-V}(\rho)$ and $\mathcal{N}_{T-W}(\sigma)$ have identical first moments and covariance, they are actually the same state: $\mathcal{N}_{T-V}(\rho) = \mathcal{N}_{T-W}(\sigma)$. Then, we can write

$$\begin{aligned}
\frac{1}{2}\left\|\rho-\sigma\right\|_1 &\leq \frac{1}{2}\left\|\rho-\mathcal{N}_{T-V}(\rho)+\mathcal{N}_{T-W}(\sigma)-\sigma\right\|_1 \\
&\leq \frac{1}{2}\left\|\rho-\mathcal{N}_{T-V}(\rho)\right\|_1+\frac{1}{2}\left\|\mathcal{N}_{T-W}(\sigma)-\sigma\right\|_1 \\
&= \frac{1}{2}\left\|(\mathcal{N}_{T-V}-\mathrm{Id})(\rho)\right\|_1+\frac{1}{2}\left\|(\mathcal{N}_{T-W}-\mathrm{Id})(\sigma)\right\|_1 \\
&\leq \frac{1}{2}\left\|\mathcal{N}_{T-V}-\mathrm{Id}\right\|_{\diamond N}+\frac{1}{2}\left\|\mathcal{N}_{T-W}-\mathrm{Id}\right\|_{\diamond N} \\
&\overset{\text{(i)}}{\leq} f(N)\left(\sqrt{\mathrm{Tr}[T-V]}+\sqrt{\mathrm{Tr}[T-W]}\right) \\
&\overset{\text{(ii)}}{\leq} 2\,f(N)\sqrt{\frac{1}{2}\,\mathrm{Tr}[T-V]+\frac{1}{2}\,\mathrm{Tr}[T-W]} \\
&= \sqrt{2}\,f(N)\sqrt{\mathrm{Tr}\left|V-W\right|} \\
&= \sqrt{2}\,f(N)\,\sqrt{\left\|V-W\right\|_1}\,,
\end{aligned} \tag{2.265}$$

where in (i) we have used Lemma 59, and in (ii) we have used the concavity of the square root. This proves the claim under the assumption that the Gaussian states ρ and σ have zero first moment.

For the general case, we can denote by $\delta \coloneqq \mathbf{m}(\rho) - \mathbf{m}(\sigma)$ the difference between the first moments of ρ and σ. Then

$$\begin{aligned}
\frac{1}{2}\left\|\rho-\sigma\right\|_1 &\overset{\text{(iii)}}{=} \frac{1}{2}\left\|\hat{D}_{-\mathbf{m}(\sigma)}\rho\hat{D}^\dagger_{-\mathbf{m}(\sigma)}-\hat{D}_{-\mathbf{m}(\sigma)}\sigma\hat{D}^\dagger_{-\mathbf{m}(\sigma)}\right\|_1 \\
&= \frac{1}{2}\left\|\hat{D}_\delta\,\hat{D}_{-\mathbf{m}(\rho)}\rho\hat{D}^\dagger_{-\mathbf{m}(\rho)}\,\hat{D}^\dagger_\delta-\hat{D}_{-\mathbf{m}(\sigma)}\sigma\hat{D}^\dagger_{-\mathbf{m}(\sigma)}\right\|_1 \\
&\leq \frac{1}{2}\left\|\hat{D}_\delta\,\hat{D}_{-\mathbf{m}(\rho)}\rho\hat{D}^\dagger_{-\mathbf{m}(\rho)}\,\hat{D}^\dagger_\delta-\hat{D}_{-\mathbf{m}(\rho)}\rho\hat{D}^\dagger_{-\mathbf{m}(\rho)}\right\|_1+\frac{1}{2}\left\|\hat{D}_{-\mathbf{m}(\rho)}\rho\hat{D}^\dagger_{-\mathbf{m}(\rho)}-\hat{D}_{-\mathbf{m}(\sigma)}\sigma\hat{D}^\dagger_{-\mathbf{m}(\sigma)}\right\|_1 \\
&\overset{\text{(iv)}}{\leq} \frac{1}{2}\left\|\mathcal{D}_\delta-\mathrm{Id}\right\|_{\diamond N}+\frac{1}{2}\left\|\hat{D}_{-\mathbf{m}(\rho)}\rho\hat{D}^\dagger_{-\mathbf{m}(\rho)}-\hat{D}_{-\mathbf{m}(\sigma)}\sigma\hat{D}^\dagger_{-\mathbf{m}(\sigma)}\right\|_1 \\
&\overset{\text{(v)}}{\leq} \frac{1}{2}\left\|\mathcal{D}_\delta-\mathrm{Id}\right\|_{\diamond N}+\sqrt{2}\,f(N)\,\sqrt{\left\|V(\rho)-V(\sigma)\right\|_1} \\
&\overset{\text{(vi)}}{\leq} f(N)\left(\left\|\delta\right\|_2+\sqrt{2}\,\sqrt{\left\|V(\rho)-V(\sigma)\right\|_1}\right).
\end{aligned} \tag{2.266}$$

Here, in (iii), we have used the unitary invariance of the trace norm. In (iv), we exploited the definition of energy-constrained diamond norm in Eq. (2.257), together with the fact that

$$\mathrm{Tr}\left[\hat{N}_n\hat{D}_{-\mathbf{m}(\rho)}\rho\hat{D}^\dagger_{-\mathbf{m}(\rho)}\right] = \frac{\mathrm{Tr}[V(\rho)-\mathbb{1}]}{4} \leq \frac{\mathrm{Tr}[V(\rho)-\mathbb{1}]}{4}+\frac{\left\|\mathbf{m}(\rho)\right\|_2^2}{2} = \mathrm{Tr}\left[\hat{N}_n\rho\right] \leq N\,, \tag{2.267}$$

where we exploited Lemma 60. In (v), we leveraged (2.265), as the two Gaussian states $\hat{D}_{-\mathbf{m}(\rho)}\rho\hat{D}^\dagger_{-\mathbf{m}(\rho)}$ and $\hat{D}_{-\mathbf{m}(\sigma)}\sigma\hat{D}^\dagger_{-\mathbf{m}(\sigma)}$ have zero first moment. Finally, in (vi) we exploited Lemma 58, together with the inequality $\sin(x) \leq x$ for all $x \geq 0$. □

Upper bound in the pure-state setting Theorem 54 provides an upper bound on the trace distance between two Gaussian states ρ and σ, which are possibly mixed. We can obtain a tighter bound if we assume that one of the two states is pure, as proved in the following theorem.

Theorem 61. *Let $\psi := |\psi\rangle\langle\psi|$ be a pure n-mode Gaussian state with first moment $\boldsymbol{m}(\psi)$ and second moment $V(\psi)$. Let ρ be an n-mode possibly non-Gaussian state with first moment $\boldsymbol{m}(\rho)$ and second moment $V(\rho)$. Then*

$$\| \rho - |\psi\rangle\langle\psi| \|_1 \leq \sqrt{\operatorname{Tr} V(\psi)} \sqrt{\|V(\rho) - V(\psi)\|_\infty + 2\|\boldsymbol{m}(\rho) - \boldsymbol{m}(\psi)\|_2^2} \tag{2.268}$$

In particular, if the mean energy of ψ is upper bounded by a constant E, that is,

$$\operatorname{Tr}\left[\rho \hat{E}_n\right] \leq E \tag{2.269}$$

where $\hat{E}_n := \frac{\hat{\mathbf{R}}^\intercal \hat{\mathbf{R}}}{2}$, then

$$\frac{1}{2} \| \rho - |\psi\rangle\langle\psi| \|_1 \leq \sqrt{E} \sqrt{\|V(\rho) - V(\psi)\|_\infty + 2\|\boldsymbol{m}(\rho) - \boldsymbol{m}(\psi)\|_2^2} \tag{2.270}$$

Before proving Theorem 61, let us show some useful lemmas.

Lemma 62. *Let ψ be an n-mode pure Gaussian state. The inverse of the covariance matrix reads*

$$V(\psi)^{-1} = \Omega_n V(\psi) \Omega_n^\intercal \,, \tag{2.271}$$

where Ω_n is the symplectic form defined in (2.64).

Proof. The covariance matrix $V(\psi)$ of the pure Gaussian state ψ can be written as $V(\psi) = SS^\intercal$ with $S \in \mathrm{Sp}(2n)$ being a symplectic matrix (due to (2.95)). Consequently, it holds that

$$\Omega_n^\intercal V(\psi)^{-1} \Omega_n = -\Omega_n V(\psi)^{-1} \Omega_n = -\Omega_n (S^\intercal)^{-1} S^{-1} \Omega_n = -S \Omega_n^2 S^\intercal = SS^\intercal = V(\psi) \,, \tag{2.272}$$

where we have used that $S^{-1} \in \mathrm{Sp}(2n)$. □

Lemma 63. *For any symplectic $S \in \mathrm{Sp}(2n)$ and for any unitarily invariant norm $\|\cdot\|$, it holds that*

$$\|S\| = \|S^{-1}\| \,. \tag{2.273}$$

Proof. By exploiting that $S^{-1} \in \mathrm{Sp}(2n)$, we have that

$$S^{-1} = -S^{-1} \Omega_n \Omega_n = -\Omega_n S^\intercal \Omega_n = \Omega_n S^\intercal \Omega_n^\intercal \,. \tag{2.274}$$

In particular, since Ω_n is orthogonal, it holds that $\|S^{-1}\| = \|S^\intercal\| = \|S\|$. □

We are now ready to prove Theorem 61.

Proof of Theorem 61. Since ψ is a pure Gaussian state, the symplectic eigenvalues of its covariance matrix $V(\psi)$ are all equal to 1. Consequently, there exists a symplectic matrix S such that

$$\begin{aligned} V(\psi) &= SS^\intercal \,, \\ |\psi\rangle &= \hat{D}_{\mathbf{m}(\psi)} U_S |0\rangle \,, \end{aligned} \tag{2.275}$$

where we have denoted as $|0\rangle$ the n-mode vacuum state vector. Hence, it holds that

$$\langle\psi| \rho |\psi\rangle = \langle 0| U_S^\dagger \hat{D}_{\mathbf{m}(\psi)}^\dagger \rho \hat{D}_{\mathbf{m}(\psi)} U_S |0\rangle = \langle 0| \omega |0\rangle \,, \tag{2.276}$$

where we have introduced the state

$$\omega := U_S^\dagger \hat{D}_{\mathbf{m}(\psi)}^\dagger \rho \hat{D}_{\mathbf{m}(\psi)} U_S \,. \tag{2.277}$$

By exploiting the operator inequality

$$|0\rangle\langle 0| \geq \mathbb{1} - \sum_{i=1}^{n} a_i^\dagger a_i \,, \tag{2.278}$$

we find that

$$\begin{aligned} \langle\psi|\,\rho\,|\psi\rangle &= \langle 0|\,\omega\,|0\rangle \\ &= \mathrm{Tr}[\omega\,|0\rangle\langle 0|] \\ &\geq 1 - \mathrm{Tr}\left[\omega \sum_{i=1}^{n} a_i^\dagger a_i\right] \\ &= 1 - \frac{\mathrm{Tr}[V(\omega) - \mathbb{1}]}{4} - \frac{\|\mathbf{m}(\omega)\|_2^2}{2} \,, \end{aligned} \tag{2.279}$$

where we also used Lemma 60. Note that the first moment and covariance matrix of ω are given by

$$\begin{aligned} \mathbf{m}(\omega) &= S^{-1}\left(\mathbf{m}(\rho) - \mathbf{m}(\psi)\right) \,, \\ V(\omega) &= S^{-1} V(\rho) (S^{-1})^\intercal \,. \end{aligned} \tag{2.280}$$

Consequently, it holds that

$$\begin{aligned} |\mathrm{Tr}[V(\omega) - \mathbb{1}]| &= \left|\mathrm{Tr}[S^{-1}V(\rho)(S^{-1})^\intercal - \mathbb{1}]\right| \\ &= \left|\mathrm{Tr}[S^{-1}(V(\rho) - V(\psi))(S^{-1})^\intercal]\right| \\ &\overset{\text{(i)}}{\leq} \|(S^{-1})^\intercal S^{-1}\|_1 \, \|V(\rho) - V(\psi)\|_\infty \\ &\overset{\text{(ii)}}{=} \mathrm{Tr}\left[(S^{-1})^\intercal S^{-1}\right] \|V(\rho) - V(\psi)\|_\infty \\ &= \mathrm{Tr}[V(\psi)^{-1}] \, \|V(\rho) - V(\psi)\|_\infty \\ &\overset{\text{(iii)}}{=} \mathrm{Tr}\, V(\psi) \, \|V(\rho) - V(\psi)\|_\infty \,. \end{aligned} \tag{2.281}$$

Here, in (i), we have used Hölder inequality $|\,\mathrm{Tr}[AB]| \leq \|A\|_1 \|B\|_\infty$. In (ii), we have exploited that $(S^{-1})^\intercal S^{-1}$ is positive semi-definite. Moreover, in (iii) we utilised Lemma 62, which asserts that the inverse of the covariance matrix $V(\psi)$ of the pure Gaussian state ψ can be written as $V(\psi)^{-1} = -\Omega_n V(\psi) \Omega_n$, and, in particular, this implies $\mathrm{Tr}\, V(\psi) = \mathrm{Tr}[V(\psi)^{-1}]$. In addition, it holds that

$$\begin{aligned} \|\mathbf{m}(\omega)\|_2 &= \|S^{-1}\left(\mathbf{m}(\rho) - \mathbf{m}(\psi)\right)\|_2 \\ &\leq \|S^{-1}\|_\infty \, \|\mathbf{m}(\rho) - \mathbf{m}(\psi)\|_2 \\ &\overset{\text{(iv)}}{=} \|S\|_\infty \, \|\mathbf{m}(\rho) - \mathbf{m}(\psi)\|_2 \\ &\leq \sqrt{\mathrm{Tr}[SS^\intercal]} \, \|\mathbf{m}(\rho) - \mathbf{m}(\psi)\|_2 \\ &= \sqrt{\mathrm{Tr}\, V(\psi)} \, \|\mathbf{m}(\rho) - \mathbf{m}(\psi)\|_2 \,, \end{aligned} \tag{2.282}$$

where in (iv) we have exploited Lemma 63. Consequently, (2.279) implies that

$$\langle\psi|\,\rho\,|\psi\rangle \geq 1 - \frac{\mathrm{Tr}\, V(\psi)}{4}\left(\|V(\rho) - V(\psi)\|_\infty + 2\|\mathbf{m}(\rho) - \mathbf{m}(\psi)\|_2^2\right) \tag{2.283}$$

Hence, Fuchs-van de Graaf inequality implies that

$$\begin{aligned} \|\rho - \psi\|_1 &\leq 2\sqrt{1 - \langle\psi|\,\rho\,|\psi\rangle} \\ &\leq \sqrt{\mathrm{Tr}\, V(\psi)}\sqrt{\|V(\rho) - V(\psi)\|_\infty + 2\|\mathbf{m}(\rho) - \mathbf{m}(\psi)\|_2^2} \,. \end{aligned} \tag{2.284}$$

In particular, by exploiting Lemma 60 and the assumption that the mean energy of ψ is upper bounded by E, we conclude that

$$\mathrm{Tr}\, V(\psi) = 4\,\mathrm{Tr}[\hat{E}_n \psi] - 2\|\mathbf{m}(\psi)\|_2^2 \leq 4E \,. \tag{2.285}$$

□

Lower bound In this subsection we obtain a lower bound on the trace distance between Gaussian states in terms of the norm distance between their covariance matrices and their first moments.

Theorem 64. *Let ρ_1 and ρ_2 be n-mode Gaussian states with mean energy upper bounded by E, i.e.,*

$$\begin{aligned} \operatorname{Tr}\left[\rho_1 \hat{E}_n\right] &\leq E\,, \\ \operatorname{Tr}\left[\rho_2 \hat{E}_n\right] &\leq E\,. \end{aligned} \tag{2.286}$$

Then, the trace distance between ρ_1 and ρ_2 can be lower bounded in terms of the norm distance between their first moments as

$$\frac{1}{2}\|\rho_1 - \rho_2\|_1 \geq \frac{1}{200}\min\left\{1, \frac{\|\boldsymbol{m}(\rho_1) - \boldsymbol{m}(\rho_2)\|_2}{\sqrt{4E+1}}\right\}, \tag{2.287}$$

and in terms of the norm distance between their covariance matrices as

$$\frac{1}{2}\|\rho_1 - \rho_2\|_1 \geq \frac{1}{200}\min\left\{1, \frac{\|V(\rho_2) - V(\rho_1)\|_2}{4E+1}\right\}. \tag{2.288}$$

Proof. Recall that the Husimi function $Q_\rho : \mathbb{R}^{2n} \longmapsto \mathbf{R}$ of an n-mode quantum state ρ is defined as [7]

$$Q_\rho(\mathbf{r}) \coloneqq \frac{1}{(2\pi)^n}\langle \mathbf{r}|\,\rho\,|\mathbf{r}\rangle\,, \tag{2.289}$$

where $|\mathbf{r}\rangle \coloneqq \hat{D}_{\mathbf{r}}\,|0\rangle$ is a coherent state. The Husimi function Q_ρ is the probability distribution of the outcome $\mathbf{r}$ of the heterodyne measurement performed on ρ [7]. Hence, by exploiting the monotonicity of the trace norm under quantum channels, the trace distance between ρ_1 and ρ_2 can be lower bounded in terms of the TV distance between their Husimi functions as

$$\frac{1}{2}\|\rho_1 - \rho_2\|_1 \geq \frac{1}{2}\int_{\mathbf{r}\in\mathbb{R}^{2n}} \mathrm{d}^{2n}\mathbf{r}\,|Q_{\rho_1}(\mathbf{r}) - Q_{\rho_2}(\mathbf{r})| = d_{\mathrm{TV}}\left(Q_{\rho_1}, Q_{\rho_2}\right), \tag{2.290}$$

where in the last equality we have used the definition of TV distance. The Husimi function of a Gaussian state ρ is a Gaussian probability distribution with first moment $\mathbf{m}(\rho)$ and covariance matrix $\frac{V(\rho)+\mathbb{1}}{2}$ (see (3.229) in the preliminaries section):

$$Q_\rho(\mathbf{r}) = \mathcal{N}\left[\mathbf{m}(\rho), \frac{V(\rho)+\mathbb{1}}{2}\right](\mathbf{r})\,, \tag{2.291}$$

where $\mathcal{N}$ denotes a Gaussian probability distribution defined as follows. Given a real vector $\mathbf{m}$ and a positive matrix V, the Gaussian probability distribution $\mathcal{N}[\mathbf{m}, V]$ with first moment $\mathbf{m}$ and covariance matrix V is defined as

$$\mathcal{N}[\mathbf{m}, V](\mathbf{r}) \coloneqq \frac{e^{-\frac{1}{2}(\mathbf{r}-\mathbf{m})^\intercal V^{-1}(\mathbf{r}-\mathbf{m})}}{(2\pi)^n\sqrt{\det V}}\,. \tag{2.292}$$

Consequently, (2.290) implies that

$$\frac{1}{2}\|\rho_1 - \rho_2\|_1 \geq d_{\mathrm{TV}}\left(\mathcal{N}\left[\mathbf{m}(\rho_1), \frac{V(\rho_1)+\mathbb{1}}{2}\right], \mathcal{N}\left[\mathbf{m}(\rho_2), \frac{V(\rho_2)+\mathbb{1}}{2}\right]\right). \tag{2.293}$$

Now, let us leverage recent results about tight bounds on the TV distance between two arbitrary Gaussian probability distributions [54]. By exploiting [54, Theorem 1.1], together with [84, Lemma B.3], the TV distance between two Gaussian probability distributions $\mathcal{N}[\mathbf{m}_1, V_1]$ and $\mathcal{N}[\mathbf{m}_2, V_2]$ can be lower bounded in terms of the norm distance between covariance matrices as

$$d_{\mathrm{TV}}\left(\mathcal{N}[\mathbf{m}_1, V_1], \mathcal{N}[\mathbf{m}_2, V_2]\right) \geq \frac{\min\left\{1, \left\|V_1^{-1/2} V_2 V_1^{-1/2} - \mathbb{1}\right\|_2\right\}}{200}\,. \tag{2.294}$$

Note that we can lower bound the term $\left\|V_1^{-1/2}V_2V_1^{-1/2}-\mathbb{1}\right\|_2$ in terms of $\|V_2-V_1\|_2$ as

$$\begin{aligned}\left\|V_1^{-1/2}V_2V_1^{-1/2}-\mathbb{1}\right\|_2 &= \left\|V_1^{-1/2}\left(V_2-V_1\right)V_1^{-1/2}\right\|_2 \\ &= \sqrt{\mathrm{Tr}\left[V_1^{-1}(V_2-V_1)V_1^{-1}(V_2-V_1)\right]} \\ &\geq \sqrt{\mathrm{Tr}\left[V_1^{-1}(V_2-V_1)\frac{\mathbb{1}}{\|V_1\|_\infty}(V_2-V_1)\right]} \\ &\geq \frac{\sqrt{\mathrm{Tr}[(V_2-V_1)^2]}}{\|V_1\|_\infty} \\ &= \frac{\|V_2-V_1\|_2}{\|V_1\|_\infty}.\end{aligned} \tag{2.295}$$

Consequently, we obtain the following lower bound on the TV distance between two Gaussian probability distributions

$$d_{\mathrm{TV}}\left(\mathcal{N}[\mathbf{m}_1,V_1],\mathcal{N}[\mathbf{m}_2,V_2]\right) \geq \frac{1}{200}\min\left\{1,\frac{\|V_2-V_1\|_2}{\|V_1\|_\infty}\right\}. \tag{2.296}$$

In addition, thanks to Ref. [54, Theorem 1.1], the TV distance between two Gaussian probability distributions can also be lower bounded as

$$d_{\mathrm{TV}}\left(\mathcal{N}[\mathbf{m}_1,V_1],\mathcal{N}[\mathbf{m}_2,V_2]\right) \geq \frac{1}{200}\min\left\{1,\frac{\|\mathbf{m}_1-\mathbf{m}_2\|_2^2}{\sqrt{(\mathbf{m}_1-\mathbf{m}_2)^\intercal V_1\left(\mathbf{m}_1-\mathbf{m}_2\right)}}\right\}, \tag{2.297}$$

which implies that

$$d_{\mathrm{TV}}\left(\mathcal{N}[\mathbf{m}_1,V_1],\mathcal{N}[\mathbf{m}_2,V_2]\right) \geq \frac{1}{200}\min\left\{1,\frac{\|\mathbf{m}_1-\mathbf{m}_2\|_2}{\sqrt{\|V_1\|_\infty}}\right\}. \tag{2.298}$$

Consequently, by using (2.293), together with (2.296) and (2.298), we obtain that

$$\begin{aligned}\frac{1}{2}\|\rho_1-\rho_2\|_1 &\geq \frac{1}{200}\min\left\{1,\frac{\|\mathbf{m}(\rho_1)-\mathbf{m}(\rho_2)\|_2}{\sqrt{\|V(\rho_1)+\mathbb{1}\|_\infty}}\right\}, \\ \frac{1}{2}\|\rho_1-\rho_2\|_1 &\geq \frac{1}{200}\min\left\{1,\frac{\|V(\rho_2)-V(\rho_1)\|_2}{\|V(\rho_1)+\mathbb{1}\|_\infty}\right\}.\end{aligned} \tag{2.299}$$

Hence, since any n-mode state ρ satisfies $\|V(\rho)\|_\infty \leq 4\,\mathrm{Tr}[\rho\hat{E}_n]$, the assumption on the energy constraints in (2.286) implies that

$$\begin{aligned}\frac{1}{2}\|\rho_1-\rho_2\|_1 &\geq \frac{1}{200}\min\left\{1,\frac{\|\mathbf{m}(\rho_1)-\mathbf{m}(\rho_2)\|_2}{\sqrt{4E+1}}\right\}, \\ \frac{1}{2}\|\rho_1-\rho_2\|_1 &\geq \frac{1}{200}\min\left\{1,\frac{\|V(\rho_2)-V(\rho_1)\|_2}{4E+1}\right\}.\end{aligned} \tag{2.300}$$

□

Learning first moments and covariance matrices

In this subsection, we conduct the sample complexity analysis of an algorithm for estimating the first moment and covariance matrix of a (possibly non-Gaussian) quantum state. Recall that the covariance matrix $V(\rho)$ is defined by

$$V(\rho) = \mathrm{Tr}[\{\hat{\mathbf{R}},\hat{\mathbf{R}}^\intercal\}\rho] - 2\mathbf{m}(\rho)\mathbf{m}(\rho)^\intercal, \tag{2.301}$$

where $\mathbf{m}(\rho) := \mathrm{Tr}[\rho\hat{\mathbf{R}}]$ represents the first moment vector, and $\hat{\mathbf{R}} = (\hat{x}_1, \hat{p}_1, \hat{x}_2, \hat{p}_2, \ldots, \hat{x}_n, \hat{p}_n)$ denotes the quadrature vector. Additionally, the covariance matrix of any state must satisfy the uncertainty relation:

$$V(\rho) + i\Omega_n \geq 0. \tag{2.302}$$

A direct method for estimating the covariance matrix involves estimating the first moment and the expectation values of all observables $(\{\hat{R}_i, \hat{R}_j\})_{i,j\in[2n]}$. However, there is a more efficient algorithm, as detailed by Aolita et al. [55], which requires fewer copies of ρ. We adopt this approach and provide a sample complexity analysis that extends beyond what was outlined in Ref. [55].

The algorithm's strategy involves grouping commuting observables whose expectation values form entries of the covariance matrix and performing joint measurements on these groups of commuting observables. This estimation procedure yields a symmetric matrix $\tilde{V} \in \mathbb{R}^{2n\times 2n}$ as an approximation of $V(\rho)$. However, the estimated matrix $\tilde{V}$ may not satisfy the uncertainty relation $\tilde{V} + i\Omega_n \geq 0$, implying that there might not exist a quantum state $\tilde{\rho}$ with covariance matrix $\tilde{V}$. We address this issue by showing that slightly perturbing $\tilde{V}$ suffices to obtain a valid covariance matrix. Ensuring that the estimated matrix is a valid covariance matrix is crucial, for example, when we need to consider a Gaussian state with covariance matrix $\tilde{V}$ (as we will do in Subsection 2.7.3), or when applying the Williamson decomposition to $\tilde{V}$ (e.g., we will need this in Section 2.7.3).

The algorithm's detailed steps are provided in Table 2.2. Notably, this algorithm exclusively relies on homodyne measurements, which are experimentally feasible in photonic platforms. In the forthcoming Theorem 65, we establish the algorithm's correctness and demonstrate that its sample complexity is $O\Big(\log\left(\frac{n^2}{\delta}\right)\frac{n^3E^2}{\varepsilon^2}\Big)$.

Theorem 65. *Let $\varepsilon, \delta \in (0,1)$ and $E \geq 0$. Let ρ be an n-mode quantum state satisfying the second moment constraint $\sqrt{\mathrm{Tr}\big[\hat{E}_n^2\rho\big]} \leq nE$. Then, N copies of ρ, satisfying*

$$\begin{aligned} N &:= (n+3)\left\lceil 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(8n^2E^2+3n)}{\varepsilon^2}\right\rceil \\ &= O\left(\log\left(\frac{n^2}{\delta}\right)\frac{n^3E^2}{\varepsilon^2}\right), \end{aligned} \tag{2.305}$$

are sufficient to build a vector $\tilde{\boldsymbol{m}} \in \mathbb{R}^{2n}$ and a symmetric matrix $\tilde{V}' \in \mathbb{R}^{2n,2n}$ such that

$$\Pr\left(\|\tilde{V}' - V(\rho)\|_\infty \leq \varepsilon \quad \textit{and} \quad \tilde{V}' + i\Omega_n \geq 0 \quad \textit{and} \quad \|\tilde{\boldsymbol{m}} - \boldsymbol{m}(\rho)\|_2 \leq \frac{\varepsilon}{10\sqrt{8En}}\right) \geq 1-\delta\,. \tag{2.306}$$

Before proving Theorem 65, let us show the following lemma.

Lemma 66. *For all $n \in \mathbb{N}$, it holds that*

$$\sum_{j,k=1}^{2n}\{\hat{R}_j, \hat{R}_k\}^2 = 16\hat{E}_n^2 + 6n\hat{\mathbb{1}}\,, \tag{2.307}$$

where $\hat{E}_n := \frac{\hat{\mathbf{R}}^\intercal\hat{\mathbf{R}}}{2}$ is the energy operator.

Table 2.2: Estimation algorithm for the first moment $\mathbf{m}(\rho)$ and covariance matrix $V(\rho)$ of an arbitrary n-mode quantum state ρ satisfying a second moment constraint.

Input: Accuracy ε, failure probability δ, second moment constraint E, N copies of the unknown n-mode quantum state ρ satisfying the moment constraint $\sqrt{\mathrm{Tr}\left[\hat{E}_n^2\rho\right]} \leq nE$, where

$$N \coloneqq (n+3)\left\lceil 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(8n^2E^2+3n)}{\varepsilon^2}\right\rceil = O\left(\log\left(\frac{n^2}{\delta}\right)\frac{n^3E^2}{\varepsilon^2}\right). \tag{2.303}$$

Output: A vector $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$ and a symmetric matrix $\tilde{V}' \in \mathbb{R}^{2n,2n}$ such that

$$\Pr\left(\|\tilde{V}' - V(\rho)\|_\infty \leq \varepsilon \quad \text{and} \quad \|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \leq \frac{\varepsilon}{10\sqrt{8En}}\right) \geq 1-\delta\,. \tag{2.304}$$

1: Set $N' \coloneqq N/(n+3)$.
2: Query N' copies of ρ and, for each, perform a joint measurement of the position observables $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$. Then, construct median-of-means estimators (Lemma 16) of the expectation values of $(\{\hat{x}_i, \hat{x}_j\})_{i\leq j}$ and of $(\hat{x}_i)_{i\in[n]}$.
3: Query N' copies of ρ and, for each, perform a joint measurement of the momentum observables $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n$. Then, construct median-of-means estimators of the expectation values of $(\{\hat{p}_i, \hat{p}_j\})_{i\leq j}$ and of $(\hat{p}_i)_{i\in[n]}$.
4: Query N' copies of ρ and, for each, perform a joint measurement of $\{\hat{x}_1, \hat{p}_1\}, \{\hat{x}_2, \hat{p}_2\}, \ldots, \{\hat{x}_n, \hat{p}_n\}$. Then, construct median-of-means estimators of the expectation value of $(\{\hat{x}_i, \hat{p}_i\})_{i\in[n]}$.
5: **for** $k \leftarrow 1$ **to** n **do**
6: Query N' copies of ρ and, for each, perform a joint measurement of $(\hat{x}_i)_{i\neq k}$ and $\hat{p}_k$. Then, construct median-of-means estimators of the expectation values of $(\{\hat{p}_k, \hat{x}_i\})_{i\neq k}$.
7: **end for**
8: Combine all the aforementioned estimates to form the estimator $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$ for the first moment $\mathbf{m}(\rho)$, and the estimator $\tilde{W} \in \mathbb{R}^{2n,2n}$ for the matrix $\mathrm{Tr}[\rho\{\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal\}]$.
9: Set $\tilde{V} \coloneqq \tilde{W} - 2\tilde{\mathbf{m}}\tilde{\mathbf{m}}^\intercal$.
10: Set $\tilde{V}' \coloneqq \tilde{V} + \frac{\varepsilon}{2}\mathbb{1}$.
11: **if** $\tilde{V}' + i\Omega_n$ is positive semi-definite **then**
12: **return** the estimator $\tilde{\mathbf{m}}$ for the first moment and the estimator $\tilde{V}'$ for the covariance matrix.
13: **else**
14: Declare *failure* and abort.
15: **end if**

Proof. Note that

$$\begin{aligned}
\sum_{j,k=1}^{2n}\{\hat{R}_j, \hat{R}_k\}^2 &= \sum_{j,k=1}^{2n}\left(\hat{R}_j\hat{R}_k + \hat{R}_k\hat{R}_j\right)^2 \\
&= 2\sum_{j,k=1}^{2n}\left(\hat{R}_j\hat{R}_k\hat{R}_j\hat{R}_k + \hat{R}_j\hat{R}_k^2\hat{R}_j\right) \\
&\overset{\text{(i)}}{=} 2\sum_{j,k=1}^{2n}\left(2\hat{R}_j^2\hat{R}_k^2 - 3i\Omega_{jk}\hat{R}_j\hat{R}_k\right) \\
&\overset{\text{(ii)}}{=} 4\sum_{j,k=1}^{2n}\hat{R}_j^2\hat{R}_k^2 + 6n\hat{\mathbb{1}} \\
&= 4\left(\hat{\mathbf{R}}^\intercal\hat{\mathbf{R}}\right)^2 + 6n\hat{\mathbb{1}} \\
&= 16\hat{E}^2 + 6n\hat{\mathbb{1}}\,,
\end{aligned} \tag{2.308}$$

Here, in (i), we have used three times that $\hat{R}_k\hat{R}_j = \hat{R}_j\hat{R}_k - i\Omega_{jk}\hat{\mathbb{1}}$ and, in (ii), we have used that $\sum_{j,k=1}^{2n}\Omega_{jk}\hat{R}_j\hat{R}_k = in\hat{\mathbb{1}}$. □

Proof of Theorem 65. We aim to establish the correctness of the algorithm presented in Table 2.2. In Line 2, we estimate the expectation value of $\{\hat{x}_i, \hat{x}_j\}$ for each $i \leq j \in [n]$ and the expectation value of $\hat{x}_i$ for each $i \in [n]$, amounting to $\frac{n(n+1)}{2} + n$ quantities. In Line 3, we estimate the expectation value of $\{\hat{p}_i, \hat{p}_j\}$ for each $i \leq j \in [n]$ and the expectation value of $\hat{p}_i$ for each $i \in [n]$, which are also $\frac{n(n+1)}{2} + n$ quantities. In Line 4, we estimate the expectation value of $\{\hat{x}_i, \hat{p}_i\}$ for each $i \in [n]$, constituting n quantities. Finally, in all iterations of Line 6, we estimate the expectation value of $\{\hat{p}_k, \hat{x}_i\}$ for each $i \neq k \in [n]$, resulting in $n(n-1)$ quantities. Consequently, across Lines 2–6, we estimate a total of $2n^2 + 3n$ quantities. By combining all these estimates, we can construct an estimator $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$ for the first moment $\mathbf{m}(\rho)$, and an estimator $\tilde{W} \in \mathbb{R}^{2n,2n}$ for the matrix $\mathrm{Tr}[\rho\{\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal\}]$. Therefore, by making use of Lemma 16 together with an union bound, for all $\varepsilon' \in (0,1)$ we can choose

$$N' \geq 68 \log\left(\frac{2(2n^2+3n)}{\delta}\right) \frac{16E^2 + 6/n}{\varepsilon'^2} \tag{2.309}$$

to ensure that with probability at least $1-\delta$ it holds that

$$\left|\tilde{W}_{i,j} - \mathrm{Tr}[\rho\{\hat{R}_i, \hat{R}_j\}]\right| \leq \varepsilon' \sqrt{\frac{\mathrm{Tr}[\rho\{\hat{R}_i, \hat{R}_j\}^2]}{16E^2 + 6/n}} \qquad \forall\, i,j \in [2n]\,, \tag{2.310}$$

and

$$|\tilde{m}_i - m_i(\rho)| \leq \varepsilon' \sqrt{\frac{\mathrm{Tr}[\rho\hat{R}_i^2]}{16E^2 + 6/n}} \qquad \forall\, i \in [2n]\,. \tag{2.311}$$

Here, we have used the fact that the variance of the random variable considered to construct the estimator $\tilde{W}_{i,j}$ can be upper bounded by $\mathrm{Tr}[\rho\{\hat{R}_i, \hat{R}_j\}^2]$ for each $i,j \in [2n]$, and the variance of the random variable considered to construct the estimator $\tilde{m}_i$ can be upper bounded by $\mathrm{Tr}[\rho\hat{R}_i^2]$ for each $i \in [2n]$. Moreover, if (2.310) holds, then we have that

$$\begin{aligned}
\left\|\tilde{W} - \mathrm{Tr}[\rho\{\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal\}]\right\|_\infty &\leq \left\|\tilde{W} - \mathrm{Tr}[\rho\{\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal\}]\right\|_2 \\
&= \sqrt{\sum_{i,j=1}^{2n} \left|\tilde{W}_{i,j} - \mathrm{Tr}[\rho\{\hat{R}_i, \hat{R}_j\}]\right|^2} \\
&\leq \varepsilon' \sqrt{\frac{\sum_{i,j=1}^{2n} \mathrm{Tr}[\rho\{\hat{R}_i, \hat{R}_j\}^2]}{16E^2 + 6/n}} \\
&= n\varepsilon'\,,
\end{aligned} \tag{2.312}$$

where the last equality is a consequence of Lemma 66. In addition, if (2.311) holds, we have that

$$\begin{aligned}
\|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 &\leq \varepsilon' \sqrt{\frac{\sum_{i=1}^{2n} \mathrm{Tr}[\rho\hat{R}_i^2]}{16E^2 + 6/n}} \\
&= \varepsilon' \sqrt{\frac{2\,\mathrm{Tr}[\rho\hat{E}_n]}{16E^2 + 6/n}} \\
&\overset{(i)}{\leq} \varepsilon' \sqrt{\frac{2\sqrt{\mathrm{Tr}[\rho\hat{E}_n^2]}}{16E^2 + 6/n}} \\
&\leq \varepsilon' \sqrt{\frac{2n\sqrt{E^2}}{16E^2 + 6/n}} \\
&\leq \frac{\sqrt{n}\varepsilon'}{\sqrt{8}E^{1/2}}
\end{aligned} \tag{2.313}$$

where (i) is a consequence of the fact that $\mathrm{Tr}[\hat{E}_n\rho] = \mathrm{Tr}[(\hat{E}_n^2)^{\frac{1}{2}}\rho]$ and exploited the concavity of $x \mapsto \sqrt{x}$ for $x > 0$ (see Eq. (2.41)). Furthermore, if (2.311) holds, we have also that

$$\begin{aligned}
\|\tilde{\mathbf{m}}\tilde{\mathbf{m}}^\intercal - \mathbf{m}(\rho)\mathbf{m}(\rho)^\intercal\|_\infty &\leq \|(\tilde{\mathbf{m}} - \mathbf{m}(\rho))\tilde{\mathbf{m}}^\intercal\|_\infty + \|\mathbf{m}(\rho)(\tilde{\mathbf{m}} - \mathbf{m}(\rho))^\intercal\|_\infty \\
&= \|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2\|\tilde{\mathbf{m}}\|_2 + \|\mathbf{m}(\rho)\|_2\|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \\
&\leq \|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \left(\|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 + 2\|\mathbf{m}(\rho)\|_2\right) \\
&\overset{\text{(ii)}}{\leq} \frac{\sqrt{n}\varepsilon'}{\sqrt{8}E^{1/2}}\left(\frac{\sqrt{n}\varepsilon'}{\sqrt{8}E^{1/2}} + 2\|\mathbf{m}(\rho)\|_2\right) \\
&\overset{\text{(iii)}}{\leq} \frac{\sqrt{n}\varepsilon'}{\sqrt{8}E^{1/2}}\left(\frac{\sqrt{n}\varepsilon'}{\sqrt{8}E^{1/2}} + \sqrt{8n}E^{1/2}\right) \\
&= \left(\frac{\varepsilon'}{8E} + 1\right) n\varepsilon' \\
&\overset{\text{(iv)}}{\leq} 2n\varepsilon'
\end{aligned} \tag{2.314}$$

Here, in (ii) we have used (2.313), while in (iii) we have exploited that

$$\|\mathbf{m}(\rho)\|_2 = \sqrt{\sum_{i=1}^{2n} \mathrm{Tr}[\rho\hat{R}_i]^2} \leq \sqrt{\sum_{i=1}^{2n} \mathrm{Tr}[\rho\hat{R}_i^2]} = \sqrt{2\,\mathrm{Tr}[\rho\hat{E}_n]} \leq \sqrt{2\sqrt{\mathrm{Tr}[\rho\hat{E}_n^2]}} \leq \sqrt{2n}E^{1/2}\,, \tag{2.315}$$

where the first inequality and second inequality are a consequence of Eq. (2.41). In (iv) we have used that the second moment satisfies $E \geq \frac{1}{2}$ (due to (2.66)). Therefore, if (2.310) and (2.311) hold, then it holds that

$$\|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \leq \frac{\sqrt{n}\varepsilon'}{\sqrt{8}E^{1/2}} \tag{2.316}$$

and

$$\|\tilde{V} - V(\rho)\|_\infty \leq 5n\varepsilon'\,, \tag{2.317}$$

where we defined $\tilde{V} \coloneqq \tilde{W} - 2\tilde{\mathbf{m}}\tilde{\mathbf{m}}^\intercal$ and we have used triangle inequality. Consequently, by setting $\varepsilon' \coloneqq \frac{\varepsilon}{10n}$, we have that the choice

$$N' \geq 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(8n^2E^2+3n)}{\varepsilon^2}\,, \tag{2.318}$$

allows us to guarantee that

$$\Pr\left(\|\tilde{V} - V(\rho)\|_\infty \leq \frac{\varepsilon}{2} \quad \text{and} \quad \|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \leq \frac{\varepsilon}{10\sqrt{8}E^{1/2}\sqrt{n}}\right) \geq 1 - \delta\,. \tag{2.319}$$

The total number of copies of ρ used in the algorithm in Table 2.2 is $N = (n+3)N'$. Then, the algorithm defines the matrix $\tilde{V}' \coloneqq \tilde{V} + \frac{\varepsilon}{2}\mathbb{1}$. Consequently, in order to conclude the proof it suffices to show that if the condition $\|\tilde{V} - V(\rho)\|_\infty \leq \frac{\varepsilon}{2}$ is satisfied, then $\tilde{V}'$ is a proper covariance matrix satisfying $\|\tilde{V}' - V(\rho)\|_\infty \leq \varepsilon$. The latter condition follows by the triangle inequality. In order to show that $\tilde{V}'$ is a proper covariance matrix, let us observe that $\tilde{V}' + i\Omega_n \geq 0$:

$$\begin{aligned}
\tilde{V}' + i\Omega_n &= \tilde{V} + i\Omega_n + \frac{\varepsilon}{2}\mathbb{1} \\
&\overset{\text{(i)}}{\geq} \tilde{V} - V(\rho) + \frac{\varepsilon}{2}\mathbb{1} \\
&\overset{\text{(ii)}}{\geq} -\|\tilde{V} - V(\rho)\|_\infty\mathbb{1} + \frac{\varepsilon}{2}\mathbb{1} \\
&\geq 0\,,
\end{aligned} \tag{2.320}$$

where in (i) we have used that $V(\rho) + i\Omega_n \geq 0$, while in (ii) we have used that for any operator Θ it holds that $\Theta \geq -\|\Theta\|_\infty\mathbb{1}$. □

The sample complexity for estimating the covariance matrix of the above algorithm is $O\Big(\frac{n^3E^2}{\varepsilon^2}\log\Big(\frac{n^2}{\delta}\Big)\Big)$, outperforming the straightforward method of estimating indipendently the expectation values of all the observables $(\{\hat{R}_i, \hat{R}_j\})_{i,j\in[2n]}$, which leads to a worse sample complexity of $O\Big(\frac{n^4E^2}{\varepsilon^2}\log\Big(\frac{n^2}{\delta}\Big)\Big)$.

Learning Gaussian states

In this subsection, we analyse the algorithm for learning Gaussian states. In Subsubsection 2.7.3, we discuss the relation between the second moment energy constraint of a Gaussian state and its first moment and covariance matrix. Subsequently, in Subsubsection 2.7.3, we detail the algorithm for learning mixed Gaussian states and provide the associated recovery guarantees. Moreover, in Subsubsection 2.7.3, we rigorously show that this algorithm is noise-robust, if the state deviates slightly from the set of Gaussian states, the same tomography algorithm can still be applied. Finally, in Subsubsection 2.7.3, we outline the algorithm for learning pure Gaussian states and prove its recovery guarantees.

Second moment of the energy of a Gaussian state A Gaussian state ρ is uniquely identified by its first moment $\mathbf{m}(\rho)$ and its covariance matrix $V(\rho)$. In the forthcoming lemma we will see how to express the second moment of the energy in terms of these quantities.

Lemma 67. *Let ρ be an n-mode Gaussian state with first moment $\boldsymbol{m}$ and covariance matrix V. The first and the second moment of the energy can be expressed in terms of the covariance matrix and the first moment as*

$$\mathrm{Tr}\Big[\rho\hat{E}_n\Big] = \frac{\mathrm{Tr}V}{4} + \frac{\|\boldsymbol{m}\|_2^2}{2}\,, \tag{2.321}$$

$$\mathrm{Tr}\Big[\rho\hat{E}_n^2\Big] = \Big(\frac{1}{4}\,\mathrm{Tr}\,V + \frac{1}{2}\|\boldsymbol{m}\|_2^2\Big)^2 + \frac{1}{8}\,\mathrm{Tr}[V^2] + \frac{1}{2}\boldsymbol{m}^\intercal V\boldsymbol{m} - \frac{n}{4}\,, \tag{2.322}$$

where $\hat{E}_n \coloneqq \frac{\hat{\mathbf{R}}^\intercal\hat{\mathbf{R}}}{2}$ is the energy operator. In particular, this implies that the second moment of the energy can be upper bounded in terms of the mean energy as

$$\mathrm{Tr}[\rho\hat{E}_n^2] \leq 3(\mathrm{Tr}[\rho\hat{E}_n])^2\,. \tag{2.323}$$

Proof. (2.321) holds because of Lemma 60. In order to prove (2.322), let us employ [85, Lemma 7], which states that: for any $2n\times 2n$ real symmetric matrix X, it holds that

$$\mathrm{Tr}\Big[(\hat{\mathbf{R}}^\intercal X\hat{\mathbf{R}})^2\,\rho(V,\mathbf{m})\Big] = \Big(\frac{1}{2}\,\mathrm{Tr}[VX] + \mathbf{m}^\intercal X\mathbf{m}\Big)^2 + \frac{1}{2}\,\mathrm{Tr}[XVXV] + \frac{1}{2}\,\mathrm{Tr}[\Omega X\Omega X] + 2\mathbf{m}^\intercal XVX\mathbf{m}\,. \tag{2.324}$$

By setting $X = \frac{\mathbb{1}}{2}$, this proves (2.322). Moreover, note that

$$\begin{aligned}
\mathrm{Tr}\Big[\rho\hat{E}_n^2\Big] &= (\mathrm{Tr}[\rho\hat{E}_n])^2 + \frac{1}{8}\,\mathrm{Tr}[V^2] + \frac{1}{2}\mathbf{m}^\intercal V\mathbf{m} - \frac{n}{4} \\
&\leq (\mathrm{Tr}[\rho\hat{E}_n])^2 + \frac{1}{8}\,[\mathrm{Tr}V]^2 + \frac{1}{2}\|\mathbf{m}\|_2^2\,\mathrm{Tr}V \\
&= (\mathrm{Tr}[\rho\hat{E}_n])^2 + \frac{1}{2}\,\mathrm{Tr}[\rho\hat{E}_n]\,\mathrm{Tr}V + \frac{1}{4}\|\mathbf{m}\|_2^2\,\mathrm{Tr}V \\
&= 3(\mathrm{Tr}[\rho\hat{E}_n])^2 - \mathrm{Tr}[\rho\hat{E}_n]\|\mathbf{m}\|_2^2 - \frac{1}{2}\|\mathbf{m}\|_2^4 \\
&\leq 3(\mathrm{Tr}[\rho\hat{E}_n])^2\,,
\end{aligned} \tag{2.325}$$

where we have used that $\mathrm{Tr}\Big[\rho\hat{E}_n\Big] = \frac{\mathrm{Tr}V}{4} + \frac{\|\mathbf{m}\|_2^2}{2}$. □

Table 2.3: Estimation algorithm of an unknown n-mode Gaussian state ρ satisfying the energy constraint $\mathrm{Tr}[\rho\hat{E}_n] \leq nE$.

Input: Accuracy ε, failure probability δ, mean energy per mode upper bound E, N of copies of the unknown n-mode Gaussian state ρ satisfying the energy constraint $\mathrm{Tr}[\rho\hat{E}_n] \leq nE$, where

$$\begin{aligned} N &:= (n+3)\left\lceil 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(24n^2E^2+3n)}{\varepsilon^4}2^{14}E^2n^4\right\rceil \\ &= O\left(\log\left(\frac{n^2}{\delta}\right)\frac{n^7E^4}{\varepsilon^4}\right). \end{aligned} \tag{2.326}$$

Output: With probability less than δ, the output is *failure*. Otherwise, with probability $\geq 1-\delta$, the output is a classical description of a Gaussian state $\tilde{\rho}$, such that

$$\frac{1}{2}\|\tilde{\rho}-\rho\|_1 \leq \varepsilon\,. \tag{2.327}$$

The classical description consists of the first moment $\mathbf{m}(\tilde{\rho}) \in \mathbb{R}^{2n}$ and in the covariance matrix $V(\tilde{\rho}) \in \mathbb{R}^{2n,2n}$, which uniquely characterise the Gaussian state $\tilde{\rho}$ and satisfy

$$\Pr\left(\|\tilde{V}-V(\rho)\|_\infty \leq \frac{\varepsilon^2}{2^7En^2} \quad \text{and} \quad \|\tilde{\mathbf{m}}-\mathbf{m}(\rho)\|_2 \leq \frac{\varepsilon^2}{2^7E^{3/2}n^{5/2}}\right) \geq 1-\delta\,. \tag{2.328}$$

1: Query N copies of ρ and apply the algorithm in Table 2.2, which outputs, with probability $\geq 1-\delta$, a vector $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$ and a covariance matrix $\tilde{V} \in \mathbb{R}^{2n,2n}$. In case of failure of this algorithm, which happens with probability less than δ, declare *failure* and abort.
2: Define $\tilde{\rho}$ as the Gaussian state with first moment $\mathbf{m}(\tilde{\rho}) = \tilde{\mathbf{m}}$ and covariance matrix $V(\tilde{\rho}) = \tilde{V}$.
3: **return** the vector $\tilde{\mathbf{m}}$ and the matrix $\tilde{V}$, which form a classical description of the Gaussian state $\tilde{\rho}$.

Tomography algorithm for mixed Gaussian states In Table 2.3 we present a tomography algorithm to learn a classical description of an unknown n-mode Gaussian state ρ, by possessing the prior knowledge that ρ has mean energy per mode bounded by some known constant $E > 0$. The correctness of such a tomography algorithm is proved in Theorem 68. The idea of such a tomography algorithm is trivial: estimate first moment and covariance matrix, and then outputs the Gaussian state with these first moment and covariance matrix. The non-trivial aspect of the algorithm concerns the underlying sample complexity analysis, which crucially relies on upper bounding the error in trace distance in terms of the errors of first moment and covariance matrix. We have proved such an upper bound in Theorem 54 above.

Theorem 68. *Let $\varepsilon, \delta \in (0,1)$ and $E \geq 0$. Let ρ be an n-mode Gaussian state satisfying the energy constraint $\mathrm{Tr}[\rho\hat{E}_n] \leq nE$. A number of copies N of ρ, such that*

$$\begin{aligned} N &:= (n+3)\left\lceil 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(24n^2E^2+3n)}{\varepsilon^4}2^{14}E^2n^4\right\rceil \\ &= O\left(\log\left(\frac{n^2}{\delta}\right)\frac{n^7E^4}{\varepsilon^4}\right), \end{aligned} \tag{2.329}$$

are sufficient to build a classical description of a Gaussian state $\tilde{\rho}$ such that

$$\Pr\left(\frac{1}{2}\|\tilde{\rho}-\rho\|_1 \leq \varepsilon\right) \geq 1-\delta\,. \tag{2.330}$$

The classical description consists of the first moment $\boldsymbol{m}(\tilde{\rho}) \in \mathbb{R}^{2n}$ and in the covariance matrix $V(\tilde{\rho}) \in \mathbb{R}^{2n,2n}$, which uniquely characterise the Gaussian state $\tilde{\rho}$ and satisfy

$$\Pr\left(\|\tilde{V} - V(\rho)\|_\infty \leq \frac{\varepsilon^2}{2^7 E n^2} \quad \textit{and} \quad \|\tilde{\boldsymbol{m}} - \boldsymbol{m}(\rho)\|_2 \leq \frac{\varepsilon^2}{2^7 E^{3/2} n^{5/2}}\right) \geq 1 - \delta\,. \tag{2.331}$$

Proof. We aim to establish the correctness of the algorithm presented in Table 2.3. First, let us observe that the second moment of the energy of the Gaussian state ρ, which satisfies the energy constraint $\operatorname{Tr}[\rho \hat{E}_n] \leq nE$, can be upper bounded as

$$\frac{1}{n}\sqrt{\operatorname{Tr}[\rho \hat{E}_n^2]} \leq \sqrt{3}\frac{\operatorname{Tr}[\rho \hat{E}_n]}{n} \leq \sqrt{3}E\,, \tag{2.332}$$

where we have used Lemma 67. Consequently, for any $\varepsilon' \in (0, \frac{1}{2})$ Theorem 65 establishes that a number

$$N \coloneqq (n+3)\left\lceil 68 \log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(24n^2E^2+3n)}{\varepsilon'^2}\right\rceil \tag{2.333}$$

of copies of ρ suffices to construct a vector $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$ and a covariance matrix $\tilde{V} \in \mathbb{R}^{2n,2n}$ such that

$$\Pr\left(\|\tilde{V} - V(\rho)\|_\infty \leq \varepsilon' \quad \text{and} \quad \|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \leq \frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}\right) \geq 1 - \delta\,. \tag{2.334}$$

Let $\tilde{\rho}$ be the Gaussian state with first moment $\mathbf{m}(\tilde{\rho}) = \tilde{\mathbf{m}}$ and covariance matrix $V(\tilde{\rho}) = \tilde{V}$. Let us show that, with probability at least $1-\delta$, the energy of the estimator $\tilde{\rho}$ is at most twice the energy of the true state ρ:

$$\begin{aligned}
\operatorname{Tr}[\tilde{\rho}\hat{E}_n] &\overset{\text{(i)}}{=} \frac{\operatorname{Tr}\tilde{V}}{4} + \frac{\|\tilde{\mathbf{m}}\|_2^2}{2} \\
&\overset{\text{(ii)}}{\leq} \frac{\operatorname{Tr} V(\rho)}{4} + \frac{2n\|\tilde{V} - V(\rho)\|_\infty}{4} + \frac{(\|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 + \|\mathbf{m}(\rho)\|_2)^2}{2} \\
&\overset{\text{(iii)}}{\leq} \operatorname{Tr}[\rho\hat{E}_n] + \frac{n\varepsilon'}{2} + \frac{\left(\frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}\right)^2 + 2\|\mathbf{m}(\rho)\|_2\left(\frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}\right)}{2} \\
&\leq \operatorname{Tr}[\rho\hat{E}_n] + \frac{n\varepsilon'}{2} + \frac{1 + 2\|\mathbf{m}(\rho)\|_2}{2}\left(\frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}\right) \\
&\overset{\text{(iv)}}{\leq} \operatorname{Tr}[\rho\hat{E}_n] + \frac{n\varepsilon'}{2} + \frac{1 + 2\sqrt{2nE}}{2}\left(\frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}\right) \\
&\leq \operatorname{Tr}[\rho\hat{E}_n] + \frac{n\varepsilon'}{2} + \frac{\varepsilon'}{10(3)^{1/4}} \\
&\leq \operatorname{Tr}[\rho\hat{E}_n] + n\varepsilon' \\
&\leq \operatorname{Tr}[\rho\hat{E}_n] + \frac{n}{2} \\
&\leq 2\operatorname{Tr}[\rho\hat{E}_n]\,.
\end{aligned} \tag{2.335}$$

Here, in (i), we exploited Lemma 60. In (ii), we have used Hölder inequality, triangle inequality, and the fact that $\|\tilde{V} - V(\rho)\|_\infty \leq \varepsilon'$. In (iii), we exploited again Lemma 60 and we have used that $\|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \leq \frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}$. The inequality in (iv) follows from (2.315). Then, building on

Theorem 54, we obtain that, with probability at least $1-\delta$, it holds that

$$\begin{aligned}
\frac{1}{2}\left\|\rho-\tilde{\rho}\right\|_1 &\leq 2\sqrt{\max\left(\mathrm{Tr}[\rho\hat{E}_n],\mathrm{Tr}[\tilde{\rho}\hat{E}_n]\right)}\left(\left\|\mathbf{m}(\rho)-\mathbf{m}(\tilde{\rho})\right\|_2+\sqrt{2}\sqrt{\left\|V(\rho)-V(\tilde{\rho})\right\|_1}\right)\\
&\leq 2\sqrt{2nE}\left(\left\|\mathbf{m}(\rho)-\mathbf{m}(\tilde{\rho})\right\|_2+\sqrt{2}\sqrt{\left\|V(\rho)-V(\tilde{\rho})\right\|_1}\right)\\
&\leq \sqrt{8nE}\left(\frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}+2\sqrt{n\left\|V(\rho)-V(\tilde{\rho})\right\|_\infty}\right)\\
&\leq \sqrt{8nE}\left(\frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}+2\sqrt{n\varepsilon'}\right)\\
&\leq \sqrt{8nE}\left(4\sqrt{n\varepsilon'}\right)\\
&= \sqrt{2^7En^2\varepsilon'}\,.
\end{aligned} \tag{2.336}$$

Consequently, by setting $\varepsilon' \coloneqq \frac{\varepsilon^2}{2^7En^2} \in (0,\frac{1}{2})$, we have that the choice

$$N \coloneqq (n+3)\left\lceil 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(24n^2E^2+3n)}{\varepsilon^4}2^{14}E^2n^4\right\rceil \tag{2.337}$$

allows us to guarantee that

$$\Pr\left(\frac{1}{2}\|\tilde{\rho}-\rho\|_1\leq\varepsilon\right)\geq 1-\delta\,. \tag{2.338}$$

With this choice of ε', we have also that

$$\Pr\left(\|\tilde{V}-V(\rho)\|_\infty\leq\frac{\varepsilon^2}{2^7En^2}\quad\text{and}\quad\|\tilde{\mathbf{m}}-\mathbf{m}(\rho)\|_2\leq\frac{\varepsilon^2}{10(3)^{1/4}2^7\sqrt{8}E^{3/2}n^{5/2}}\right)\geq 1-\delta\,, \tag{2.339}$$

which, in particular, implies that

$$\Pr\left(\|\tilde{V}-V(\rho)\|_\infty\leq\frac{\varepsilon^2}{2^7En^2}\quad\text{and}\quad\|\tilde{\mathbf{m}}-\mathbf{m}(\rho)\|_2\leq\frac{\varepsilon^2}{2^7E^{3/2}n^{5/2}}\right)\geq 1-\delta\,, \tag{2.340}$$

□

Noise robustness of the algorithm Above we have seen that learning an unknown Gaussian states is efficient, as the sample complexity scales polynomially in the number of modes. Here, we address the questions: *What happens if the state is not exactly Gaussian? Is our tomography procedure stable under little perturbations of the set of Gaussian states?* Answering these questions is conceptually relevant, as experimental imperfections may transform Gaussian states into slightly-perturbed Gaussian states.

Here, in order to quantify the non-Gaussian character of a quantum state, we use the *relative entropy of non-Gaussianity* [38, 80]. For any state ρ, the relative entropy of non-Gaussianity $d_{\mathscr{G}}(\rho)$ is defined as the minimum relative entropy between ρ and any Gaussian state:

$$d_{\mathscr{G}}(\rho)\coloneqq\min_{\sigma\in\mathscr{G}}S(\rho\|\sigma)\,, \tag{2.341}$$

where $\mathscr{G}$ denotes the set of Gaussian states and

$$S(\rho\|\sigma)\coloneqq\mathrm{Tr}[\rho\log_2\rho]-\mathrm{Tr}[\rho\log_2\sigma] \tag{2.342}$$

denotes the quantum relative entropy between ρ and σ. The quantum relative entropy $S(\rho\|\sigma)$ has a strong operation meaning in terms of the optimal rate in asymmetric quantum hypothesis testing between ρ and σ, due to the quantum Stein's lemma [78, 79].

As proved by [80], the relative entropy of non-Gaussianity satisfies desirable properties that make it a meaningful measure of non-Gaussianity. In particular, note that $d_G(\rho)$ is faithful, in the sense that $d_G(\rho) \geq 0$ and it vanishes if and only if ρ is Gaussian. Notably, the minimum in the definition of $d_{\mathscr{G}}(\rho)$ is achieved by the *Gaussianification* of ρ [38]. The Gaussianification $G(\rho)$ of a state ρ is the Gaussian state with the same first moment and covariance matrix of ρ. In formula, the relative entropy of non-Gaussianity is given by [38]:

$$d_{\mathscr{G}}(\rho) = S(\rho \| G(\rho)) \ . \tag{2.343}$$

Let us now analyse the stability of our tomography algorithm for Gaussian states. We make use the of the following simple observation.

Lemma 69. *If the unknown state ρ is not Gaussian, the algorithm designed for learning Gaussian states in Table 2.3 effectively learns the Gaussianification $G(\rho)$. Mathematically, $O\Big(\log\Big(\frac{n^2}{\delta}\Big)\frac{n^7E^4}{\varepsilon^4}\Big)$ copies of ρ suffices in order to build a classical description of a Gaussian state $\tilde{\rho}$ such that $\Pr\Big(\frac{1}{2}\|\tilde{\rho} - G(\rho)\|_1 \leq \varepsilon\Big) \geq 1-\delta$.*

Proof. The claim follows by two main observations: (i) the algorithm in Table 2.3 only involves the first moment and the covariance matrix of the unknown state, and (ii) the mean photon number of a state ρ and the one of its Gaussianification $G(\rho)$ coincide (because of Lemma 60). □

Now we show that *quantum state tomography of slightly-perturbed Gaussian states is efficient*. Here, by 'slightly-perturbed Gaussian state' we mean that the relative entropy of non-Gaussianity is sufficiently small.

Theorem 70 (Quantum state tomography of slightly-perturbed Gaussian states is efficient)**.** *Let $\varepsilon, \delta \in (0,1)$ and $E \geq 0$. Let ρ be an unknown n-mode (possibly non-Gaussian) state such that its relative entropy of non-Gaussianity satisfies $d_{\mathscr{G}}(\rho) \leq \varepsilon^2$. Assume that ρ satisfies the energy constraint $\mathrm{Tr}[\rho \hat{E}_n] \leq nE$. Then, $O\Big(\log\Big(\frac{n^2}{\delta}\Big)\frac{n^7E^4}{\varepsilon^4}\Big)$ copies of ρ suffices to build a classical description of a Gaussian state $\tilde{\rho}$ such that $\Pr\Big(\frac{1}{2}\|\tilde{\rho} - \rho\|_1 \leq \varepsilon\Big) \geq 1-\delta$.*

Proof. Thanks to Remark 69, $O\Big(\log\Big(\frac{n^2}{\delta}\Big)\frac{n^7E^4}{\varepsilon^4}\Big)$ copies of ρ suffices to build a classical description of a Gaussian state $\tilde{\rho}$ such that $\frac{1}{2}\|\tilde{\rho} - G(\rho)\|_1 \leq \Big(1 - \sqrt{\frac{\ln 2}{2}}\Big)\varepsilon$ with probability at least $1-\delta$, where $G(\rho)$ denotes the Gaussianification of ρ. If this event happens, then

$$\begin{aligned}
\frac{1}{2}\|\rho - \tilde{\rho}\|_1 &\leq \frac{1}{2}\|G(\rho) - \tilde{\rho}\|_1 + \frac{1}{2}\|G(\rho) - \rho\|_1 \\
&\leq \left(1 - \sqrt{\frac{\ln 2}{2}}\right)\varepsilon + \frac{1}{2}\|G(\rho) - \rho\|_1 \\
&\overset{\text{(i)}}{\leq} \left(1 - \sqrt{\frac{\ln 2}{2}}\right)\varepsilon + \sqrt{\frac{\ln 2}{2}}\sqrt{S(\rho\|G(\rho))} \\
&\overset{\text{(ii)}}{=} \left(1 - \sqrt{\frac{\ln 2}{2}}\right)\varepsilon + \sqrt{\frac{\ln 2}{2}}\sqrt{d_{\mathscr{G}(\rho)}} \\
&\leq \varepsilon \, .
\end{aligned} \tag{2.344}$$

Here, in (i), we employed the quantum Pinsker inequality [86, Theorem 11.9.1], which states that for any τ and σ the trace distance can be upper bounded in terms of the relative entropy as $\frac{1}{2}\|\tau - \sigma\|_1 \leq \sqrt{\frac{\ln 2}{2} S(\tau\|\sigma)}$. Finally, in (ii), we just employed the characterisation of the relative entropy of non-Gaussianity in (2.343). □

Improved tomography algorithm for pure Gaussian states In this subsubsection we present an improved upper bound on the sample complexity of tomography of pure Gaussian states. We can show that $O\Big(\log\Big(\frac{n^2}{\delta}\Big)\frac{n^5E^4}{\varepsilon^4}\Big)$ state copies suffices to achieve tomography of pure Gaussian states, thus obtaining an improvement with respect the scaling $O\Big(\log\Big(\frac{n^2}{\delta}\Big)\frac{n^7E^4}{\varepsilon^4}\Big)$ of the mixed setting analysed in subsubsection 2.7.3. The main technical tool employed here is the improved upper bound on the trace distance between a pure Gaussian state and an arbitrary (possibly-mixed) state presented in Theorem 61.

Theorem 71 (Tomography of pure Gaussian states)**.** *Let $\varepsilon, \delta \in (0,1)$ and $E \geq 0$. Let ψ be an n-mode pure Gaussian state satisfying the energy constraint $\mathrm{Tr}[\psi \hat{E}_n] \leq nE$. A number of copies N of ψ, such that*

$$\begin{aligned} N &:= (n+3)\left\lceil 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(24n^2E^2+3n)}{\varepsilon^4}16E^2n^2\right\rceil \\ &= O\left(\log\left(\frac{n^2}{\delta}\right)\frac{n^5E^4}{\varepsilon^4}\right), \end{aligned} \tag{2.345}$$

are sufficient to build a classical description of a Gaussian state $\tilde{\rho}$ such that

$$\Pr\left(\frac{1}{2}\|\tilde{\rho}-\psi\|_1 \leq \varepsilon\right) \geq 1-\delta\,. \tag{2.346}$$

Proof. We proceed as in Lemma 68. First, let us observe that Lemma 67 establishes that the second moment of the energy of the Gaussian state ψ, which satisfies the energy constraint $\mathrm{Tr}[\psi\hat{E}_n] \leq nE$, can be upper bounded as

$$\frac{1}{n}\sqrt{\mathrm{Tr}[\psi\hat{E}_n^2]} \leq \sqrt{3}\frac{\mathrm{Tr}[\psi\hat{E}_n]}{n} \leq \sqrt{3}E\,. \tag{2.347}$$

Consequently, for any $\varepsilon' \in (0, \frac{1}{2})$ Theorem 65 establishes that a number

$$N := (n+3)\left\lceil 68\log\left(\frac{2(2n^2+3n)}{\delta}\right)\frac{200(24n^2E^2+3n)}{\varepsilon'^2}\right\rceil \tag{2.348}$$

of copies of ψ suffices to construct a vector $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$ and a covariance matrix $\tilde{V} \in \mathbb{R}^{2n,2n}$ such that

$$\Pr\left(\|\tilde{V}-V(\psi)\|_\infty \leq \varepsilon' \quad \text{and} \quad \|\tilde{\mathbf{m}}-\mathbf{m}(\psi)\|_2 \leq \frac{\varepsilon'}{10(3)^{1/4}\sqrt{8En}}\right) \geq 1-\delta\,. \tag{2.349}$$

Let $\tilde{\rho}$ be the Gaussian state with first moment $\mathbf{m}(\tilde{\rho}) = \tilde{\mathbf{m}}$ and covariance matrix $V(\tilde{\rho}) = \tilde{V}$. By exploiting the same reasoning used in (2.335), we have that the energy of the estimator $\tilde{\rho}$ is at most twice the energy of ψ, i.e.

$$\mathrm{Tr}[\tilde{\rho}\hat{E}_n] \leq 2\,\mathrm{Tr}[\psi\hat{E}_n]\,, \tag{2.350}$$

with probability at least $1-\delta$. Then, by exploiting Theorem 61, we have that, with probability at least $1-\delta$, it holds that

$$\begin{aligned} \frac{1}{2}\|\,\tilde{\rho}-\psi\,\|_1 &\leq \sqrt{\max\left(\mathrm{Tr}[\tilde{\rho}\hat{E}_n], \mathrm{Tr}[\psi\hat{E}_n]\right)}\sqrt{\|V(\tilde{\rho})-V(\psi)\|_\infty + 2\|\mathbf{m}(\tilde{\rho})-\mathbf{m}(\psi)\|_2^2} \\ &\leq \sqrt{2nE}\sqrt{\|V(\tilde{\rho})-V(\psi)\|_\infty + 2\|\mathbf{m}(\tilde{\rho})-\mathbf{m}(\psi)\|_2^2} \\ &\leq \sqrt{2nE}\sqrt{\varepsilon' + \frac{\varepsilon'^2}{400\sqrt{3}\,En}} \\ &\leq \sqrt{4nE\varepsilon'}\,. \end{aligned} \tag{2.351}$$

Consequently, by setting $\varepsilon' \coloneqq \frac{\varepsilon^2}{4nE} \in (0, \frac{1}{2})$, we have that the choice

$$N \coloneqq (n+3)\left\lceil 68 \log\left(\frac{2(2n^2+3n)}{\delta}\right) \frac{200(24n^2E^2+3n)}{\varepsilon^4} 16E^2n^2 \right\rceil \tag{2.352}$$

allows us to guarantee that

$$\Pr\left(\frac{1}{2}\|\tilde{\rho} - \psi\|_1 \leq \varepsilon\right) \geq 1 - \delta\,. \tag{2.353}$$

□

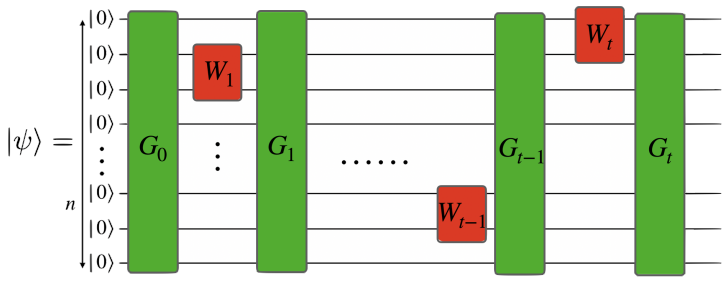


Figure 2.7: Depiction of a t-doped Gaussian state. The state vector $|\psi\rangle$ is prepared by applying Gaussian unitaries (green gates) and at most t local non-Gaussian unitaries (red gates). In the specific example depicted in the figure, the locality is $\kappa = 4$ as each of the non-Gaussian unitaries acts non-trivially on two modes (thus 4 quadratures).

Tomography of t-doped bosonic Gaussian states

In the preceding section, we rigorously proved that quantum state tomography of bosonic Gaussian states is efficient. In this section, we delve into the analysis of *t-doped bosonic Gaussian states*, which are states prepared by Gaussian unitaries and at most t non-Gaussian local unitaries. They encompass a much broader class of efficiently learnable states than exact bosonic Gaussian states.

The results we will show in this section can be seen as an extension to the bosonic setting of what was previously shown for t-doped stabiliser states [40, 45, 69] (states prepared by Clifford gates and at most t T-gates) and t-doped fermionic Gaussian states of Ref. [46] (states prepared by fermionic Gaussian unitaries and at most t fermionic non-Gaussian local unitaries). However, extending these results is far from trivial, as in the bosonic setting one must deal not only with different commutation relations than in the fermionic setting, but also with subtleties arising from the energy constraints and from the infinite-dimensional Hilbert space. Just as T-gates are considered *magic* gates for Clifford circuits, which are classically simulable [87], local non-Gaussian gates can similarly be viewed as *magic* gates for Gaussian circuits, which are also classically simulable [88]. Recent works have also been focusing on the classical simulability of states prepared by a Gaussian evolution applied to an input state that is a superposition of a bounded number of Gaussian states [89–91].

An n-mode unitary U is said to be a (t, κ)-*doped Gaussian unitary* if it is a composition of Gaussian unitaries and at most t non-Gaussian κ-local unitaries, where κ-local means that each non-Gaussian gate involves at most κ quadrature operators. In other words, U is of the form

$$U = G_t W_t \cdots G_1 W_1 G_0 \,, \tag{2.354}$$

where each G_i is an n-mode Gaussian unitary and W_i is a κ-local non-Gaussian unitary. Strictly speaking, we assume each W_i to be a unitary generated by a Hamiltonian which is a (non-quadratic) polynomial in at most κ quadrature operators. An n-mode state vector $|\psi\rangle$ is said to be a (t, κ)-*doped Gaussian state* if it can be prepared by applying a (t, κ)-doped Gaussian unitary to the vacuum, that is,

$$|\psi\rangle = U\,|0\rangle^{\otimes n} \,. \tag{2.355}$$

Informally, we sometimes say that a state is a t-doped Gaussian state (by omitting the dependence on κ) if it is a (t, κ)-doped Gaussian state for $\kappa = O(1)$ in the number of modes. Fig. 2.7.3 shows a depiction of a t-doped Gaussian state.

The forthcoming Theorem 72 provides a remarkable decomposition of (t, κ)-doped unitaries and states, which shows that, if $\kappa t \leq n$, one can turn any t-doped state into a tensor product between a κt-mode non-Gaussian state and the $(n - \kappa t)$-mode vacuum state via a suitable Gaussian unitary. We provide a depiction of such a decomposition in Fig. 2.7.3

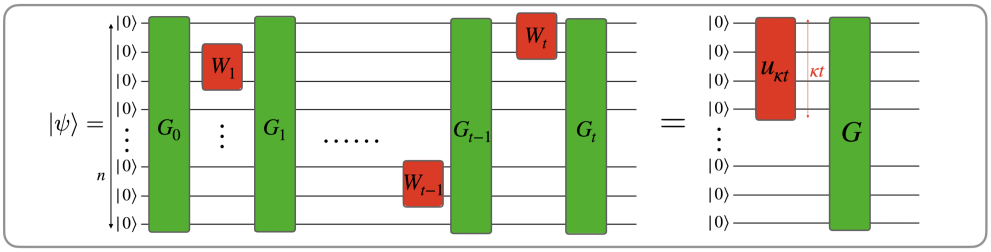


Figure 2.8: Depiction of the decomposition of a t-doped Gaussian state vector $|\psi\rangle$ proved in Theorem 72. All the 'non-Gaussianity' of $|\psi\rangle$ can be compressed on the first κt modes by applying to $|\psi\rangle$ a suitable Gaussian unitary $G^\dagger$.

Theorem 72 (Non-Gaussianity compression in t-doped Gaussian unitaries and states)**.** *If $n \geq \kappa t$, any n-mode (t, κ)-doped Gaussian unitary U can be decomposed as*

$$U = G(u_{\kappa t} \otimes \mathbb{1}_{n-\kappa t}) G_{passive} \tag{2.356}$$

for some suitable Gaussian unitary G, passive Gaussian unitary $G_{passive}$, and κt-mode (non-Gaussian) unitary $u_{\kappa t}$. In particular, any n-mode (t, κ)-doped Gaussian state can be decomposed as

$$|\psi\rangle = G\left(|\phi_{\kappa t}\rangle \otimes |0\rangle^{\otimes(n-\kappa t)}\right) \tag{2.357}$$

for some suitable Gaussian unitary G and κt-mode (non-Gaussian) state vector $|\phi_{\kappa t}\rangle$.

The decomposition in (2.357) shows that it is possible to compress all the non-Gaussianity of a t-doped Gaussian state via a suitable Gaussian unitary. By leveraging such a property, we devise a tomography algorithm for t-doped Gaussian states which has a sample and time complexity that scales polynomially in n as long as $\kappa t = O(1)$, *thereby establishing that tomography of (energy-constrained) t-doped states is efficient* in this regime. This establishes the robustness of tomography of Gaussian states, in the sense that, even if few non-Gaussian unitaries are applied to a Gaussian state, the resulting state remains efficiently learnable.

The rough idea behind our tomography algorithm for unknown t-doped Gaussian states $|\psi\rangle$ involves first estimating the Gaussian unitary G (see (2.357)), then applying its inverse to the state in order to compress the non-Gaussianity, and finally performing tomography of the first κt modes. Remarkably, our tomography algorithm is experimentally feasible, as it uses only Gaussian unitaries and easily implementable Gaussian measurements, such as homodyne and heterodyne detection [7, 55].

In the following theorem, we analyse the performance guarantees of our tomography algorithm.

Theorem 73 (Informal version)**.** *Let $|\psi\rangle$ be an unknown n-mode (t, κ)-doped Gaussian state vector, which satisfies the second-moment constraint $\mathrm{Tr}[\psi \hat{N}_n^2] \leq n^2 N_{phot}^2$. Then, $O\left(\left(\frac{n^2 N_{phot}^2}{\varepsilon^2}\right)^{\kappa t}\right)$ state copies suffices to construct a succint classical description of an estimator $|\tilde{\psi}\rangle$ which is ε-close in trace distance to $|\psi\rangle$ with high probability. Such a classical description consists of a triplet $(\tilde{\boldsymbol{m}}, \tilde{S}, |\tilde{\phi}_1\rangle)$, which defines the estimator via the relation*

$$|\tilde{\psi}\rangle := \hat{D}_{\tilde{\boldsymbol{m}}} U_{\tilde{S}} |\tilde{\phi}_1\rangle \otimes |0\rangle^{\otimes(n-\kappa t)} , \tag{2.358}$$

where $\hat{D}_{\tilde{\boldsymbol{m}}}$ is the displacement operator with $\tilde{\boldsymbol{m}} \in \mathbb{R}^{2n}$, $U_{\tilde{S}}$ is a Gaussian unitary associated with the $2n \times 2n$ symplectic matrix $\tilde{S}$, and $|\tilde{\phi}_1\rangle$ is a κt-mode state contained in the d_{eff}-dimensional subspace spanned by all the n-mode Fock states with total photon number less than $O\left(\left(\frac{n^2 N_{phot}^2}{\varepsilon^2}\right)\right)$, where

$$d_{\mathit{eff}} := O\left(\left(\frac{n^2 N_{phot}^2}{\varepsilon^2}\right)^{\kappa t}\right) . \tag{2.359}$$

In particular, tomography of t-doped Gaussian states is efficient in the regime $\kappa t = O(1)$, as its sample, time, and memory complexity scales polynomially in n.

We conclude by showing that any tomography algorithm that aims to learn quantum states which can be written as $G(|\phi_t\rangle \otimes |0\rangle^{\otimes(n-t)})$, where G is a Gaussian unitary and $|\phi_t\rangle$ a t-mode quantum state vector, must be inefficient if t scales *slightly* more than a constant in the number of modes.

2.7.4 t-doped Gaussian states and unitaries

We now present the formal definitions crucial for our analysis.

Definition 74 (Non-Gaussian k-local unitary)**.** *Let $n \in \mathbb{N}$ and $\kappa \in [2n]$. A κ-local non-Gaussian unitary is an n-mode unitary generated by a Hamiltonian which is a polynomial in at most k quadrature operators $\{\hat{R}_{\mu(r)}\}_{r=1}^{\kappa}$, where $\mu(1), \ldots, \mu(\kappa) \in [2n]$.*

For instance, unitaries like $\exp[i\hat{x}_1^4]$, $\exp[i(\hat{x}_1^4 + \hat{p}_1^4)]$, $\exp[i(\hat{x}_1^4 + \hat{p}_1^4 + \hat{x}_3^6)]$, and $\exp\left[i0.7\hat{a}_1^\dagger \hat{a}_2^\dagger \hat{a}_3 \hat{a}_4\right]$ are examples of 1-, 2-, 3-local, and 4-local non-Gaussian unitaries, respectively.

Definition 75 (t-doped Gaussian unitary)**.** *A unitary U_t is a (t, κ)-doped Gaussian unitary if it can be prepared by Gaussian unitaries $\{G_i\}_{i=0}^{t}$ and t non-Gaussian κ-local unitaries $\{W_i\}_{i=1}^{t}$, specifically given by*

$$U_t = G_t W_t \cdots G_1 W_1 G_0 \,. \tag{2.360}$$

Definition 76 (t-doped Gaussian state)**.** *A (t, κ)-doped Gaussian state vector $|\psi\rangle$ is obtained by the action of a (t, κ)-doped Gaussian unitary U_t on the vacuum, i.e., $|\psi\rangle := U_t |0\rangle^{\otimes n}$.*

Compression of the non-Gaussianity

In this section, we introduce a finding that sheds light on the structure of t-doped Gaussian unitaries: the 'non-Gaussianity' can be compressed into the first modes through a Gaussian operation. This result is an extension to the bosonic setting of a result proved in Ref. [46], formulated for t-doped *fermionic* unitaries. The proof for bosons can be easily generalised from the one for fermions of Ref. [46], by making use of the following parallelism between the theory of bosonic Gaussian and fermionic Gaussian unitaries. Specifically, in the bosonic setting, Gaussian unitaries correspond to symplectic matrices, and Gaussian passive unitaries are associated with symplectic-orthogonal matrices. In the fermionic context, the role of symplectic and orthogonal matrices is exchanged: Gaussian unitaries correspond to orthogonal matrices, and Gaussian passive unitaries are associated with symplectic-orthogonal matrices. Consequently, when handling Gaussian passive transformations, the theory of Gaussian bosons aligns with the theory of Gaussian fermions. In the following proof, we follow the construction presented in Ref. [46].

Theorem 77 (Compression of non-Gaussianity in t-doped unitaries)**.** *Let $n \in \mathbb{N}$ be the number of modes. Let $\kappa, t \in N$ such that $\kappa t \leq n$. For any (t, κ)-doped bosonic Gaussian unitary U_t there exist a Gaussian unitary G, a passive Gaussian unitary $G_{passive}$, and a (possibly non-Gaussian) unitary $u_{\kappa t}$ supported exclusively on the first κt modes such that*

$$U_t = G(u_{\kappa t} \otimes \mathbb{1}_{n-\kappa t}) G_{passive}, \tag{2.361}$$

where $\mathbb{1}_{n-\kappa t}$ denotes the identity on the last $n - \kappa t$ modes.

Proof. By definition, the t-doped unitary can be expressed as $U_t = (\prod_{t'=1}^{t} G_{t'} W_{t'}) G_0$. Rearranging it, we express it as

$$U_t = \tilde{G}_t \prod_{t'=1}^{t} \tilde{W}_{t'}, \tag{2.362}$$

where $\tilde{W}_{t'} \coloneqq \tilde{G}_{t'-1}^{\dagger} W_{t'} \tilde{G}_{t'-1}$ and $\tilde{G}_{t'} \coloneqq G_{t'} .. G_0$. Let G_{aux} be a passive Gaussian unitary, which we will fix later in order to compress all the non-Gaussianity to the first κt modes. We can rewrite U_t as

$$U_t = \tilde{G}_t G_{\text{aux}} \prod_{t'=1}^{t} (G_{\text{aux}}^{\dagger} \tilde{W}_{t'} G_{\text{aux}}) G_{\text{aux}}^{\dagger} \,. \tag{2.363}$$

Moreover, by defining

$$\begin{aligned} G &\coloneqq \tilde{G}_t G_{\text{aux}} \,, \\ u_{\kappa t} &\coloneqq \prod_{t'=1}^{t} (G_{\text{aux}}^{\dagger} \tilde{W}_{t'} G_{\text{aux}}) \,, \\ G_{\text{passive}} &\coloneqq G_{\text{aux}}^{\dagger} \,, \end{aligned} \tag{2.364}$$

we have

$$U_t = G u_{\kappa t} G_{\text{passive}} \,. \tag{2.365}$$

Observe that G is a Gaussian unitary, as the product of Gaussian unitaries remains Gaussian. Similarly, G_{passive} is a passive Gaussian unitary, given that the adjoint of a passive Gaussian unitary remains passive and Gaussian. In order to conclude the proof, we have to show that we can choose G_{aux} such that $u_{\kappa t}$ is supported only on the first κt modes. It suffices to show that the Heisenberg evolution, under the Gaussian unitary $\tilde{G}_{t'-1} G_{\text{aux}}$, of the Hamiltonian generating $W_{t'}$ involves only the first κt quadratures $(\hat{R}_i)_{i \in [\kappa t]}$. In other words, we need to show that the Heisenberg evolution of each quadrature operator involved in the Hamiltonian generating $W_{t'}$ involves only the first κt quadratures. Let O_{aux} be the symplectic orthogonal matrix associated with G_{aux}. For each $t' \in [t]$ let $\mu(t',1), \mu(t',2), \ldots, \mu(t',\kappa)$, with $\mu(t',1) \leq \mu(t',2) \leq \ldots \leq \mu(t',\kappa)$, be the quadratures involved in the Hamiltonian generating $W_{t'}$. For each $t' \in [t]$ and $r \in [\kappa]$ it holds that

$$G_{\text{aux}}^{\dagger} \tilde{G}_{t'-1}^{\dagger} \hat{R}_{\mu(t',r)} \tilde{G}_{t'-1} G_{\text{aux}} = \sum_{m=1}^{2n} (S_{t'-1} O_{\text{aux}})_{\mu(t',r),m} \hat{R}_m \,, \tag{2.366}$$

where $S_{t'-1}$ is the symplectic matrix associated with $\tilde{G}_{t'-1}$. Consequently, we have to choose O_{aux} so that

$$(S_{t'-1} O_{\text{aux}})_{\mu(t',r),m} = (O_{\text{aux}}^{\intercal} S_{t'-1}^{\intercal})_{m,\mu(t',r)} = 0 \tag{2.367}$$

for all $t' \in [t]$, $r \in [\kappa]$, and all $m \in \{2\kappa t + 1, \ldots, 2n\}$. Let $(\mathbf{e}_i)_{i \in [2n]}$ be the canonical basis vectors of $\mathbb{R}^{2n}$. For simplicity, let us denote the vectors $(S_{t'-1}^{\intercal} \mathbf{e}_{\mu(t',r)})_{t' \in [t], r \in [\kappa]}$ with $(\mathbf{v}_j)_{j \in [\kappa t]}$. It suffices to show that there exists an orthogonal symplectic matrix O_{aux} such that for all $j \in [\kappa t]$ and all $m \in \{2\kappa t + 1, \ldots, 2n\}$ it holds that

$$\mathbf{e}_m^T O_{\text{aux}}^{\intercal} \mathbf{v}_j = 0 \,. \tag{2.368}$$

For any arbitrary vectors $\{\mathbf{v}_j\}_{j \in [\kappa t]}$, there exist such an orthogonal matrix O_{aux}. This is a consequence of the forthcoming Lemma 78, pointed out in [46], which is a consequence of the well-known isomorphism between $2n \times 2n$ symplectic orthogonal real matrices and $n \times n$ unitaries [7]. □

Lemma 78. *Let $\{\mathbf{e}_i\}_{i=1}^{2n}$ be the canonical basis of $\mathbb{R}^{2n}$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_M \in \mathbb{R}^{2n}$ be a set of real vectors, where $M \leq 2n$. There exists an orthogonal symplectic matrix $O \in \mathrm{O}(2n) \cap \mathrm{Sp}(2n, \mathbb{R})$ such that*

$$\mathbf{e}_i^T O \mathbf{v}_j = 0, \tag{2.369}$$

for all $i \in \{2M + 1, \ldots, 2n\}$ and $j \in [M]$, meaning that all $\{O\mathbf{v}_j\}_{j=1}^{M}$ are exclusively supported on the span of the first $2M$ canonical basis vectors.

The proof of Lemma 78 can be found in [46]. The decomposition for t-doped Gaussian unitaries proved in Theorem 77 implies a similar for t-doped states, as established by the following theorem.

Theorem 79. *Let $n \in \mathbb{N}$ be the number of modes. Let $\kappa, t \in \mathbb{N}$ such that $\kappa t \leq n$. For any (t,κ)-doped Gaussian state vector $|\psi\rangle$ there exists a Gaussian unitary G and a (possibly non-Gaussian) κt-mode state vector $|\phi\rangle$ such that*

$$|\psi\rangle = G\left(|\phi\rangle \otimes |0\rangle^{\otimes(n-\kappa t)}\right) . \tag{2.370}$$

Moreover, G can be expressed in terms of the first moment and the covariance matrix of $|\psi\rangle$ as follows. The symplectic diagonalisation of the covariance matrix is of the form

$$V(|\psi\rangle\langle\psi|) = S(D \oplus I_{2(n-\kappa t)})S^{\intercal} , \tag{2.371}$$

with D being a $2\kappa t \times 2\kappa t$ diagonal matrix and S being a symplectic matrix. Then, $|\psi\rangle$ can be expressed as

$$|\psi\rangle = \hat{D}_{\boldsymbol{m}(\psi)} U_S |\varphi_{\kappa t}\rangle \otimes |0\rangle^{\otimes(n-\kappa t)} , \tag{2.372}$$

for some κt-mode (possibly non-Gaussian) state vector $|\varphi_{\kappa t}\rangle$, where $\hat{D}_{\boldsymbol{m}(\psi)}$ is the displacement operator and U_S is the Gaussian unitary associated with the symplectic S.

Proof. By definition, any (t,κ)-doped Gaussian state vector $|\psi\rangle$ can be written in terms of a (t,κ)-doped Gaussian unitary U_t as $|\psi\rangle := U_t |0\rangle^{\otimes n}$. Consequently, Theorem 77 implies that

$$\begin{aligned} |\psi\rangle &= G(u_{\kappa t} \otimes \mathbb{1}_{n-\kappa t}) G_{\text{passive}} |0\rangle^{\otimes n} \\ &= G(u_{\kappa t} \otimes \mathbb{1}_{n-\kappa t}) |0\rangle^{\otimes n} , \end{aligned} \tag{2.373}$$

where we have used that G_{passive} is passive and thus it preserves the vacuum state, i.e., $G_{\text{passive}} |0\rangle^{\otimes n} = |0\rangle^{\otimes n}$. Hence, (2.370) follows by defining $|\phi\rangle := u_{\kappa t} |0\rangle^{\otimes \kappa t}$. Now, let us prove (2.372). Note that the first moment and the covariance matrix of the state vector $U_S^\dagger \hat{D}^\dagger_{\mathbf{m}(\psi)} |\psi\rangle$ are given by

$$\begin{aligned} \mathbf{m}\left(U_S^\dagger \hat{D}^\dagger_{\mathbf{m}(\psi)} |\psi\rangle\langle\psi| \hat{D}_{\mathbf{m}(\psi)} U_S\right) &= \mathbf{0}_n , \\ V\left(U_S^\dagger \hat{D}^\dagger_{\mathbf{m}(\psi)} |\psi\rangle\langle\psi| \hat{D}_{\mathbf{m}(\psi)} U_S\right) &= D \otimes I_{n-\kappa t} . \end{aligned} \tag{2.374}$$

and thus

$$\begin{aligned} \mathbf{m}\left(\mathrm{Tr}_{[\kappa t]}\left[U_S^\dagger \hat{D}^\dagger_{\mathbf{m}(\psi)} |\psi\rangle\langle\psi| \hat{D}_{\mathbf{m}(\psi)} U_S\right]\right) &= \mathbf{0}_{n-\kappa t} , \\ V\left(\mathrm{Tr}_{[\kappa t]}\left[U_S^\dagger \hat{D}^\dagger_{\mathbf{m}(\psi)} |\psi\rangle\langle\psi| \hat{D}_{\mathbf{m}(\psi)} U_S\right]\right) &= I_{n-\kappa t} . \end{aligned} \tag{2.375}$$

Hence, the forthcoming Lemma 80 establishes that

$$\mathrm{Tr}_{[\kappa t]}\left[U_S^\dagger \hat{D}^\dagger_{\mathbf{m}(\psi)} |\psi\rangle\langle\psi| \hat{D}_{\mathbf{m}(\psi)} U_S\right] = |0\rangle\langle 0|^{\otimes(n-\kappa t)} . \tag{2.376}$$

Consequently, there exists a κt-mode state vector $|\phi\rangle$ such that

$$U_S^\dagger \hat{D}^\dagger_{\mathbf{m}(\psi)} |\psi\rangle = |\phi\rangle \otimes |0\rangle^{\otimes(n-\kappa t)} , \tag{2.377}$$

which concludes the proof. □

Lemma 80. *Let ρ be an n-mode state with covariance matrix equal to the identity and first moment equal to $\boldsymbol{r} \in \mathbb{R}^{2n}$,*

$$\begin{aligned} V(\rho) &= I_{2n} , \\ \boldsymbol{m}(\rho) &= \boldsymbol{r} . \end{aligned} \tag{2.378}$$

Then, ρ is the coherent state $\rho = |\mathbf{r}\rangle\langle\mathbf{r}|$, where $|\mathbf{r}\rangle := \hat{D}_{\mathbf{r}} |0\rangle^{\otimes n}$.

Proof. It suffices to show that $\mathrm{Tr}[|\mathbf{r}\rangle\langle\mathbf{r}|\,\rho] = 1$, which is valid since it holds that

$$\begin{aligned} 1 &\geq \mathrm{Tr}[|\mathbf{r}\rangle\langle\mathbf{r}|\,\rho] \\ &= \mathrm{Tr}\Big[|0\rangle\langle 0|^{\otimes n}\,\hat{D}_{\mathbf{r}}^{\dagger}\rho\hat{D}_{\mathbf{r}}\Big] \\ &\geq \mathrm{Tr}\left[\left(\mathbb{1} - \sum_{i=1}^{n} a_i^{\dagger}a_i\right)\hat{D}_{\mathbf{r}}^{\dagger}\rho\hat{D}_{\mathbf{r}}\right] \\ &\overset{\text{(i)}}{=} 1 - \frac{\mathrm{Tr}\Big[V\Big(\hat{D}_{\mathbf{r}}^{\dagger}\rho\hat{D}_{\mathbf{r}}\Big) - I_{2n}\Big]}{4} - \frac{\|\mathbf{m}(\hat{D}_{\mathbf{r}}^{\dagger}\rho\hat{D}_{\mathbf{r}})\|_2^2}{2} \\ &= 1 - \frac{\mathrm{Tr}[V(\rho) - I_{2n}]}{4} - \frac{\|\mathbf{m}(\rho) - \mathbf{r}\|_2^2}{2} \\ &= 1\,, \end{aligned} \tag{2.379}$$

where in (i) we have used Lemma 60. □

t-compressible Gaussian states

In this section, we present the notion of t-compressible states.

Definition 81 (t-compressible states)**.** *Let $t, n \in \mathbb{N}$ with $t \leq n$. An n-mode state vector $|\psi\rangle$ is said to be t-compressible when there exists a Gaussian unitary G and a t-mode state vector $|\phi\rangle$ such that*

$$|\psi\rangle = G\,|\phi\rangle \otimes |0\rangle^{\otimes(n-t)}\,. \tag{2.380}$$

The above Gaussian unitary G can be chosen as follows.

Lemma 82 (Choice of states)**.** *Let $|\psi\rangle$ be an n-mode t-compressible state vector. Let*

$$V(|\psi\rangle\langle\psi|) = S\left(D_t \oplus I_{2(n-t)}\right)S^{\intercal}\,. \tag{2.381}$$

be the symplectic diagonalisation of the covariance matrix of $V(|\psi\rangle\langle\psi|)$. Then, $|\psi\rangle$ can be written as

$$|\psi\rangle = \hat{D}_{\boldsymbol{m}(|\psi\rangle\langle\psi|)}U_S\,|\phi\rangle \otimes |0\rangle^{\otimes(n-t)}\,, \tag{2.382}$$

where $|\phi\rangle$ is a t-mode state vector with first moment and covariance matrix given by

$$\begin{aligned} V(|\phi\rangle\langle\phi|) &= D_t\,, \\ \boldsymbol{m}(|\phi\rangle\langle\phi|) &= 0\,. \end{aligned} \tag{2.383}$$

Proof. The proof is the same as the one of Theorem 79. □

As a consequence of Theorem 79, we have the following.

Corollary 83 (Compressibility)**.** *A (t, κ)-doped Gaussian state is κt-compressible.*

Note that the covariance matrix of a t-compressible state has at least $n-t$ symplectic eigenvalues equal to 1. This motivates the following definition, in analogy to the stabiliser dimension [45] and fermionic Gaussian dimension [46].

Definition 84 (Gaussian dimension of a state)**.** *The Gaussian dimension of a state vector $|\psi\rangle$ is the number of symplectic eigenvalues of the covariance matrix $V(|\psi\rangle\langle\psi|)$ which are equal to one.*

We now observe the following fact.

Lemma 85 (Characterization of t-compressible pure states)**.** *An n-mode pure state is t-compressible if and only if its Gaussian dimension is at least $n - t$.*

Proof. First, let us assume that the Gaussian dimension of the pure state vector $|\psi\rangle$ is $n-t$. Hence, the symplectic diagonalisation of its covariance matrix is of the form

$$V(|\psi\rangle\langle\psi|) = S\left(D_t \oplus I_{2(n-t)}\right) S^\intercal \,. \tag{2.384}$$

Consequently, it holds that

$$V\Big(\mathrm{Tr}_{[t]}[U_S^\dagger\,|\psi\rangle\langle\psi|\,U_S]\Big) = I_{2(n-t)}\,. \tag{2.385}$$

Hence, Lemma 80 establishes that $\mathrm{Tr}_{[t]}[U_S^\dagger\,|\psi\rangle\langle\psi|\,U_S] = |\mathbf{r}\rangle\langle\mathbf{r}|$ for some $\mathbf{r}\in\mathbb{R}^{2(n-t)}$. Consequently, there exists a t-mode state vector $|\phi\rangle\langle\phi|$ such that

$$U_S^\dagger\,|\psi\rangle = |\phi\rangle\otimes|\mathbf{r}\rangle\ . \tag{2.386}$$

This implies that

$$|\psi\rangle = U_S(\mathbb{1}_t\otimes\hat{D}_{\mathbf{r}})\,|\phi\rangle\otimes|0\rangle^{\otimes(n-t)}\ , \tag{2.387}$$

and in particular that $|\psi\rangle$ is t-compressible. On the other hand, the fact that the Gaussian dimension of a t-compressible state is at least $n-t$ is trivial. □

Tomography algorithm for t-compressible Gaussian states

In this subsection, we present a tomography algorithm for learning t-compressible n-mode bosonic states. Thus, because of Theorem 83, this algorithm can also be applied to doped quantum states. Our algorithm is outlined in Table 2.4, and its correctness is proved in the forthcoming Theorem 86. The algorithm turns out to be efficient for $t = O(1)$ (i.e., it runs in polynomial time in the number of modes), while super-poly (yet sub-exponential) for $t = O(\log n)$. Conversely, in Theorem 92 we show that, in the regime $t = O(\log n)$ *any* tomography algorithm designed to learn t-compressible states must be inefficient.

Our tomography algorithm is motivated by Theorem 79 and consists of two main parts, summarised as follows: In the first part, it estimates the covariance matrix of the unknown state, expresses it in its Williamson decomposition (3.33), and uses this decomposition to find a Gaussian unitary $\tilde{G}^\dagger$ that can *approximately* transform the n-mode state into a tensor product between an arbitrary state on the first t modes and the vacuum state in the last $n-t$ modes. In the second part, the algorithm applies $\tilde{G}^\dagger$ to the unknown state. It then measures the occupation number of the last $n-t$ modes, and if the vacuum state is obtained, it performs full state tomography on only the first t modes, resulting in a state vector $|\tilde{\phi}\rangle$. The output of the algorithm will then be a classical representation of the state $\tilde{G}(|\tilde{\phi}\rangle\otimes|0\rangle^{\otimes(n-t)})$. Refer to Table 2.4 for more precise details of the algorithm.

Remarkably, our tomography algorithm is experimentally feasible, as it requires only Gaussian evolutions, avalanche photodiodes (devices capable of reliably distinguishing between zero and one or more photons, commonly referred to as 'on/off detectors' [7]), and easily implementable Gaussian measurements, such as homodyne and heterodyne detection [7]. Specifically, the estimation of the first moment and covariance matrix in Line 1 relies solely on homodyne measurements [7, 55]. Line 4 involves only Gaussian unitaries, while Line 5 uses a photon counting that post-selects on the vacuum of the last $n-t$ modes when no photons are detected. This latter measurement can be practically achieved using photodetectors, like avalanche photodiodes, capable of discriminating between zero and one or more photons. Lastly, in Line 8, quantum state tomography of the first t modes is performed. To accomplish this in an experimentally feasible manner, we may utilise the continuous-variable classical shadow algorithm proposed in [17], which relies solely on randomised Gaussian unitaries and homodyne and heterodyne measurements. (Alternatively, in order to improve the performance, we may also consider utilising our tomography algorithm for moment-constrained pure states detailed in Section 2.7.2. However, despite its better performance, the latter algorithm is not feasible for experimental implementation with our current technology.)

Before delving into the correctness proof of the algorithm, let us establish the notation. Let $N_{\mathrm{cov}}(n, \varepsilon, \delta, E_2)$ be the sample complexity of the algorithm outlined in Table 2.2 and Theorem 65 to estimate the first moment and the covariance matrix of an unknown n-mode state ρ, subject to the second-moment energy constraint $\mathrm{Tr}[\rho \hat{E}_n^2] \leq n^2 E_2^2$, with precision ε and failure probability δ. Specifically, thanks to Theorem 65, a number

$$N_{\mathrm{cov}}(n, \varepsilon, \delta, E_2) = O\left(\log\left(\frac{n^2}{\delta}\right) \frac{n^3 E_2^2}{\varepsilon^2}\right) \tag{2.388}$$

of copies of ρ are sufficient to build a covariance matrix $\tilde{V}$ and a vector $\tilde{m}$ such that

$$\Pr\left(\|\tilde{V}' - V(\rho)\|_\infty \leq \varepsilon \quad \text{and} \quad \|\tilde{\mathbf{m}} - \mathbf{m}(\rho)\|_2 \leq \varepsilon\right) \geq 1 - \delta\,. \tag{2.389}$$

Furthermore, we denote as $N_{\mathrm{tom,CV}}(n, \varepsilon, \delta, E_1)$ the sample complexity of the full state tomography algorithm outlined in Table 2.1 and Theorem 48 to learn an unknown n-mode pure state ψ, subject to the energy constraint $\mathrm{Tr}[\psi \hat{E}_n] \leq n E_1$, with accuracy ε in trace distance and failure probability δ. Specifically, Theorem 48 implies that

$$N_{\mathrm{tom,CV}}(n, \varepsilon, \delta, E_1) = \left\lceil 2^{21} \frac{d_{\mathrm{eff}}(n, \varepsilon, E_1)}{\varepsilon^2} \log\left(\frac{4}{\delta}\right)\right\rceil = O\left(\frac{E_1}{\varepsilon^2}\right)^n \tag{2.390}$$

copies of ψ are sufficient to generate a classical representation (i.e., $d_{\mathrm{eff}}(n, \varepsilon, E_1)$-sized vector) of a pure state $\tilde{\psi}$ such that

$$\Pr\left[\frac{1}{2}\|\psi - \tilde{\psi}\|_1 \leq \varepsilon\right] \geq 1 - \delta, \tag{2.391}$$

where $d_{\mathrm{eff}}(n, \varepsilon, E_1) \leq \left(\frac{eE_1}{\varepsilon^2} + 2e\right)^n$.

Theorem 86 (Number of copies required)**.** *Let $n, t \in \mathbb{N}$ with $t \leq n$, and $E \geq 0$. Let $|\psi\rangle$ be an n-mode t-compressible state satisfying the second-moment constraint $\mathrm{Tr}[|\psi\rangle\langle\psi| \hat{E}_n^2] \leq n^2 E^2$. Let $\varepsilon, \delta \in (0, 1)$, and N such that*

$$N := N_1 + \left\lceil 2N_2 + 24 \log\left(\frac{3}{\delta}\right)\right\rceil = O\left(\frac{n^9 E^6}{\varepsilon^4}\right) + O\left(\left(\frac{nE^2}{\varepsilon^2}\right)^t\right), \tag{2.395}$$

where

$$N_1 := N_{\mathrm{cov}}\left(n, \frac{\varepsilon^2}{2(n+1)(1+4nE)^2}, \frac{\delta}{3}, E\right) \tag{2.396}$$

is reported in (2.388) and $N_2 := N_{\mathrm{tom,CV}}\left(t, \frac{\varepsilon}{2}, \frac{\delta}{3}, 80n^2E^2\right)$ is reported in (2.390). Then, N copies of $|\psi\rangle$ are sufficient to build a classical description of a t-compressible pure state vector $|\hat{\psi}\rangle$ such that

$$\Pr\left(\frac{1}{2}\left\||\hat{\psi}\rangle\langle\hat{\psi}| - |\psi\rangle\langle\psi|\right\|_1 \leq \varepsilon\right) \geq 1 - \delta\,. \tag{2.397}$$

Such a classical description consists of the triplet $\left(\tilde{\boldsymbol{m}}, \tilde{S}, |\tilde{\phi}_1\rangle\right)$, which defines $|\hat{\psi}\rangle$ as

$$|\hat{\psi}\rangle := \hat{D}_{\tilde{\boldsymbol{m}}} U_{\tilde{S}} |\tilde{\phi}_1\rangle \otimes |0\rangle^{\otimes(n-t)}\,, \tag{2.398}$$

where $\tilde{\boldsymbol{m}} \in \mathbb{R}^{2n}$, $\tilde{S} \in \mathrm{Sp}(\mathbb{R}^{2n})$, and $|\tilde{\phi}_1\rangle$ is a t-mode pure state contained in a $\left\lceil\left(\frac{e(80n^2E^2 - \frac{1}{2})}{\varepsilon^2}\right)^t\right\rceil$-dimensional subspace.

Before proving Theorem 86 let us state some useful lemmas.

Lemma 87 (Bounds to symplectic matrices)**.** *Let $V(\rho)$ be a covariance matrix associated with an n-mode state ρ and let $V(\rho) = SDS^\intercal$ its symplectic diagonalisation. It holds that*

$$\|S\|_\infty \leq \sqrt{\|V(\rho)\|_\infty} \tag{2.399}$$

In particular, it holds that

$$\|V(\rho)\|_\infty \leq 4\,\mathrm{Tr}[\rho \hat{E}_n] \tag{2.400}$$

and consequently $\|S\|_\infty \leq \sqrt{4\,\mathrm{Tr}[\rho \hat{E}_n]}$.

Proof. Since $D \geq \mathbb{1}$, we have that $V(\rho) \geq SS^\intercal$. Consequently, it holds that

$$\begin{aligned} \|S\|_\infty^2 &= \|SS^\intercal\|_\infty \\ &\leq \|V(\rho)\|_\infty \\ &\leq \mathrm{Tr}\, V(\rho) \\ &= 4\,\mathrm{Tr}[\rho \hat{E}_n] - 2\|\mathbf{m}(\rho)\|_2^2 \\ &\leq 4\,\mathrm{Tr}[\rho \hat{E}_n]\,, \end{aligned} \tag{2.401}$$

where we have exploited Lemma 60. This implies $\|S\|_\infty \leq \sqrt{4\,\mathrm{Tr}[\rho \hat{E}_n]}$, where $\hat{E}_n \coloneqq \frac{1}{2}\hat{\mathbf{R}}^\intercal \hat{\mathbf{R}}$ is the energy operator. □

Lemma 88 (Perturbation on symplectic diagonalisation [92])**.** *Let $V_1, V_2 \in \mathbb{R}^{2n\times 2n}$ be two covariance matrices with symplectic diagonalisations $V_1 = S_1 D_1 S_1^\intercal$ and $V_2 = S_2 D_2 S_2^\intercal$, where the elements on the diagonal of D_1 and D_2 are arranged in descending order. Then*

$$\|D_1 - D_2\|_\infty \leq \sqrt{K(V_1)K(V_2)}\|V_1 - V_2\|_\infty\,, \tag{2.402}$$

where $K(V)$ is the condition number of the covariance matrix V, defined as $K(V) \coloneqq \|V\|_\infty \|V^{-1}\|_\infty$.

Lemma 89 Inverse of a covariance matrix**.** *Let V be a covariance matrix. The inverse of the covariance matrix V^{-1} satisfies*

$$\|V^{-1}\| \leq \|V\| \tag{2.403}$$

where $\|\cdot\|$ is an orthogonal invariant matrix norm. As a consequence, the condition number of V, which is defined as $K(V) \coloneqq \|V\|_\infty \|V^{-1}\|_\infty$, can be upper bounded as

$$K(V) \leq \|V\|_\infty^2\,. \tag{2.404}$$

Proof. Let $V = SDS^\intercal$ be the symplectic diagonalisation of V with S symplectic and D diagonal of the form $D = \mathrm{diag}(d_1, d_1, d_2, d_2, \ldots, d_n, d_n)$. In particular, we recall that $S\Omega_n S^\intercal = \Omega_n$ and that D commutes with Ω_n. Then it holds that

$$\begin{aligned} V^{-1} &= (S^{-1})^\intercal D^{-1} S^{-1} \\ &= \Omega_n\Omega_n (S^{-1})^\intercal D^{-1} S^{-1} \Omega_n \Omega_n \\ &= \Omega_n S \Omega_n D^{-1} \Omega_n S^\intercal \Omega_n \\ &= -\Omega_n S D^{-1} S^\intercal \Omega_n \\ &= \Omega_n S D^{-1} (\Omega_n S)^\intercal\,. \end{aligned} \tag{2.405}$$

Since V is a covariance matrix, then $D \geq \mathbb{1}$ and thus $D^{-1} \leq D$. Consequently, the matrix inequality

$$\begin{aligned} V^{-1} &\leq \Omega_n S D (\Omega_n S)^\intercal \\ &= \Omega_n V \Omega_n^\intercal \end{aligned} \tag{2.406}$$

holds. Since Ω_n is orthogonal, then for any orthogonal invariant norm $\|\cdot\|$ it holds that $\|V^{-1}\| \leq \|V\|$[75], that concludes the proof. □

Lemma 90 (Upper bound to mean energy)**.** *Let ρ be an n-mode state. Let $S \in \mathrm{Sp}(2n)$ be a symplectic matrix. Let U_S be the Gaussian symplectic unitary associated with S. Then, the mean energy of $U_S \rho U_S^\dagger$ can be upper bounded as*

$$E\big(U_S \rho U_S^\dagger\big) \leq \|S\|_\infty^2 E(\rho)\,, \tag{2.407}$$

where $E(\rho) \coloneqq \mathrm{Tr}\Big[\frac{\hat{\boldsymbol{R}}^\intercal \hat{\boldsymbol{R}}}{2}\,\rho\Big]$.

Proof. It holds that

$$\begin{aligned} E\big(U_S \rho U_S^\dagger\big) &= \frac{1}{2}\,\mathrm{Tr}\big[U_S \rho U_S^\dagger \hat{\mathbf{R}}^\intercal \hat{\mathbf{R}}\big] \\ &\overset{\text{(i)}}{=} \frac{1}{2}\,\mathrm{Tr}\big[\rho \hat{\mathbf{R}}^\intercal S^\intercal S \hat{\mathbf{R}}\big] \\ &\overset{\text{(ii)}}{=} \frac{1}{2}\,\mathrm{Tr}\big[\rho \hat{\mathbf{R}}^\intercal O_1^\intercal Z^2 O_1 \hat{\mathbf{R}}\big] \\ &\overset{\text{(iii)}}{=} \frac{1}{2}\,\mathrm{Tr}\big[U_{O_1} \rho U_{O_1}^\dagger \hat{\mathbf{R}}^\intercal Z^2 \hat{\mathbf{R}}\big] \\ &\overset{\text{(iv)}}{\leq} \|S\|_\infty^2\,\mathrm{Tr}\big[U_{O_1} \rho U_{O_1}^\dagger \hat{E}\big] \\ &\overset{\text{(v)}}{=} \|S\|_\infty^2\,\mathrm{Tr}\big[\rho \hat{E}\big]\,. \end{aligned} \tag{2.408}$$

Here, in (i), we have used that $U_S^\dagger \hat{\mathbf{R}} U_S = S\mathbf{R}$. In (ii), we have exploited the Euler decomposition $S = O_1 Z O_2$ given in (2.97). In (iii), we have used $U_{O_1}^\dagger \hat{\mathbf{R}} U_{O_1} = O_1 \mathbf{R}$. In (iv), we first used that $\hat{\mathbf{R}}^\intercal Z^2 \hat{\mathbf{R}} = \sum_{i=1}^{2n} z_i^2 \hat{R}_i^2 \leq \|Z\|_\infty^2 \hat{\mathbf{R}}^\intercal \hat{\mathbf{R}}$, and, second, we have exploited that $\|S\|_\infty^2 = \|S^\intercal S\|_\infty = \|O_2^\intercal Z^2 O_2\|_\infty = \|Z\|_\infty^2$ to obtain that $\hat{\mathbf{R}}^\intercal Z^2 \hat{\mathbf{R}} \leq 2\|S\|_\infty^2 \hat{E}$. Finally, in (v) we have used again $U_{O_1}^\dagger \hat{\mathbf{R}} U_{O_1} = O_1 \mathbf{R}$ and the fact that O_1 is orthogonal. □

Lemma 91 (Refined mean energy bound)**.** *Let ρ be an n-mode state. Let $\mathbf{r} \in \mathbb{R}^{2n}$ and let $\hat{D}_{\mathbf{r}}$ be the associated displacement operator. Then, the mean energy of $\hat{D}_{\mathbf{r}} \rho \hat{D}_{\mathbf{r}}^\dagger$ reads*

$$E\big(\hat{D}_{\mathbf{r}} \rho \hat{D}_{\mathbf{r}}^\dagger\big) = E(\rho) + \mathbf{r}^\intercal \mathbf{m}(\rho) + \frac{1}{2}\|\mathbf{r}\|_2^2\,, \tag{2.409}$$

where $E(\rho) \coloneqq \mathrm{Tr}\Big[\frac{\hat{\boldsymbol{R}}^\intercal \hat{\boldsymbol{R}}}{2}\,\rho\Big]$. In particular, it can be upper bounded as

$$E\big(\hat{D}_{\mathbf{r}} \rho \hat{D}_{\mathbf{r}}^\dagger\big) \leq E(\rho) + \sqrt{2E(\rho)}\|\boldsymbol{r}\|_2 + \frac{1}{2}\|\boldsymbol{r}\|_2^2\,. \tag{2.410}$$

Proof. It holds that

$$\begin{aligned} E\big(\hat{D}_{\mathbf{r}} \rho \hat{D}_{\mathbf{r}}^\dagger\big) &= \frac{1}{2}\,\mathrm{Tr}[\rho(\hat{\mathbf{R}} + \mathbf{r}\mathbb{1})^\intercal(\hat{\mathbf{R}} + \mathbf{r}\mathbb{1})] \\ &= E(\rho) + \mathbf{r}^\intercal \mathbf{m}(\rho) + \frac{1}{2}\|\mathbf{r}\|_2^2\,. \end{aligned} \tag{2.411}$$

Moreover, note that

$$m_i(\rho) = \mathrm{Tr}[\rho R_i] \leq \sqrt{\mathrm{Tr}[\rho R_i^2]} \tag{2.412}$$

for each $i \in [2n]$, where we have exploited (2.41). Consequently, it holds that

$$\|\mathbf{m}(\rho)\|_2 \leq \sqrt{\sum_{i=1}^{n} \mathrm{Tr}[\rho R_i^2]} = \sqrt{2E(\rho)}\,. \tag{2.413}$$

Thus, it follows that

$$\begin{aligned} E\big(\hat{D}_{\mathbf{r}} \rho \hat{D}_{\mathbf{r}}^\dagger\big) &\leq E(\rho) + \|\mathbf{r}\|_2 \|\mathbf{m}(\rho)\|_2 + \frac{1}{2}\|\mathbf{r}\|_2^2 \\ &\leq E(\rho) + \sqrt{2E(\rho)}\|\mathbf{r}\|_2 + \frac{1}{2}\|\mathbf{r}\|_2^2\,. \end{aligned} \tag{2.414}$$

□

Now, we are ready to prove Theorem 86.

Proof of Theorem 86. We aim to establish the correctness of the algorithm presented in Table 2.4.

In Line 1, we use $N_{\mathrm{cov}}\big(n, \varepsilon_{\mathrm{cov}}, \frac{\delta}{3}, E\big)$ copies of $|\psi\rangle$ in order to build a vector $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$ and a covariance matrix $\tilde{V} \in \mathbb{R}^{2n,2n}$ such that

$$\Pr\Big(\|\tilde{V} - V(|\psi\rangle\langle\psi|)\|_\infty \leq \varepsilon_{\mathrm{cov}} \quad \text{and} \quad \|\tilde{\mathbf{m}} - \mathbf{m}(|\psi\rangle\langle\psi|)\|_2 \leq \varepsilon_{\mathrm{cov}}\Big) \geq 1 - \frac{\delta}{3}\,, \tag{2.415}$$

where $N_{\mathrm{cov}}(\cdot)$ is defined in (2.388) and the accuracy $\varepsilon_{\mathrm{cov}}$ will be fixed later. From now on, let us assume that we are in the probability event in which

$$\|\tilde{V} - V(|\psi\rangle\langle\psi|)\|_\infty \leq \varepsilon_{\mathrm{cov}} \quad \text{and} \quad \|\tilde{\mathbf{m}} - \mathbf{m}(|\psi\rangle\langle\psi|)\|_2 \leq \varepsilon_{\mathrm{cov}}\,. \tag{2.416}$$

In Line 2, we compute the symplectic diagonalisation $\tilde{V} = \tilde{S}\tilde{D}\tilde{S}^\intercal$, where $\tilde{S} \in \mathrm{Sp}(2n)$ and

$$\tilde{D} := \mathrm{diag}\left(\tilde{d}_1, \tilde{d}_1, \tilde{d}_2, \tilde{d}_2, \ldots, \tilde{d}_n, \tilde{d}_n\right) \tag{2.417}$$

with $\tilde{d}_1 \geq \tilde{d}_2 \geq \ldots \geq \tilde{d}_n \geq 1$.

The sequence of steps in Lines 4, 5, and 6 have to be repeated a number

$$\left\lceil 2N_{\mathrm{tom,CV}}\left(t, \varepsilon_{\mathrm{tom}}, \frac{\delta}{3}, E\right) + 24\log\left(\frac{3}{\delta}\right)\right\rceil \tag{2.418}$$

of times, where the quantity $N_{\mathrm{tom,CV}}\big(t, \varepsilon_{\mathrm{tom}}, \frac{\delta}{3}, E\big)$ is the number of copies sufficient for tomography of a t-mode pure state with mean energy per mode at most E with accuracy $\varepsilon_{\mathrm{tom}}$ and failure probability $\frac{\delta}{3}$ (see (2.390)). The accuracy $\varepsilon_{\mathrm{tom}}$ and the energy E will be fixed later.

In Line 4, we query one copy of $|\psi\rangle$ and apply $\hat{D}^\dagger_{\tilde{\mathbf{m}}} U^\dagger_{\tilde{S}}$, obtaining the state vector $|\tilde{\psi}_t\rangle := \hat{D}^\dagger_{\tilde{\mathbf{m}}} U^\dagger_{\tilde{S}} |\psi\rangle$. Since $|\psi\rangle$ is t-compressible and since it holds that $\tilde{\mathbf{m}} \simeq \mathbf{m}(|\psi\rangle\langle\psi|)$ and $\tilde{V} = \tilde{S}\tilde{D}\tilde{S}^\intercal \simeq V(|\psi\rangle\langle\psi|)$, we intuitively expect that the reduced state of $|\tilde{\psi}_t\rangle$ onto the last $n - t$ modes is very close to the vacuum $|0\rangle^{\otimes(n-t)}$ (as we are going to rigorously show).

In Line 5, we measure $|\tilde{\psi}_t\rangle$ with respect the POVM $\{\mathbb{1}_t \otimes |0\rangle\langle 0|^{\otimes(n-t)}, \mathbb{1} - \mathbb{1}_t \otimes |0\rangle\langle 0|^{\otimes(n-t)}\}$, and discard the copies associated with the outcome corresponding to the POVM element $\mathbb{1} - \mathbb{1}_t \otimes |0\rangle\langle 0|^{\otimes(n-t)}$. The post-measurement state is not discarded with probability

$$P_{\mathrm{succ}} := \mathrm{Tr}\Big[\mathbb{1}_t \otimes |0\rangle\langle 0|^{\otimes(n-t)}\ |\tilde{\psi}_t\rangle\langle\tilde{\psi}_t|\Big]\,, \tag{2.419}$$

and the remaining copies are thus in a state of the form $|\tilde{\phi}\rangle \otimes |0\rangle^{\otimes(n-t)}$, where $|\tilde{\phi}\rangle$ is given by

$$|\tilde{\phi}\rangle := \frac{\langle 0|^{\otimes(n-t)}\,|\tilde{\psi}_t\rangle}{\sqrt{P_{\mathrm{succ}}}}\,. \tag{2.420}$$

Let us now find a lower bound on the probability of success P_{succ}. For simplicity in the following we will use the notation $\psi := |\psi\rangle\langle\psi|$, $\tilde{\psi}_t := |\psi\rangle\langle\psi|$ and $\tilde{\phi} := |\tilde{\phi}\rangle\langle\tilde{\phi}|$. Note that

$$\begin{aligned}
P_{\mathrm{succ}} &= \mathrm{Tr}[\mathbb{1}_t \otimes |0\rangle\langle 0|^{\otimes(n-t)}\ \tilde{\psi}_t] \\
&= \mathrm{Tr}[|0\rangle\langle 0|^{\otimes(n-t)}\ \mathrm{Tr}_{[t]}\,\tilde{\psi}_t] \\
&\overset{\text{(i)}}{\geq} \mathrm{Tr}\left[\left(\mathbb{1} - \sum_{i=t+1}^{n} a_i^\dagger a_i\right)\ \mathrm{Tr}_{[t]}\,\tilde{\psi}_t\right] \\
&\overset{\text{(ii)}}{=} 1 - \frac{\mathrm{Tr}[V(\mathrm{Tr}_{[t]}\,\tilde{\psi}_t) - I_{n-t}]}{4} - \frac{\|\mathbf{m}(\mathrm{Tr}_{[t]}\,\tilde{\psi}_t)\|_2^2}{2} \\
&\geq 1 - \frac{2(n-t)\|V(\mathrm{Tr}_{[t]}\,\tilde{\psi}_t) - I_{n-t}\|_\infty}{4} - \frac{\|\mathbf{m}(\tilde{\psi}_t)\|_2^2}{2}\,,
\end{aligned} \tag{2.421}$$

Here, in (i), we have used the following operator inequality

$$\mathbb{1}_t \otimes |0\rangle\langle 0|^{\otimes(n-t)} \geq \mathbb{1} - \sum_{i=t+1}^{n} a_i^\dagger a_i \,. \tag{2.422}$$

which can be readily verified. In (ii), we have exploited Lemma 60. Moreover, note that

$$\begin{aligned} V(\tilde{\psi}_t) &= V\left(U_{\tilde{S}}^\dagger \hat{D}_{\tilde{\mathbf{m}}}^\dagger \psi \hat{D}_{\tilde{\mathbf{m}}} U_{\tilde{S}}\right) \\ &= \tilde{S}^{-1} V(\psi) (\tilde{S}^{-1})^\intercal \\ &= \tilde{S}^{-1} \left(V(\psi) - \tilde{V}\right) (\tilde{S}^{-1})^\intercal + \tilde{D} \,. \end{aligned} \tag{2.423}$$

Since ψ is t-compressible, Lemma 82 establishes that $|\psi\rangle$ can be written as

$$|\psi\rangle = \hat{D}_{\mathbf{m}(\psi)} U_S \, |\phi_t\rangle \otimes |0\rangle^{\otimes(n-t)} \,, \tag{2.424}$$

where $|\phi_t\rangle$ is some t-mode state, S is a symplectic matrix such that $V(\psi) = S \left(D_t \oplus \mathbb{1}_{n-t}\right) S^\intercal$, and D_t is $2t \times 2t$ diagonal matrix of the form

$$D_t \coloneqq \operatorname{diag}\left(d_1, d_1, d_2, d_2, \ldots, d_n, d_n\right) \tag{2.425}$$

with $d_1 \geq d_2 \geq \ldots \geq d_n \geq 1$. Let us analyse the first term in the r.h.s. of (2.421). It holds that

$$\begin{aligned} \|V(\operatorname{Tr}_{[t]} \tilde{\psi}_t) - I_{n-t}\|_\infty &\overset{\text{(iii)}}{=} \left\| \left[\tilde{S}^{-1} \left(V(\psi) - \tilde{V}\right) (\tilde{S}^{-1})^\intercal\right]_{n-t} + [\tilde{D}]_{n-t} - I_{n-t} \right\|_\infty \\ &\leq \left\| \left[\tilde{S}^{-1} \left(V(\psi) - \tilde{V}\right) (\tilde{S}^{-1})^\intercal\right]_{n-t} \right\|_\infty + \left\| [\tilde{D}]_{n-t} - I_{n-t} \right\|_\infty \\ &\overset{\text{(iv)}}{\leq} \left\| \tilde{S}^{-1} \left(V(\psi) - \tilde{V}\right) (\tilde{S}^{-1})^\intercal \right\|_\infty + \left\| \tilde{D} - D \right\|_\infty \\ &\overset{\text{(v)}}{\leq} \|\tilde{S}\|_\infty^2 \|V(\psi) - \tilde{V}\|_\infty + \left\| \tilde{D} - D \right\|_\infty \\ &\overset{\text{(vi)}}{\leq} \|\tilde{V}\|_\infty \varepsilon_{\text{cov}} + \left\| \tilde{D} - D \right\|_\infty \\ &\overset{\text{(vii)}}{\leq} (\varepsilon_{\text{cov}} + \|V(\psi)\|_\infty)\varepsilon_{\text{cov}} + \sqrt{K(V(\psi))K(\tilde{V})} \|\tilde{V} - V(\psi)\|_\infty \\ &\overset{\text{(viii)}}{\leq} \left(\varepsilon_{\text{cov}} + 4 \operatorname{Tr}[\psi \hat{E}_n]\right) \varepsilon_{\text{cov}} + \|V(\psi)\|_\infty \|\tilde{V}\|_\infty \varepsilon_{\text{cov}} \\ &\overset{\text{(ix)}}{\leq} \left(\varepsilon_{\text{cov}} + 4 \operatorname{Tr}[\psi \hat{E}_n]\right) \varepsilon_{\text{cov}} + 4 \operatorname{Tr}[\psi \hat{E}_n] \left(4 \operatorname{Tr}[\psi \hat{E}_n] + \varepsilon_{\text{cov}}\right) \varepsilon_{\text{cov}} \\ &\overset{\text{(x)}}{\leq} (\varepsilon_{\text{cov}} + 4nE)\,(1 + 4nE)\,\varepsilon_{\text{cov}} \\ &\leq (1 + 4nE)^2 \,\varepsilon_{\text{cov}} \,. \end{aligned} \tag{2.426}$$

Here, in (iii), we have exploited (2.426) and we have introduced the notation $[A]_k$ to denote the $k \times k$ lower-right block of the matrix A. In (iv), first, we have used that for any $n \times n$ matrix A and any $k \leq n$ it holds that $\|[A]_k\|_\infty \leq \|A\|_\infty$, and, second, we have exploited that $[D]_{n-t} = I_{n-t}$. In (v), we first used the submultiplicativity of the operator norm, and second, we applied Lemma 63, which implies that $\|\tilde{S}^{-1}\|_\infty = \|\tilde{S}\|_\infty$. In (vi), we have used that $\|\tilde{S}\|_\infty \leq \sqrt{\|\tilde{V}\|_\infty}$, as established by Lemma 87, and that $\|V(\psi) - \tilde{V}\|_\infty \leq \varepsilon_V$. In (vii), first we have used that $\|\tilde{V}\|_\infty \leq \|V(\psi)\|_\infty + \varepsilon_V$, and second, we have exploited Lemma 88 that establishes that the maximum difference between the ordered symplectic eigenvalues $\|\tilde{D} - D\|_\infty$ can be upper bounded in terms of the *condition numbers* $K(V(\psi)), K(\tilde{V})$ (see Lemma 88) and in terms of $\|V(\psi) - \tilde{V}\|_\infty$. In (viii), we have used that $K(V) \leq \|V\|_\infty^2$, as stated by Lemma 89. In (ix), we have exploited that $\|V(\psi)\|_\infty \leq 4 \operatorname{Tr}[\psi \hat{E}_n]$, as

established by Lemma 87. Finally, in (x), we have used Cauchy Schwarz inequality to deduce that $\mathrm{Tr}[\psi \hat{E}_n] \leq \sqrt{\mathrm{Tr}[\psi \hat{E}_n^2]}$ and then we have used the assumption that $\sqrt{\mathrm{Tr}[\psi \hat{E}_n^2]} \leq nE$.

Now, let us analyse the second term in the r.h.s. of (2.421). With similar techniques of (2.426), we can upper bound $\|\mathbf{m}(\tilde{\psi}_t)\|_2$ as

$$\begin{aligned}
\|\mathbf{m}(\tilde{\psi}_t)\|_2 &= \left\|\mathbf{m}\Big(U_{\tilde{S}}^\dagger \hat{D}_{\tilde{\mathbf{m}}}^\dagger \psi \hat{D}_{\tilde{\mathbf{m}}} U_{\tilde{S}}\Big)\right\|_2 \\
&= \left\|\tilde{S}^{-1}\left(\mathbf{m}(\psi) - \tilde{\mathbf{m}}\right)\right\|_2 \\
&\leq \|\tilde{S}^{-1}\|_\infty \|\mathbf{m}(\psi) - \tilde{\mathbf{m}}\|_2 \\
&\leq \|\tilde{S}\|_\infty \, \varepsilon_{\mathrm{cov}} \\
&\leq \sqrt{\|\tilde{V}\|_\infty} \varepsilon_{\mathrm{cov}} \\
&\leq \sqrt{4nE + \varepsilon_{\mathrm{cov}}} \, \varepsilon_{\mathrm{cov}} \\
&\leq \sqrt{4nE + 1} \, \varepsilon_{\mathrm{cov}} \, .
\end{aligned} \tag{2.427}$$

By putting (2.421), (2.426), and (2.427) together, we deduce that P_{succ} is lower bounded by

$$\begin{aligned}
P_{\mathrm{succ}} &\geq 1 - \frac{2(n-t)\|V(\mathrm{Tr}_{[t]}\, \tilde{\psi}_t) - \mathbb{1}\|_\infty}{4} - \frac{\|\mathbf{m}(\tilde{\psi}_t)\|_2^2}{2} \\
&\geq 1 - \frac{n-t}{2}(1+4nE)^2 \varepsilon_{\mathrm{cov}} - \frac{1}{2}(1+4nE)\varepsilon_{\mathrm{cov}}^2 \\
&\geq 1 - \frac{(1+4nE)^2}{2}(n\, \varepsilon_{\mathrm{cov}} + \varepsilon_{\mathrm{cov}}^2) \\
&\geq 1 - \frac{(1+4nE)^2 (n+1)}{2} \varepsilon_{\mathrm{cov}} \, .
\end{aligned} \tag{2.428}$$

The choice of $\varepsilon_{\mathrm{cov}}$, which we will make later in (2.436), implies that $P_{\mathrm{succ}} \geq \frac{3}{4}$. Consequently, thanks Lemma 17, since we conduct in total $\left\lceil 2N_{\mathrm{tom,CV}}\left(t, \varepsilon_{\mathrm{tom}}, \frac{\delta}{3}, E\right) + 24 \log\left(\frac{3}{\delta}\right)\right\rceil$ measurements, each of which with a success probability of $P_{\mathrm{succ}} \geq \frac{3}{4}$, the probability of obtaining at least $N_{\mathrm{tom,CV}}\left(t, \varepsilon_{\mathrm{tom}}, \frac{\delta}{3}, E\right)$ successful outcomes is $\geq 1 - \frac{\delta}{3}$. In other words, the probability of getting at least $N_{\mathrm{tom,CV}}\left(t, \varepsilon_{\mathrm{tom}}, \frac{\delta}{3}, E\right)$ post-measurement states $|\tilde{\phi}\rangle \otimes |0\rangle^{\otimes(n-t)}$ is $\geq 1 - \frac{\delta}{3}$. From now on, let us assume that we are in such a probability event. In Line 6, we do a step of the algorithm in Table 2.1 to perform pure-state tomography of the energy-constrained t-mode state vector $|\tilde{\phi}\rangle$, by considering the first t modes of the post-measurement state vector $|\tilde{\phi}\rangle \otimes |0\rangle^{\otimes(n-t)}$. In order to apply such an algorithm we need to find an upper bound on the mean energy of $|\tilde{\phi}\rangle$. By denoting

the mean energy of a state ψ as $E(\psi)$, let us note that

$$\begin{aligned}
E(\tilde{\phi}) &\overset{\text{(i)}}{\leq} \frac{1}{P_{\text{succ}}} E(\tilde{\psi}_t) \\
&\overset{\text{(ii)}}{\leq} \frac{\|\tilde{S}\|_\infty^2}{P_{\text{succ}}} E\Big(\hat{D}_{\tilde{\mathbf{m}}}^\dagger \left|\psi\right\rangle\!\left\langle\psi\right| \hat{D}_{\tilde{\mathbf{m}}}\Big) \\
&\overset{\text{(iii)}}{\leq} \frac{\|\tilde{S}\|_\infty^2}{P_{\text{succ}}} \left[E(\psi) + \sqrt{2E(\psi)}\|\tilde{\mathbf{m}}\|_2 + \frac{1}{2}\|\tilde{\mathbf{m}}\|_2^2\right] \\
&\overset{\text{(iv)}}{\leq} \frac{\|\tilde{S}\|_\infty^2}{P_{\text{succ}}} \left[E(\psi) + \sqrt{2E(\psi)}(\|\mathbf{m}(\psi)\|_2 + \varepsilon_m) + \frac{1}{2}(\|\mathbf{m}(\psi)\|_2 + \varepsilon_m)^2\right] \\
&\overset{\text{(v)}}{\leq} \frac{4E(\psi) + \varepsilon_V}{P_{\text{succ}}} \left[E(\psi) + \sqrt{2E(\psi)}(\sqrt{2E(\psi)} + \varepsilon_m) + \frac{1}{2}(\sqrt{2E(\psi)} + \varepsilon_m)^2\right] \\
&\overset{\text{(vi)}}{\leq} \frac{4}{3}[4E(\psi) + \varepsilon_V] \left[4E(\psi) + 2\sqrt{2E(\psi)}\varepsilon_m + \frac{\varepsilon_m^2}{2}\right] \\
&\overset{\text{(vii)}}{\leq} 80\,[E(\psi)]^2 \\
&\leq 80\ \mathrm{Tr}[\psi \hat{E}_n^2] \\
&\leq 80\, E^2 n^2.
\end{aligned} \tag{2.429}$$

Here, in (i), we have used that

$$\begin{aligned}
E(\tilde{\phi}) &= \frac{1}{P_{\text{succ}}} E\Big(\left\langle 0\right|^{\otimes(n-t)} \tilde{\psi}_t \left|0\right\rangle^{\otimes(n-t)}\Big) \\
&= \frac{1}{2P_{\text{succ}}} \mathrm{Tr}\left[\sum_{i=1}^t R_i^2 \left\langle 0\right|^{\otimes(n-t)} \tilde{\psi}_t \left|0\right\rangle^{\otimes(n-t)}\right] \\
&= \frac{1}{2P_{\text{succ}}} \mathrm{Tr}\left[\left(\sum_{i=1}^t R_i^2\right) \otimes \left|0\right\rangle\!\left\langle 0\right|^{\otimes(n-t)} \tilde{\psi}_t\right] \\
&\leq \frac{1}{2P_{\text{succ}}} \mathrm{Tr}\left[\left(\sum_{i=1}^t R_i^2\right) \otimes \mathbb{1}_{n-t}\, \tilde{\psi}_t\right] \\
&= \frac{1}{2P_{\text{succ}}} \mathrm{Tr}\left[\left(\sum_{i=1}^t R_i^2\right) \tilde{\psi}_t\right] \\
&\leq \frac{1}{2P_{\text{succ}}} \mathrm{Tr}\left[\left(\sum_{i=1}^n R_i^2\right) \tilde{\psi}_t\right] \\
&= \frac{E(\tilde{\psi}_t)}{P_{\text{succ}}}.
\end{aligned} \tag{2.430}$$

In (ii), first we have used that $\tilde{\psi} = U_{\tilde{S}}^\dagger \hat{D}_{\tilde{\mathbf{m}}}^\dagger \psi \hat{D}_{\tilde{\mathbf{m}}} U_{\tilde{S}}$, and then we have used Lemma 90. In (iii), we have used Lemma 91. Then, in (iv), we have used triangular inequality and the bound $\|\mathbf{m}(\psi) - \tilde{\mathbf{m}}\|_2 \leq \varepsilon_{\text{cov}}$. In (v), we have used Lemma 87, together with $\|\tilde{V} - V\|_\infty \leq \varepsilon_{\text{cov}}$ and (2.413). In (vi), we have exploited that $P_{\text{succ}} \geq \frac{3}{4}$. In (vii), we have exploited that $\varepsilon_{\text{cov}} \leq 1$ and $E(\psi) \geq \frac{1}{2}$. Consequently, the mean energy per mode of the state $\tilde{\phi}$ is upper bounded by

$$\frac{E(\tilde{\phi})}{t} \leq \frac{80\, E^2 n^2}{t} \leq 80\, E^2 n^2\,, \tag{2.431}$$

and hence we can set $E := 80\, E^2 n^2$.

In Line 8, the tomography algorithm called in the iterations of Line 6 outputs a t-mode state vector $|\tilde{\phi}_1\rangle$, which is supported on a $\left\lceil \left(\frac{e(80n^2E^2 - \frac{1}{2})}{\varepsilon^2}\right)^t \right\rceil$-dimensional subspace of $L^2(\mathbb{R}^t)$. Since the

total number of copies of $|\tilde{\phi}\rangle$ employed in such a tomography algorithm is $\geq N_{\mathrm{tom,CV}}\Big(t, \varepsilon_{\mathrm{tom}}, \frac{\delta}{3}, E\Big)$, the state vector $|\tilde{\phi}_1\rangle$ satisfies

$$\Pr\left(\frac{1}{2}\left\||\tilde{\phi}\rangle\langle\tilde{\phi}| - |\tilde{\phi}_1\rangle\langle\tilde{\phi}_1|\right\|_1 < \varepsilon_{\mathrm{tom}}\right) \geq 1 - \frac{\delta}{3}\,. \tag{2.432}$$

From now on, let us assume that we are in the probability event in which

$$\frac{1}{2}\left\||\tilde{\phi}\rangle\langle\tilde{\phi}| - |\tilde{\phi}_1\rangle\langle\tilde{\phi}_1|\right\|_1 < \varepsilon_{\mathrm{tom}}\,. \tag{2.433}$$

In Line 9, we output a classical description of the n-mode state vector $|\hat{\psi}\rangle$, which is defined as

$$|\hat{\psi}\rangle := \hat{D}_{\tilde{\mathbf{m}}} U_{\tilde{S}} \left(|\tilde{\phi}_1\rangle \otimes |0\rangle^{\otimes(n-t)}\right)\,. \tag{2.434}$$

Note that

$$\begin{aligned}
\frac{1}{2}\left\||\hat{\psi}\rangle\langle\hat{\psi}| - |\psi\rangle\langle\psi|\right\|_1 &\overset{\text{(i)}}{=} \frac{1}{2}\left\||\tilde{\phi}_1\rangle\langle\tilde{\phi}_1| \otimes |0\rangle\langle 0|^{\otimes(n-t)} - |\tilde{\psi}_t\rangle\langle\tilde{\psi}_t|\right\|_1 \\
&\overset{\text{(ii)}}{\leq} \frac{1}{2}\left\||\tilde{\phi}_1\rangle\langle\tilde{\phi}_1| - |\tilde{\phi}\rangle\langle\tilde{\phi}|\right\|_1 + \frac{1}{2}\left\||\tilde{\phi}\rangle\langle\tilde{\phi}| \otimes |0\rangle\langle 0|^{\otimes(n-t)} - |\tilde{\psi}_t\rangle\langle\tilde{\psi}_t|\right\|_1 \\
&\overset{\text{(iii)}}{\leq} \varepsilon_{\mathrm{tom}} + \sqrt{1 - P_{\mathrm{succ}}} \\
&\overset{\text{(iv)}}{\leq} \varepsilon_{\mathrm{tom}} + (1 + 4nE)\sqrt{\frac{n+1}{2}\varepsilon_{\mathrm{cov}}}\,.
\end{aligned} \tag{2.435}$$

Here, in (i), we have used unitarily invariance of the trace norm. In (ii), we have used the triangular inequality of the trace norm. In (iii), we have used (2.433) and the *gentle measurement lemma* [86]. Finally, in (iv) we have used (2.428). Consequently, by choosing

$$\begin{aligned}
\varepsilon_{\mathrm{cov}} &:= \frac{\varepsilon^2}{2(n+1)(1+4nE)^2}\,, \\
\varepsilon_{\mathrm{tom}} &:= \frac{\varepsilon}{2}\,,
\end{aligned} \tag{2.436}$$

we have that

$$\frac{1}{2}\left\||\hat{\psi}\rangle\langle\hat{\psi}| - |\psi\rangle\langle\psi|\right\|_1 \leq \varepsilon\,. \tag{2.437}$$

By using an union bound, we can conclude that a total number of

$$N_{\mathrm{cov}}\left(n, \varepsilon_{\mathrm{cov}}, \frac{\delta}{3}, E\right) + \left\lceil 2N_{\mathrm{tom,CV}}\left(t, \varepsilon_{\mathrm{tom}}, \frac{\delta}{3}, E\right) + 24\log\left(\frac{3}{\delta}\right)\right\rceil \tag{2.438}$$

copies of $|\psi\rangle$ suffices to construct a classical representation of a state vector $|\hat{\psi}\rangle$ such that

$$\Pr\left(\frac{1}{2}\left\||\hat{\psi}\rangle\langle\hat{\psi}| - |\psi\rangle\langle\psi|\right\|_1 \leq \varepsilon\right) \geq 1 - \delta\,. \tag{2.439}$$

□

We now show that any algorithm to learn t-compressible states should be inefficient if t scales faster than a constant in the number of modes.

Theorem 92 (Lower bound on sample complexity of tomography of t-compressible states)**.** *Let us consider a tomography algorithm that learns, within a trace distance $\leq \varepsilon$ and failure probability $\leq \delta$, an unknwon n-mode t-compressible state ψ satisfying the second-moment constraint*

$\sqrt{\mathrm{Tr}[\psi \hat{E}_n^2]} \leq nE$. *Then, such a tomography algorithm must use a number of state copies N satisfying*

$$\begin{aligned} N &\geq \frac{1}{t\, g\Big(\frac{n}{t}\Big(E-\frac{1}{2}\Big)\Big)} \left[2(1-\delta)\left(\frac{\frac{n}{t}\Big(E-\frac{1}{2}\Big)}{12\varepsilon}-\frac{1}{t}\right)^t - (1-\delta)\log_2(32\pi) - H_2(\delta)\right] \\ &= \Theta\Big(\frac{nE}{t\varepsilon}\Big)^t . \end{aligned} \tag{2.440}$$

Here, $g(x) \coloneqq (x+1)\log_2(x+1) - x\log_2 x$ is the bosonic entropy, and $H_2(x) \coloneqq -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy.

Proof. Since the tomography algorithm can learn arbitrary n-mode t-compressible states ψ satisfying the second-moment constraint $\sqrt{\mathrm{Tr}[\psi \hat{E}_n^2]} \leq nE$, then it can also learn *arbitrary* t-mode states $|\phi_t\rangle$ satisfying

$$\sqrt{\mathrm{Tr}[|\phi_t\rangle\langle\phi_t|\, \hat{N}_t^2]} \leq n\Big(E-\frac{1}{2}\Big) . \tag{2.441}$$

Indeed, $|\psi\rangle \coloneqq |\phi_t\rangle \otimes |0\rangle^{\otimes(n-t)}$ is a t-compressible state with second-moment upper bounded by

$$\begin{aligned} \sqrt{\mathrm{Tr}\Big[|\psi\rangle\langle\psi|\, \hat{E}_n^2\Big]} &= \sqrt{\mathrm{Tr}\Big[|\phi_t\rangle\langle\phi_t| \otimes |0\rangle\langle 0|^{\otimes(n-t)}\, \hat{E}_n^2\Big]} \\ &= \sqrt{\mathrm{Tr}\left[|\phi_t\rangle\langle\phi_t| \otimes |0\rangle\langle 0|^{\otimes(n-t)} \left(\hat{N}_n + \frac{n}{2}\mathbb{1}\right)^2\right]} \\ &= \sqrt{\mathrm{Tr}[|\phi_t\rangle\langle\phi_t|\, \hat{N}_t^2] + n\,\mathrm{Tr}[|\phi_t\rangle\langle\phi_t|\, \hat{N}_t] + \frac{n^2}{4}} \\ &\leq \sqrt{\mathrm{Tr}[|\phi_t\rangle\langle\phi_t|\, \hat{N}_t^2] + n\sqrt{\mathrm{Tr}[|\phi_t\rangle\langle\phi_t|\, \hat{N}_t^2]} + \frac{n^2}{4}} \\ &= \sqrt{\mathrm{Tr}[|\phi_t\rangle\langle\phi_t|\, \hat{N}_t^2]} + \frac{n}{2} \\ &\leq nE . \end{aligned} \tag{2.442}$$

Finally, Theorem 38 ensures that any algorithm that learns unknown t-mode states $|\phi_t\rangle$, satisfying the second-moment constraint

$$\frac{1}{t}\sqrt{\mathrm{Tr}[|\phi_t\rangle\langle\phi_t|\, \hat{N}_t^2]} \leq \frac{n}{t}\Big(E-\frac{1}{2}\Big) , \tag{2.443}$$

must use a number of state copies N that satisfies the condition in (2.440). □

Table 2.4: Tomography algorithm for bosonic t-compressible pure states

Input: Accuracy ε, failure probability δ, second moment upper bound E, N copies of the unknown n-mode t-compressible state vector $|\psi\rangle$ satisfying the 2-moment constraint $\mathrm{Tr}\left[\hat{E}_n^2 |\psi\rangle\langle\psi|\right] \leq n^2 E^2$, where

$$\begin{aligned} N &:= N_1 + \left\lceil 2N_2 + 24 \log\left(\frac{3}{\delta}\right) \right\rceil \\ &= O\left(\frac{n^9 E^6}{\varepsilon^4}\right) + O\left(\left(\frac{n^2 E^2}{\varepsilon^2}\right)^t\right) , \end{aligned} \tag{2.392}$$

where $N_1 := N_{\mathrm{cov}}\left(n, \frac{\varepsilon^2}{2(n+1)(1+4nE)^2}, \frac{\delta}{3}, E\right)$ is reported in (2.388) and $N_2 := N_{\mathrm{tom,CV}}\left(t, \frac{\varepsilon}{2}, \frac{\delta}{3}, 80n^2E^2\right)$ is reported in (2.390).

Output: A classical description of a pure state vector $|\hat{\psi}\rangle$ such that

$$\Pr\left(\frac{1}{2}\left\| |\hat{\psi}\rangle\langle\hat{\psi}| - |\psi\rangle\langle\psi| \right\|_1 \leq \varepsilon\right) \geq 1 - \delta\,. \tag{2.393}$$

Such a classical description consists of the triplet $(\tilde{\mathbf{m}}, \tilde{S}, |\tilde{\phi}_1\rangle)$ which defines $|\hat{\psi}\rangle$ via the relation

$$|\hat{\psi}\rangle := \hat{D}_{\tilde{\mathbf{m}}} U_{\tilde{S}} |\tilde{\phi}_1\rangle \otimes |0\rangle^{\otimes(n-t)} , \tag{2.394}$$

where $\tilde{\mathbf{m}} \in \mathbb{R}^{2n}$, $\tilde{S} \in \mathrm{Sp}(\mathbb{R}^{2n})$, and $|\tilde{\phi}_1\rangle$ is a t-mode pure state contained in a $\left\lceil \left(\frac{e(80n^2E^2 - \frac{1}{2})}{\varepsilon^2}\right)^t \right\rceil$-dimensional subspace.

1: Query N_1 copies of $|\psi\rangle$ and apply the algorithm in Table 2.2 to construct a vector $\tilde{\mathbf{m}}$ and a covariance matrix $\tilde{V}$ which are estimates of the first moment $\mathbf{m}(|\psi\rangle\langle\psi|)$ and of the covariance matrix $V(|\psi\rangle\langle\psi|)$, respectively.

2: Compute the symplectic diagonalisation $\tilde{V} = \tilde{S}\tilde{D}\tilde{S}^\intercal$, where $\tilde{S} \in \mathrm{Sp}(2n)$ and $\tilde{D} := \mathrm{diag}(\tilde{d}_1, \tilde{d}_1, \tilde{d}_2, \tilde{d}_2, \ldots, \tilde{d}_n, \tilde{d}_n)$, with $\tilde{d}_1 \geq \tilde{d}_2 \geq \ldots \geq \tilde{d}_n \geq 1$.

3: **for** $k \leftarrow 1$ **to** $\left\lceil 2N_2 + 24 \log\left(\frac{3}{\delta}\right) \right\rceil$ **do**

4: Query one copy of $|\psi\rangle$ and apply $\hat{D}^\dagger_{\tilde{\mathbf{m}}} U^\dagger_{\tilde{S}}$, obtaining the state vector $|\tilde{\psi}_t\rangle := \hat{D}^\dagger_{\tilde{\mathbf{m}}} U^\dagger_{\tilde{S}} |\psi\rangle$.

5: Measure $|\tilde{\psi}_t\rangle$ with respect the POVM $\{M_0 := \mathbb{1}_t \otimes |0\rangle\langle 0|^{\otimes(n-t)}, M_1 := \mathbb{1} - M_0\}$, and discard if the outcome corresponds to the POVM element M_1. The post-measurement state is thus of the form $|\tilde{\phi}\rangle \otimes |0\rangle^{\otimes(n-t)}$, where $|\tilde{\phi}\rangle$ is the pure state proportional to $\langle 0|^{\otimes(n-t)} |\tilde{\psi}_t\rangle$.

6: Do a step of the algorithm in Table 2.1 to perform pure-state tomography of the t-mode state vector $|\tilde{\phi}\rangle$, which has mean energy upper bounded by $80n^2E^2$.

7: **end for**

8: The tomography algorithm of Line 6 returns a classical description of a t-mode pure state vector $|\tilde{\phi}_1\rangle$, which is supported on a $\left\lceil \left(\frac{e(80n^2E^2 - \frac{1}{2})}{\varepsilon^2}\right)^t \right\rceil$-dimensional subspace.

9: **return** the triplet $\left(\tilde{\mathbf{m}}, \tilde{S}, |\tilde{\phi}_1\rangle\right)$.

Bibliography

[1] Anurag Anshu and Srinivasan Arunachalam. A survey on the complexity of learning quantum states. *Nature Rev. Phys.*, 6(1):59–69, 2024. doi: 10.1038/s42254-023-00662-4. URL https://doi.org/10.1038/s42254-023-00662-4.

[2] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. Quantum certification and benchmarking. *Nature Rev. Phys.*, 2(7):382–390, 2020. doi: 10.1038/s42254-020-0186-4.

[3] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature Comm.*, 1(1):149–149, 2010. doi: 10.1038/ncomms1147.

[4] Dolev Bluvstein, Simon J. Evered, Alexandra A. Geim, Sophie H. Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, J. Pablo Bonilla Ataides, Nishad Maskara, Iris Cong, Xun Gao, Pedro Sales Rodriguez, Thomas Karolyshyn, Giulia Semeghini, Michael J. Gullans, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, 2024. doi: 10.1038/s41586-023-06927-3. URL https://doi.org/10.1038/s41586-023-06927-3.

[5] Daniel T. Smithey, Mark Beck, Michael G. Raymer, and Adel Faridani. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Phys. Rev. Lett.*, 70(9): 1244–1247, 1993. doi: 10.1103/PhysRevLett.70.1244.

[6] Alex I. Lvovsky and Michael G. Raymer. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81(1):299–332, 2009. doi: 10.1103/RevModPhys.81.299.

[7] Alessio Serafini. *Quantum continuous variables: A primer of theoretical methods*. CRC Press, Taylor & Francis Group, Boca Raton, USA, 2017.

[8] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos. Efficient tomography of a quantum many-body system. *Nature Phys.*, 13(12):1158–1162, December 2017. doi: 10.1038/nphys4244.

[9] C. W. Helstrom. *Quantum detection and estimation theory*. Academic press, New York, USA, 1976.

[10] A. S. Holevo. Investigations in the general theory of statistical decisions. *Trudy Mat. Inst. Steklov*, 124:3–140, 1976. doi: 10.1070/RM9856. (English translation: Proc. Steklov Inst. Math. 124:1–140, 1978).

[11] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements, 2014.

[12] Ryan O'Donnell and John Wright. Efficient quantum tomography, 2015.

[13] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Th.*, 63:5628–5641, 2017. doi: 10.1109/tit.2017.2719044.

[14] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum Hamiltonians from high-temperature Gibbs states, 2021.

[15] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature Comm.*, 6(1):8498, 2015. ISSN 2041-1723. doi: 10.1038/ncomms9498.

[16] Srilekha Gandhari, Victor V. Albert, Thomas Gerrits, Jacob M. Taylor, and Michael J. Gullans. Precision bounds on continuous-variable state tomography using classical shadows, 2023.

[17] S. Becker, N. Datta, L. Lami, and C. Rouzé. Classical shadow tomography for continuous variables quantum systems. *IEEE Trans. Inf. Theory*, 70(5):3427–3452, 2024. doi: 10.1109/TIT.2024.3357972.

[18] Changhun Oh, Senrui Chen, Yat Wong, Sisi Zhou, Hsin-Yuan Huang, Jens A. H. Nielsen, Zheng-Hao Liu, Jonas S. Neergaard-Nielsen, Ulrik L. Andersen, Liang Jiang, and John Preskill. Entanglement-enabled advantage for learning a bosonic random displacement channel, 2024.

[19] Lars S. Madsen, Fabian Laudenbach, Mohsen Falamarzi. Askarani, Fabien Rortais, Trevor Vincent, Jacob F. F. Bulmer, Filippo M. Miatto, Leonhard Neuhaus, Lukas G. Helt, Matthew J. Collins, Adriana E. Lita, Thomas Gerrits, Sae Woo Nam, Varun D. Vaidya, Matteo Menotti, Ish Dhand, Zachary Vernon, Nicolás Quesada, and Jonathan Lavoie. Quantum computational advantage with a programmable photonic processor. *Nature*, 606(7912):75–81, 2022. doi: 10.1038/s41586-022-04725-x.

[20] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, Peng Hu, Xiao-Yan Yang, Wei-Jun Zhang, Hao Li, Yuxuan Li, Xiao Jiang, Lin Gan, Guangwen Yang, Lixing You, Zhen Wang, Li Li, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, December 2020. ISSN 1095-9203. doi: 10.1126/science.abe8770.

[21] D. Hangleiter and J. Eisert. Computational advantage of quantum random sampling. *Rev. Mod. Phys.*, 95:035001, 2023. doi: 10.1103/RevModPhys.95.035001.

[22] F. H. B. Somhorst, R. van der Meer, M. Correa Anguita, R. Schadow, H. J. Snijders, M. de Goede, B. Kassenberg, P. Venderbosch, C. Taballione, J. P. Epping, H. H. van den Vlekkert, J. F. F. Bulmer, J. Lugani, I. A. Walmsley, P. W. H. Pinkse, J. Eisert, N. Walk, and J. J. Renema. Quantum photo-thermodynamics on a programmable photonic quantum processor. *Nature Comm.*, 14:3895, 2023. doi: 10.1038/s41467-023-38413-9.

[23] Mazyar Mirrahimi, Zaki Leghtas, Victor V Albert, Steven Touzard, Robert J Schoelkopf, Liang Jiang, and Michel H Devoret. Dynamically protected cat-qubits: a new paradigm for universal quantum computation. *New J. Phys.*, 16(4):045014, April 2014. ISSN 1367-2630. doi: 10.1088/1367-2630/16/4/045014.

[24] Nissim Ofek, Andrei Petrenko, Reinier Heeres, Philip Reinhold, Zaki Leghtas, Brian Vlastakis, Yehan Liu, Luigi Frunzio, Steven Girvin, Liang Jiang, Mazyar Mirrahimi, M. Devoret, and R. Schoelkopf. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature*, 536, 07 2016. doi: 10.1038/nature18949.

[25] Marios H. Michael, Matti Silveri, R. T. Brierley, Victor V. Albert, Juha Salmilehto, Liang Jiang, and S. M. Girvin. New class of quantum error-correcting codes for a bosonic mode. *Phys. Rev. X*, 6:031006, Jul 2016. doi: 10.1103/PhysRevX.6.031006.

[26] Jérémie Guillaud and Mazyar Mirrahimi. Repetition cat qubits for fault-tolerant quantum computation. *Phys. Rev. X*, 9(4):041053, December 2019. ISSN 2160-3308. doi: 10.1103/physrevx.9.041053.

[27] Koen Alexander, Andrea Bahgat, Avishai Benyamini, Dylan Black, Damien Bonneau, Stanley Burgos, Ben Burridge, Geoff Campbell, Gabriel Catalano, Alex Ceballos, Chia-Ming Chang, CJ Chung, Fariba Danesh, Tom Dauer, Michael Davis, Eric Dudley, Ping Er-Xuan, Josep Fargas, Alessandro Farsi, Colleen Fenrich, Jonathan Frazer, Masaya Fukami, Yogeeswaran Ganesan, Gary Gibson, Mercedes Gimeno-Segovia, Sebastian Goeldi, Patrick Goley, Ryan Haislmaier, Sami Halimi, Paul Hansen, Sam Hardy, Jason Horng, Matthew House, Hong Hu, Mehdi Jadidi, Henrik Johansson, Thomas Jones, Vimal Kamineni, Nicholas Kelez, Ravi Koustuban, George Kovall, Peter Krogen, Nikhil Kumar, Yong Liang, Nicholas LiCausi, Dan Llewellyn, Kimberly Lokovic, Michael Lovelady, Vitor Manfrinato, Ann Melnichuk, Mario Souza, Gabriel Mendoza, Brad Moores, Shaunak Mukherjee, Joseph Munns, Francois-Xavier Musalem, Faraz Najafi, Jeremy L. O'Brien, J. Elliott Ortmann, Sunil Pai, Bryan Park, Hsuan-Tung Peng, Nicholas Penthorn, Brennan Peterson, Matt Poush, Geoff J. Pryde, Tarun Ramprasad, Gareth Ray, Angelita Rodriguez, Brian Roxworthy, Terry Rudolph, Dylan J. Saunders, Pete Shadbolt, Deesha Shah, Hyungki Shin, Jake Smith, Ben Sohn, Young-Ik Sohn, Gyeongho Son, Chris Sparrow, Matteo Staffaroni, Camille Stavrakas, Vijay Sukumaran, Davide Tamborini, Mark G. Thompson, Khanh Tran, Mark Triplet, Maryann Tung, Alexey Vert, Mihai D. Vidrighin, Ilya Vorobeichik, Peter Weigel, Mathhew Wingert, Jamie Wooding, and Xinran Zhou. A manufacturable platform for photonic quantum computing, 2024.

[28] M. M. Wolf, D. Pérez-García, and G. Giedke. Quantum capacities of bosonic channels. *Phys. Rev. Lett.*, 98:130501, 2007. doi: 10.1103/PhysRevLett.98.130501.

[29] M. Takeoka, S. Guha, and Mark M. Wilde. Fundamental rate-loss tradeoff for optical quantum key distribution. *Nature Comm.*, 5(1):5235, 2014. doi: 10.1038/ncomms6235.

[30] Stefanno Pirandola, Riccardo Laurenza, Carlo Ottaviani, and Leonardo Banchi. Fundamental limits of repeaterless quantum communications. *Nature Comm.*, 8(1):15043, 2017. doi: 10.1038/ncomms15043.

[31] F. A. Mele, Ludovico Lami, and V. Giovannetti. Restoring quantum communication efficiency over high loss optical fibers. *Phys. Rev. Lett.*, 129:180501, 2022. doi: 10.1103/PhysRevLett.129.180501.

[32] Francesco Anna Mele, Ludovico Lami, and Vittorio Giovannetti. Maximum tolerable excess noise in continuous-variable quantum key distribution and improved lower bound on two-way capacities. *Nature Photonics*, 19(3):329–334, January 2025. ISSN 1749-4893. doi: 10.1038/s41566-024-01595-9. URL http://dx.doi.org/10.1038/s41566-024-01595-9.

[33] Francesco Anna Mele, Farzin Salek, Vittorio Giovannetti, and Ludovico Lami. Quantum communication on the bosonic loss-dephasing channel, 2024.

[34] The LIGO Scientific Collaboration. Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light. *Nature Phot.*, 7(8):613–619, 2013. doi: 10.1038/nphoton.2013.177.

[35] Junhua Zhang, Mark Um, Dingshun Lv, Jing-Ning Zhang, Lu-Ming Duan, and Kihwan Kim. Noon states of nine quantized vibrations in two radial modes of a trapped ion. *Phys. Rev. Lett.*, 121:160502, Oct 2018. doi: 10.1103/PhysRevLett.121.160502.

[36] V. Meyer, M. A. Rowe, D. Kielpinski, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. Experimental demonstration of entanglement-enhanced rotation angle estimation using trapped ions. *Phys. Rev. Lett.*, 86:5870–5873, Jun 2001. doi: 10.1103/PhysRevLett.86.5870.

[37] K. McCormick, J. Keller, S. Burd, D. Wineland, A. Wilson, and D. Leibfried. Quantum-enhanced sensing of a single-ion mechanical oscillator. *Nature*, 572:1, 08 2019. doi: 10.1038/s41586-019-1421-y.

[38] P. Marian and T. A. Marian. Relative entropy is an exact measure of non-Gaussianity. *Phys. Rev. A*, 88:012322, 2013. doi: 10.1103/PhysRevA.88.012322.

[39] S. Lloyd and S. L. Braunstein. Quantum computation over continuous variables. *Phys. Rev. Lett.*, 82:1784–1787, 1999. doi: 10.1103/PhysRevLett.82.1784.

[40] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. Learning t-doped stabilizer states. *Quantum*, 8:1361, May 2024. ISSN 2521-327X. doi: 10.22331/q-2024-05-27-1361. URL http://dx.doi.org/10.22331/q-2024-05-27-1361.

[41] Salvatore F.E. Oliviero, Lorenzo Leone, and Alioscia Hamma. Transitions in entanglement complexity in random quantum circuits by measurements. *Phys. Lett. A*, 418:127721, December 2021. ISSN 0375-9601. doi: 10.1016/j.physleta.2021.127721. URL http://dx.doi.org/10.1016/j.physleta.2021.127721.

[42] Lorenzo Leone, Salvatore F. E. Oliviero, You Zhou, and Alioscia Hamma. Quantum chaos is quantum. *Quantum*, 5:453, May 2021. ISSN 2521-327X. doi: 10.22331/q-2021-05-04-453. URL http://dx.doi.org/10.22331/q-2021-05-04-453.

[43] Salvatore F. E. Oliviero, Lorenzo Leone, Seth Lloyd, and Alioscia Hamma. Unscrambling quantum Information with Clifford decoders. *Phys. Rev. Lett.*, 132:080402, Feb 2024. doi: 10.1103/PhysRevLett.132.080402. URL https://link.aps.org/doi/10.1103/PhysRevLett.132.080402.

[44] Lorenzo Leone, Salvatore F. E. Oliviero, Seth Lloyd, and Alioscia Hamma. Learning efficient decoders for quasichaotic quantum scramblers. *Phys. Rev. A*, 109:022429, Feb 2024. doi: 10.1103/PhysRevA.109.022429. URL https://link.aps.org/doi/10.1103/PhysRevA.109.022429.

[45] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Efficient learning of quantum states prepared with few non-Clifford gates, 2023.

[46] Antonio Anna Mele and Yaroslav Herasymenko. Efficient learning of quantum states prepared with few fermionic non-Gaussian gates, 2024.

[47] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance, 2023.

[48] Sumeet Khatri and Mark M. Wilde. Principles of quantum communication theory: A modern approach, 2020.

[49] Victor Kaftal and Gary Weiss. An infinite dimensional Schur-Horn theorem and majorization theory with applications to operator ideals, 2009.

[50] John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016.

[51] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi: 10.1017/9781108231596.

[52] Leonardo Banchi, Samuel L. Braunstein, and Stefano Pirandola. Quantum fidelity for arbitrary Gaussian states. *Phys. Rev. Lett.*, 115(26):260501, 2015. doi: 10.1103/physrevlett.115.260501.

[53] Simon Becker, Nilanjana Datta, Ludovico Lami, and Cambyse Rouzé. Energy-constrained discrimination of unitaries, quantum speed limits and a Gaussian Solovay–Kitaev theorem. *Phys. Rev. Lett.*, 126:190504, 2021. doi: 10.1103/PhysRevLett.126.190504.

[54] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional Gaussians with the same mean, 2023.

[55] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature Comm.*, 6(1):8498, 2015. ISSN 2041-1723. doi: 10.1038/ncomms9498.

[56] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Th. Comp. Sc.*, 43:169–188, 1986. ISSN 0304-3975. doi: 10.1016/0304-3975(86)90174-X.

[57] Arkadi S. Nemirovskiĭ. *Problem complexity and method efficiency in optimization*. Wiley-Interscience Series in Discrete Mathematics. New York: Wiley, 1983. ISBN 978-0-471-10345-5.

[58] Jens Eisert and Martin B. Plenio. Introduction to the basics of entanglement theory in continuous-variable systems. *Int. J. Quantum Inf.*, 01(04):479–506, 2003. doi: 10.1142/S0219749903000371.

[59] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, and Seth Lloyd. Gaussian quantum information. *Rev. Mod. Phys.*, 84(2):621–621, 2012. doi: 10.1103/RevModPhys.84.621.

[60] Ludovico Lami, B. Regula, X. Wang, R. Nichols, Andreas Winter, and Gerardo Adesso. Gaussian quantum resource theories. *Phys. Rev. A*, 98:022335, 2018. doi: 10.1103/PhysRevA.98.022335.

[61] Edwin T. Jaynes. Information theory and statistical mechanics. ii. *Phys. Rev.*, 108:171–190, 1957. doi: 10.1103/PhysRev.108.171.

[62] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning?, 2023.

[63] Madalin Guta, Jonas Kahn, Richard Kueng, and Joel A. Tropp. Fast state tomography with optimal error bounds, 2018.

[64] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature Comm.*, 1(1):1038, 2010. ISSN 2041-1723. doi: 10.1038/ncomms1147.

[65] Marco Fanizza, Niklas Galke, Josep Lumbreras, Cambyse Rouzé, and Andreas Winter. Learning finitely correlated states: stability of the spectral reconstruction, 2023.

[66] Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean. Learning shallow quantum circuits, 2024.

[67] Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. Optimal algorithms for learning quantum phase states, 2023.

[68] Ashley Montanaro. Learning stabilizer states by bell sampling, 2017.

[69] Dominik Hangleiter and Michael J. Gullans. Bell sampling from quantum circuits, 2024.

[70] Scott Aaronson and Sabee Grewal. Efficient tomography of non-interacting fermion states, 2023.

[71] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, July 2006. ISBN 0471241954.

[72] Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. 1973.

[73] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.

[74] J. Watrous. *The theory of quantum information*. Cambridge University Press, 2018. doi: 10.1017/9781316848142.

[75] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 2013.

[76] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010.

[77] X.-B. Wang, T. Hiroshima, A. Tomita, and M. Hayashi. Quantum information with Gaussian states. *Phys. Rep.*, 448(1):1–111, 2007. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2007.04.005.

[78] F. Hiai and D. Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Comm. Math. Phys.*, 143(1):99–114, 1991. doi: 10.1007/BF02100287.

[79] T. Ogawa and H. Nagaoka. Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Trans. Inf. Th.*, 46(7):2428–2433, 2000. ISSN 0018-9448. doi: 10.1109/18.887855.

[80] M. G. Genoni, M. G. A. Paris, and Konrad Banaszek. Quantifying the non-Gaussian character of a quantum state by quantum relative entropy. *Phys. Rev. A*, 78:060303, 2008. doi: 10.1103/PhysRevA.78.060303.

[81] Chris A. Fuchs and Jeroen van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inf. Th.*, 45(4):1216–1227, 1999. ISSN 0018-9448. doi: 10.1109/18.761271.

[82] Andreas Winter. Energy-constrained diamond norm with applications to the uniform continuity of continuous variable channel capacities, 2017.

[83] Maxim E. Shirokov. On the energy-constrained diamond norm and its application in quantum information theory. *Probl. Inf. Transm.*, 54(1):20–33, 2018. ISSN 1608-3253. doi: 10.1134/S0032946018010027.

[84] Jamil Arbas, Hassan Ashtiani, and Christopher Liaw. Polynomial time and private learning of unbounded Gaussian mixture models, 2023.

[85] Lennart Bittel, Francesco Anna Mele, Antonio Anna Mele, Salvatore Tirone, and Ludovico Lami. Optimal estimates of trace distance between bosonic Gaussian states and applications to learning, 2024. URL https://arxiv.org/abs/2411.02368.

[86] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2nd edition, 2017. ISBN 9781316813300.

[87] Daniel Gottesman. The Heisenberg representation of quantum computers, 1998.

[88] A. Mari and J. Eisert. Positive Wigner functions render classical simulation of quantum computation efficient. *Phys. Rev. Lett.*, 109:230503, 2012. doi: 10.1103/PhysRevLett.109.230503.

[89] Ulysse Chabaud, Giulia Ferrini, Frédéric Grosshans, and Damian Markham. Classical simulation of Gaussian quantum circuits with non-Gaussian input states. *Physical Review Research*, 3(3), July 2021. doi: 10.1103/physrevresearch.3.033018. URL http://dx.doi.org/10.1103/PhysRevResearch.3.033018.

[90] Beatriz Dias and Robert Koenig. Classical simulation of non-Gaussian bosonic circuits, 2024.

[91] Oliver Hahn, Ryuji Takagi, Giulia Ferrini, and Hayata Yamasaki. Classical simulation and quantum resource theory of non-Gaussian optics, 2024.

[92] Martin Idel, Sebastián Soto Gaona, and Michael M. Wolf. Perturbation bounds for Williamson's symplectic normal form. *Lin. Alg. Appl.*, 525:45–58, 2017. ISSN 0024-3795. doi: 10.1016/j.laa.2017.03.013.

Chapter 3

Optimal estimates of trace distance between bosonic Gaussian states and applications to learning

In the previous chapter, we spent a great deal of time studying the tomography of bosonic Gaussian states. Their significance lies in their simplicity, which in turn rests on the fact that they are uniquely determined by two experimentally accessible quantities, namely their first and second moments. The last chapter posed a question that was not optimally resolved: what if these moments are only known *approximately*, as is inevitable in any realistic experiment? What is the resulting error on the Gaussian state itself, as measured by the most operationally meaningful metric for distinguishing quantum states, namely the trace distance?

In this chapter, we fully resolve this question by demonstrating that if the first and second moments are known up to an error ε, the trace-distance error on the state also scales as ε, and that this functional dependence is optimal. To prove this, we establish tight bounds on the trace distance between two Gaussian states in terms of the norm distance of their first and second moments. As an application, we improve existing bounds on the sample complexity of tomography of Gaussian states. In our analysis, we introduce the general notion of the derivative of a Gaussian state and uncover its fundamental properties, enhancing our understanding of the structure of the set of Gaussian states.

[This chapter is based on L. Bittel, F. A. Mele, A. A. Mele, S. Tirone, L. Lami, Optimal estimates of trace distance between bosonic Gaussian states and applications to learning, Quantum 9, 1769 (2025)]

3.1 Introduction

Continuous variable (CV) systems [1–5], such as quantum optical systems, play a key role in our quantum technologies, including quantum computation [6–14], communication [15–22], and sensing [23–27]. Additionally, CV systems have become increasingly popular in efforts to demonstrate quantum advantage, particularly through boson sampling [28–30] and quantum simulation experiments [31]. Among all CV quantum states, *Gaussian states* [2] are arguably among the most important ones. Indeed, on the one hand Gaussian states are the most common class of states arising in Nature, while on the other they are relatively simple to analyse mathematically.

Gaussian states are the quantum analog of Gaussian probability distributions. Just as a Gaussian probability distribution is uniquely identified by its classical first moment and its classical covariance matrix, a Gaussian state is uniquely identified by its *first moment* $\mathbf{m}$ and its *covariance matrix* V [2]. Importantly, these quantities can be experimentally estimated through methods such as homodyne or heterodyne detection [2], which are routinely performed in quantum optical laboratories. Thus, estimating a Gaussian state requires only the estimation of its first

moment and covariance matrix. In practice, however, we can never estimate the first moment and the covariance matrix *exactly*; instead, we will always have a non-zero error ε in such estimates, meaning that we can only estimate a Gaussian state up to some error. The goal of this work is to understand how an estimation error ε in the first moment and the covariance matrix of a Gaussian state propagates into the overall error in estimating the Gaussian state itself:

Problem 93 (Informal version)**.** *Let $\rho(V, \mathbf{m})$ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Let $\tilde{V}$ and $\tilde{\mathbf{m}}$ be ε-approximations of V and $\mathbf{m}$, respectively:*

$$\tilde{\mathbf{m}} \approx_\varepsilon \mathbf{m}\,, \qquad \tilde{V} \approx_\varepsilon V\,. \tag{3.1}$$

The Gaussian state $\rho(\tilde{V}, \tilde{\mathbf{m}})$ with first moment $\tilde{\mathbf{m}}$ and covariance matrix $\tilde{V}$ constitutes an approximation of $\rho(V, \mathbf{m})$:

$$\rho(\tilde{V}, \tilde{\mathbf{m}}) \approx \rho(V, \mathbf{m})\,. \tag{3.2}$$

The problem is to understand what is the error incurred in the approximation in Eq. (3.2) in terms of the error ε incurred in the approximations of the first moment and of the covariance matrix.

How can we measure the error involved in estimating a quantum state? The most operationally meaningful approach is to define this error using the *trace distance* between the true state and the estimated state, as established by the well-known Holevo–Helstrom theorem [32, 33]. This theorem gives the trace distance a strong operational interpretation, identifying it as the appropriate metric for quantifying the probability of error in discriminating ρ and σ with quantum measurements [32, 33]. The trace distance is ubiquitous in quantum information theory, playing a central role in evaluating the performance of numerous quantum algorithms [34–37]. Mathematically, the trace distance between two quantum states ρ and σ is given by $\frac{1}{2}\|\rho - \sigma\|_1$, where $\|\Theta\|_1 = \operatorname{Tr}\sqrt{\Theta^\dagger\Theta}$ denotes the *trace norm* of a linear operator Θ.

Given the operational interpretation of the trace distance, it is thus a fundamental problem to determine what the error incurred in the trace distance is when one estimates the first moment and the covariance matrix of an (unknown) Gaussian state up to a precision ε. Equivalently, it is a fundamental problem to estimate the trace distance between two Gaussian states in terms of the norm distances between their first moments and their covariance matrices:

Problem 94 (Formal version of Problem 93)**.** *Let $\rho(V, \mathbf{m})$ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Similarly, let $\rho(\tilde{V}, \tilde{\mathbf{m}})$ be the Gaussian state with first moment $\tilde{\mathbf{m}}$ and covariance matrix $\tilde{V}$. The problem is to determine stringent bounds on the trace distance*

$$\frac{1}{2}\left\|\rho(V, \mathbf{m}) - \rho(\tilde{V}, \tilde{\mathbf{m}})\right\|_1 \tag{3.3}$$

between $\rho(V, \mathbf{m})$ and $\rho(\tilde{V}, \tilde{\mathbf{m}})$ in terms of the norm distance between their first moments $\|\mathbf{m} - \tilde{\mathbf{m}}\|$ and the norm distance between their covariance matrices $\|V - \tilde{V}\|$. (Note that, since the first moments and covariance matrices are finite-dimensional matrices, the choice of the norm $\|\cdot\|$ used to measure distances between them is not crucial, as all norms are equivalent in finite dimensions.)

Deriving such bounds is of independent interest, as observed also by Alexander S. Holevo in [38, 39]. Indeed, they can be useful in analysing tomography of Gaussian states [40], characterising quantum Gaussian observables [38], testing properties of Gaussian systems [41], and learning Gaussian processes [42]. Problem 94 was initially posed in Ref. [40], where some preliminary yet unsatisfactory results were also obtained. Recently, some further progress has been made by Holevo [38]. However, neither the results in [40] nor those in [38] could fully resolve the problem of what a tight upper bound on the trace distance might look like.

In this work, we provide a complete solution to this problem, deriving tight bounds on the trace distance between two Gaussian states based on the norm distance of their first moments and

covariance matrices. Our bounds demonstrate that estimating the first moment and covariance matrix of a Gaussian state with precision ε results in a trace distance error that scales linearly with ε (and vice versa). We expect that these new trace distance bounds will serve as an important technical tool in advancing the field of quantum optics and CV quantum information.

To date, CV researchers have primarily relied on fidelity [43] rather than trace distance to estimate the distance between two Gaussian states. This preference stems from the availability of a well-known closed formula for the fidelity between Gaussian states [43]. However, this approach has two significant drawbacks: (i) fidelity lacks the strong operational interpretation that trace distance offers, and (ii) the fidelity formula is cumbersome, making it challenging to use in analytical calculations. In contrast, the trace distance bounds introduced in this work are both operationally meaningful and analytically tractable, providing a practical alternative to the widely used fidelity formula [43] for quantifying the distance between Gaussian states.

As an application of our bounds, we establish a tighter sample complexity bound for Gaussian state tomography, outperforming current state-of-the-art result [40]. This contributes to the rapidly growing literature on quantum learning theory with CV systems [40, 44–52].

In our analysis, we rigorously define and discover fundamental properties of the *derivative* of a Gaussian state with respect to its first moment and covariance matrix. Remarkably, we demonstrate that this derivative can be expressed through a relatively simple formula, offering new insights into the underlying structure of the geometry of the set of Gaussian states. We believe that this result will have broad applications in quantum metrology with CV systems [2, 53–55]. In fact, the computation of state derivatives is a central component in the analysis of quantum metrology protocols, essentially because the celebrated quantum Cramér–Rao bound [2] inherently relies on state derivatives. Until now, derivatives of Gaussian states in the context of quantum metrology have been computed only for very specific examples [56–58]. In contrast, our results apply to arbitrary Gaussian states, providing new tools that can advance the field of quantum metrology.

3.2 Prior work

Let us recap the recent rapid advances around the solution of Problem 94. First, this problem was posed in [40], where some upper and lower bounds on the trace distance between Gaussian states in terms of the norm distance between the first moments and covariance matrices were established. Specifically, the bounds proved in [40] guarantee that an ε-error in the moments leads to a trace distance error that scales at least as ε and at most as $\sqrt{\varepsilon}$ [40]. Later, Holevo elegantly proved a new trace distance upper bound that can outperform the one in [40] in some regimes [38]. However, Holevo's new upper bound still leads to a trace distance error that scales as $\sqrt{\varepsilon}$ [38] in most cases of practical interest. This suboptimal error scaling has serious drawbacks, as it can lead to a significant overestimation of the performance of quantum algorithms involving Gaussian states, especially in the context of quantum learning [40]. Given the central role of Gaussian states in optical quantum technologies, addressing this limitation and determining the optimal error dependence is thus a crucial, pressing problem. In this work, we improve on this dependence and fully resolve the problem, establishing that the trace distance error must scale as ε.

One might hope that a good trace distance upper bound could be derived by looking at the fidelity instead, which can be computed explicitly for Gaussian states [43]. However, this approach ultimately fails because the fidelity formula derived in [43] is too complicated to yield a simple bound in terms of the differences between the moments. Also, it would be much worse than our bound due to the square root loss in the Fuchs–van de Graaf inequality [34, 35, 37].

In the fermionic setting, tight bounds have been recently derived for the trace distance between two fermionic Gaussian states [59]. However, while the latter can in principle be computed numerically (for system with a small number of fermionic modes), the trace distance between two bosonic Gaussian states is more challenging to compute numerically (even for a single-mode system),

as doing that would require the diagonalisation of an infinite-dimensional matrix. This underscores the critical importance of deriving such trace distance bounds in the bosonic setting, where they are even more essential than in the fermionic case and arguably considerably more challenging to establish due to the inherent difficulties of working with infinite-dimensional quantum systems.

3.3 Notation

By definition, a CV system with n *modes* is represented by the Hilbert space $L^2(\mathbb{R}^n)$, which comprises all square-integrable complex-valued functions over $\mathbb{R}^n$. The first moment $\mathbf{m}$ and the covariance matrix V of a quantum state ρ are defined as $\mathbf{m} \coloneqq \operatorname{Tr}\left[\hat{\mathbf{R}}\,\rho\right]$ and $V \coloneqq \operatorname{Tr}\left[\left\{(\hat{\mathbf{R}} - \mathbf{m}\,\hat{\mathbb{1}}), (\hat{\mathbf{R}} - \mathbf{m}\,\hat{\mathbb{1}})^\intercal\right\}\rho\right]$, where $(\cdot)^\intercal$ denotes the transpose operation, $\{\cdot,\cdot\}$ represents the anti-commutator, and $\hat{\mathbf{R}} \coloneqq (\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n)^\intercal$ is the quadrature operator vector, with $\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n$ being the well-known position and momentum operators of each mode [2]. The quadrature operator vector satisfies the canonical commutation relation $[\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal] = i\,\Omega\,\hat{\mathbb{1}}$, where $\Omega \coloneqq \bigoplus_{i=1}^{n} \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

The set of n-mode Gaussian states is in one-to-one correspondence with the set of pairs $(V, \mathbf{m})$, where V is a $2n \times 2n$ real matrix satisfying the *uncertainty relation* $V + i\Omega \geq 0$, and $\mathbf{m}$ is a $2n$-dimensional real vector [2]. Specifically, given such a pair $(V, \mathbf{m})$ with V satisfying the uncertainty relation, there exists a unique Gaussian state with covariance matrix V and first moment $\mathbf{m}$ [2]. Conversely, the covariance matrix of any (Gaussian) state satisfies the uncertainty relation [2]. For the rest of the paper, we will denote as $\rho(V, \mathbf{m})$ the Gaussian state with covariance matrix V and first moment $\mathbf{m}$.

Given a matrix V, we write $|V| \coloneqq \sqrt{V^\dagger V}$ for its modulus, $\|V\|_\infty$ for its operator norm (i.e. the maximum singular value of V), $\|V\|_1 \coloneqq \operatorname{Tr}|V|$ for its trace norm, and $\|V\|_2 \coloneqq \sqrt{\operatorname{Tr}[V^\dagger V]}$ for its Hilbert–Schmidt norm. The Euclidean norm of a vector $\mathbf{m}$ is denoted as $\|\mathbf{m}\|_2 \coloneqq \sqrt{\mathbf{m}^\intercal \mathbf{m}}$.

3.4 Tight upper bound

The following theorem establishes a tight upper bound on the trace distance between two Gaussian states in terms of the norm distances between their first moments and covariance matrices.

Theorem 95 (Tight upper bound on the trace distance between Gaussian states)**.** *Let $\rho(V, \mathbf{m})$ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Similarly, let $\rho(W, \mathbf{t})$ be the Gaussian state with first moment $\mathbf{t}$ and covariance matrix W. Then, their trace distance can be upper bounded as follows:*

$$\begin{aligned} \frac{1}{2}\left\|\rho(V,\mathbf{m}) - \rho(W,\mathbf{t})\right\|_1 \leq{}& \tfrac{1+\sqrt{3}}{8} \operatorname{Tr}\left[|V-W|\,\Omega^\intercal\left(\tfrac{V+W}{2}\right)\Omega\right] \\ &+ \sqrt{\tfrac{\min(\|V\|_\infty, \|W\|_\infty)}{2}}\,\|\mathbf{m}-\mathbf{t}\|_2\,. \end{aligned} \tag{3.4}$$

To prove the above result, we develop a new mathematical framework based on the notion of *derivative of a Gaussian state*, which can be of independent interest (see e.g. [60]), especially for the field of quantum metrology [2, 54, 55]. See Section 3.8.2 of the Supplemental Material (SM) [61].

Crucially, Hölder's inequality implies that the term $\operatorname{Tr}\left[|V - W|\Omega^\intercal \left(\frac{V+W}{2}\right)\Omega\right]$ in (3.4) can be further upper bounded either by $\max(\|V\|_\infty, \|W\|_\infty)\|V - W\|_1$, or by $\max(\operatorname{Tr} V, \operatorname{Tr} W)\|V - W\|_\infty$. This is important because it implies that the upper bound in Theorem 95 provides a linear dependence on the norm of $V - W$, improving upon the square root dependence previously established in [40] and [38]. Remarkably, the lower bound presented in [40, Eq. S265] also exhibits the same linear dependence as our new upper bound. Consequently, it follows that our new bound provides a complete solution to Problem 94. This fully answers a fundamental question for the

field of quantum optics and CV quantum information [40], demonstrating that if the first moment and the covariance matrix of a quantum Gaussian state are known within an error ε, the associated trace distance uncertainty scales linearly with ε (and this functional dependence is optimal).

Notably, the bound in Theorem 95 is *tight* with respect to its dependence on $\varepsilon := \|V-W\|_\infty$ and $a := \max(\|V\|_\infty, \|W\|_\infty)$. Specifically, this bound establishes that $\frac{1}{2}\|\rho(V,0)-\rho(W,0)\|_1 \leq \frac{1+\sqrt{3}}{8}a\varepsilon$. As proved in Lemma 122 in the SM, the trace distance between the two Gaussian states with covariance matrices $V := \left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right)$ and $W := \left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1}+\varepsilon \end{smallmatrix}\right)$ and zero first moments satisfies

$$\lim_{\varepsilon\to 0^+} \frac{\frac{1}{2}\|\rho(W,0)-\rho(V,0)\|_1}{\varepsilon} = \frac{1+\sqrt{3}}{8}a \quad \forall a \geq 1\,, \tag{3.5}$$

establishing that the constant $\frac{1+\sqrt{3}}{8}$ in (3.4) is optimal. We can achieve have optimality also for the first moment dependence as all approximation in (3.118) can be made tight for coherent states.

3.5 Application to quantum state tomography

As a direct application of Theorem 95, we can provide an upper bound on the sample complexity of tomography of Gaussian states [40] that outperforms the one provided in [40]. Let us start by briefly introducing the fundamental problem of tomography of Gaussian states [40].

Problem 96 (Tomography of Gaussian states)**.** *Given N copies of an unknown n-mode Gaussian state ρ satisfying the energy constraint $\mathrm{Tr}[\rho\hat{E}] \leq E$, the goal is to construct a classical description of a Gaussian state $\tilde{\rho}$ such that*

$$\Pr\left(\frac{1}{2}\|\tilde{\rho}-\rho\|_1 \leq \varepsilon\right) \geq 1-\delta\,. \tag{3.6}$$

In words, $\tilde{\rho}$ must be such that the probability that it is ε-close to ρ (in trace distance) is not smaller than $1-\delta$. Here, $\hat{E} := \frac{1}{2}\hat{\mathbf{R}}^\intercal\hat{\mathbf{R}}$ denotes the energy operator [2].

Given the number of modes n, the energy constraint E, the trace distance error ε, and the failure probability δ, the *sample complexity* of tomography of Gaussian states is defined as the minimum number of copies N that allows one to solve Problem 96. In the following theorem, we find an upper bound on such a sample complexity.

Theorem 97 (Upper bound on the sample complexity of tomography of energy-constrained Gaussian states)**.** *Let $\varepsilon, \delta \in (0,1]$. Let ρ be an unknown n-mode Gaussian state with (unknown) covariance matrix V and arbitrary first moment, satisfying the energy constraint $\mathrm{Tr}[\rho\hat{E}] \leq E$. A number*

$$N = O\left(\frac{(\|V\|_1\|V\|_\infty)^2}{\varepsilon^2}\left(n+\log(2/\delta)\right)\right) \tag{3.7}$$

$$\leq O\left(\frac{E^4}{\varepsilon^2}\left(n+\log(2/\delta)\right)\right) \tag{3.8}$$

of copies of ρ suffices to construct a classical description of a Gaussian state $\tilde{\rho}$ such that $\Pr\left(\frac{1}{2}\|\tilde{\rho}-\rho\|_1 \leq \varepsilon\right) \geq 1-\delta$.

The proof is provided in Theorem 147 and Corollary 148 in the SM [61], where explicit sample-complexity bounds (without big-O notation) are provided. The above Theorem 97 improves upon the upper bound on the sample complexity of tomography of Gaussian states presented in Ref. [40], which was $N = O\left(\frac{n^3E^4}{\varepsilon^4}\right)$. The improvement is evident in both the scaling with the number of modes n and the scaling with the trace distance error ε, ultimately attributed to our trace distance bound in Theorem 95.

3.6 Discussion and conclusions

In this work, we have presented novel and optimal bounds on the trace distance between bosonic Gaussian states in terms of the norm difference of their first moments and covariance matrices. These bounds allow one to understand how errors in estimating the first moment and covariance matrix of a Gaussian state propagate to the trace distance error, which is widely regarded as the most operationally meaningful measure of error in the estimation of a quantum state. This result fully resolves an open question posed in [40] and subsequently explored by Holevo [38]. Our analysis can be viewed as a quantum analog of the estimates of the total variation distance between Gaussian probability distributions in the classical probability literature [62–64], which play a significant role in machine learning applications [64].

As an application of our new trace distance bounds, we have improved the existing bounds on the sample complexity of tomography of Gaussian states [40]. This result contributes to the rapidly growing literature on quantum learning theory and state learning [44, 65–78], particularly at its interface with CV systems, which have been relatively unexplored until recently [45–52]. Since Gaussian states play a major role both theoretically and experimentally [2], we believe that our results on their learnability hold both fundamental as well as practical value.

Finally, we have introduced the notion of derivative of Gaussian states and uncovered its fundamental properties, deepening our understanding of the structure of the set of Gaussian states. We believe that these findings will have applications in quantum metrology and sensing [2, 54, 55]. In fact, since the first posting of this paper on arXiv, progress in this research direction has already begun [79].

As additional future research directions, we propose extending our results to derive analogous bounds for the distance between Gaussian channels [2] and investigating their learnability [42, 44]. Additionally, applying our trace-distance bounds to other learning tasks, such as property testing of bosonic Gaussian states [44, 59], offers a compelling avenue for exploration. Together with our findings and the potential applications of Gaussian state derivatives in quantum metrology and sensing, these directions present exciting opportunities to further enhance the role of CV systems in advancing quantum technologies.

3.7 Technical tools

To establish our trace distance bounds, we introduce numerous technical tools that may hold independent interest for the field of quantum optics, CV quantum information, and quantum metrology. Among these tools, we introduce the concept of the *derivative of a Gaussian state* (see Section 3.8.2). Specifically, for an n-mode Gaussian state $\rho(V, \mathbf{m})$ with covariance matrix V and first moment $\mathbf{m}$, and given a $2n \times 2n$ real symmetric matrix X and a $2n$-dimensional real vector $\mathbf{x}$, we define the derivative of $\rho(V, \mathbf{m})$ in the direction $(X, \mathbf{x})$ as the following limit of incremental ratios:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\alpha}\rho(V + \alpha X, \mathbf{m} + \alpha\mathbf{x})\Big|_{\alpha=0} \\ &:= \lim_{\alpha\to 0^+} \frac{\rho(V + \alpha X, \mathbf{m} + \alpha\mathbf{x}) - \rho(V, \mathbf{m})}{\alpha}. \end{aligned} \tag{3.9}$$

The existence of such a limit is rigorously proved in Section 3.8.3 in the SM from a functional analysis perspective. Notably, the following lemma, proved in Theorem 115 in the SM, demonstrates that the derivative of a Gaussian state can be expressed with a relatively simple formula:

Lemma 98 (Compact formula for the derivative of a Gaussian state)**.** *The derivative of the*

Gaussian state $\rho(V,\mathbf{m})$ along the direction $(X,\mathbf{x})$ can be expressed as:

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\alpha}\rho(V+\alpha X,\boldsymbol{m}+\alpha\boldsymbol{x})\Big|_{\alpha=0}\\&=-\frac{1}{4}\sum_{k,j=1}^{2n}\tilde{X}_{kj}[\hat{R}_k,[\hat{R}_j,\rho(V,\mathbf{m})]]+i\sum_{j=1}^{2n}\tilde{x}_j[\hat{R}_j,\rho(V,\mathbf{m})],\end{aligned}\tag{3.10}$$

where $\tilde{X}\coloneqq\Omega^{\intercal}X\Omega$, $\tilde{\mathbf{x}}\coloneqq\Omega\mathbf{x}$, and $\hat{R}_j$ denotes the jth quadrature operator.

Such a compact formula is crucial for the proof of our trace-distance bounds and allows to shed light on the geometry of set of Gaussian states. Remarkably, in the following lemma, proved in Theorem 119 in the SM, we show how to calculate the trace norm of the derivative of a pure Gaussian state *exactly* in terms of its covariance matrix:

Lemma 99 (Closed formula for the trace norm of the derivative of a pure Gaussian state)**.** *Let $\rho(V,\boldsymbol{m})$ be a pure Gaussian state with a (pure) covariance matrix V and first moment $\boldsymbol{m}$. Then, the trace norm of the derivative of the Gaussian state $\rho(V,\boldsymbol{m})$ along the direction $(X,0)$ is given by*

$$\begin{aligned}&\left\|\frac{\mathrm{d}}{\mathrm{d}\alpha}\rho(V+\alpha X,\boldsymbol{m})\Big|_{\alpha=0}\right\|_1\\&=\frac{1}{4}\sqrt{(\mathrm{Tr}[\tilde{X}V])^2+2\,\mathrm{Tr}[\tilde{X}V\tilde{X}V]+2\,\mathrm{Tr}[\Omega\tilde{X}\Omega\tilde{X}]}\\&\quad+\frac{1}{4}\|\sqrt{V+i\Omega}\tilde{X}\sqrt{V+i\Omega}\|_1\,,\end{aligned}\tag{3.11}$$

where $\tilde{X}\coloneqq\Omega X\Omega^{\intercal}$.

In the general case of *mixed* Gaussian states, we show that the trace norm of the derivative can be bounded in terms of the covariance matrix as follows (see Lemma 118 in the SM):

$$\left\|\frac{\mathrm{d}}{\mathrm{d}\alpha}\rho(V+\alpha X,\mathbf{m})\Big|_{\alpha=0}\right\|_1\leq\frac{1+\sqrt{3}}{4}\,\mathrm{Tr}[|X|\Omega^{\intercal}V\Omega]\,,\tag{3.12}$$

which is crucial for the proof of our main result in Theorem 95.

Previous studies have explored the derivatives of specific examples of Gaussian states (e.g., [56–58]). In contrast, our work introduces a comprehensive framework for defining the derivative of Gaussian states and establishes a general formula applicable to any Gaussian state (see Lemma 98). Additionally, another work [79], posted after the first version of this work, establishes a formula for the derivative of a Gaussian state with respect to its first moment and its Hamiltonian matrix [79], which is related to our analysis.

As additional technical contributions of independent interest, in Section 3.8.4 in the SM we discover new properties of the *Gaussian noise channel* [2, Chapter 5], which is widely recognised as one of the most important Gaussian channels. Moreover, we derive an exact formula for the second moment of a quadratic observable on a Gaussian state:

Lemma 100 (Second moment of a quadratic observable on a Gaussian state)**.** *Let $\rho(V,\mathbf{m})$ be an n-mode Gaussian state with covariance matrix V and first moment $\mathbf{m}$. Let X be an $2n\times 2n$ real symmetric matrix. Then, it holds that*

$$\mathrm{Tr}\left[(\hat{\mathbf{R}}^{\intercal}X\hat{\mathbf{R}})^2\,\rho(V,\mathbf{m})\right]\tag{3.13}$$

$$=\left(\frac{1}{2}\,\mathrm{Tr}[VX]+\mathbf{m}^{\intercal}X\mathbf{m}\right)^2+\frac{1}{2}\,\mathrm{Tr}[XVXV]\tag{3.14}$$

$$+\frac{1}{2}\,\mathrm{Tr}[\Omega X\Omega X]+2\mathbf{m}^{\intercal}XVX\mathbf{m}\,.\tag{3.15}$$

The proof is detailed in Lemma 107 in the SM. While this formula was previously established for the special case of Gaussian states with vanishing first moments [80], we present an independent proof that extends its validity to Gaussian states with non-zero first moments.

3.7.1 Alternative bounds

An improved upper bound, surpassing the one in Theorem 95, can be achieved by applying the *Williamson decomposition* to the covariance matrices [2]. Any covariance matrix V can be written in the Williamson decomposition as follows [2]:

$$V = SDS^\intercal\,, \tag{3.16}$$

where S is a *symplectic* matrix, i.e. a matrix satisfying $S\Omega S^\intercal = \Omega$, and D is a diagonal matrix of the form $D = \bigoplus_{j=1}^{n} \begin{pmatrix} d_j & 0 \\ 0 & d_j \end{pmatrix}$, with $d_1, d_2, \ldots, d_n \geq 1$.

Theorem 101 (Improved upper bound)**.** *Let $\rho(V, \mathbf{m}), \rho(W, \mathbf{t})$ be Gaussian states with covariance matrices V, W and first moments $\mathbf{m}, \mathbf{t}$, respectively. Moreover, let $V = S_1 D_1 S_1^\intercal$ and $W = S_2 D_2 S_2^\intercal$ be the Williamson decompositions of V and W, as in (3.16). Then, the following bound holds:*

$$\begin{aligned} &\frac{1}{2}\left\|\rho(V,\mathbf{m}) - \rho(W,\mathbf{t})\right\|_1 \\ &\quad\leq \frac{1+\sqrt{3}}{8}\operatorname{Tr}\left[|V-W|\Omega^\intercal\left(\frac{S_1S_1^\intercal + S_2S_2^\intercal}{2}\right)\Omega\right] \\ &\qquad + \frac{\min(\|S_1\|_\infty, \|S_2\|_\infty)}{\sqrt{2}}\|\mathbf{m}-\mathbf{t}\|_2\,. \end{aligned} \tag{3.17}$$

The proof can be found in Theorem 121 in the SM. Note that the bound in Theorem 101 is strictly stronger than the one in Theorem 95 since $S_1 S_1^\intercal \leq V$ and $S_2 S_2^\intercal \leq W$. Furthermore, the improvements are especially significant for highly mixed Gaussian states.

Additionally, in the forthcoming Lemma 102, proved in Section 3.8.4 in the SM, we derive an alternative upper bound that outperforms the above bounds in the regime where the moments are sufficiently far apart. Interestingly, this bound is always finite, even when the distance between the covariance matrices is very large.

Theorem 102 (Alternative upper bound)**.** *Let $\overline{\phi} : [0,\infty) \to \mathbb{R}^+$ be a concave function such that $\overline{\phi}(x) \geq \phi(x)$ for all $x \geq 0$, where the function $\phi : [0,\infty) \to [0,1]$ defined by*

$$\phi(x) := \frac{1-e^{-x}}{2} + e^{-x/2}\sqrt{\sinh^2(x/2) + 1 - e^{-\frac{x^2}{1+4x}}}\,. \tag{3.18}$$

Let $\rho(V, \mathbf{m})$ and $\rho(W, \mathbf{t})$ be Gaussian states with covariance matrices V, W and first moments $\mathbf{t}, \mathbf{m}$, respectively. Then, it holds that

$$\begin{aligned} &\frac{1}{2}\left\|\rho(V,\mathbf{m}) - \rho(W,\mathbf{t})\right\|_1 \\ &\quad\leq 2\,\overline{\phi}\left(\frac{1}{8}\max\left\{\|V\|_\infty, \|W\|_\infty\right\}\|V-W\|_1\right) \\ &\qquad + \sqrt{1 - e^{-\frac{1}{2}\min\{\|V\|_\infty, \|W\|_\infty\}\|\mathbf{t}-\mathbf{m}\|_2^2}}\,. \end{aligned} \tag{3.19}$$

By looking at a plot of the function ϕ, it seems clear that ϕ itself is concave. However, proving this explicitly seems quite difficult. If this could be done, then in the above theorem we could take $\overline{\phi} := \phi$.

3.8 Supplementary Material

The supplemental material is organised as follows. Section 3.8.1 provides essential preliminaries on continuous variable systems necessary for understanding the proofs of our results. In Section 3.8.2, we present our results on the upper bounds on the trace distance between Gaussian states, leveraging properties of the *derivative of a Gaussian state*. Section 3.8.3 rigorously defines the derivative of a Gaussian state from a functional analysis perspective, providing the foundation for certain steps used in Section 3.8.2. Moreover, Section 3.8.4 introduces an alternative, insightful method for obtaining upper bounds on the trace distance between Gaussian states by exploring new properties of the Gaussian noise channel, which may be of independent interest. Finally, in Section 3.8.5 we apply our trace-distance bounds to improve on the existing bounds on the sample complexity of tomography of Gaussian states [40],

3.8.1 Preliminaries on continuous variable systems

In this section, we briefly review the basics of quantum information with continuous variable systems [2]. By definition, a continuous variable system is a quantum system associated with the Hilbert space $L^2(\mathbb{R}^n)$, which comprises all square-integrable complex-valued functions over $\mathbb{R}^n$. Since $L^2(\mathbb{R}^n) = (L^2(\mathbb{R}))^{\otimes n}$, a continuous variable system can be viewed as comprising n subsystems, or *modes*. In this setting, n is called the *number of modes*, and a quantum state on $L^2(\mathbb{R}^n)$ is referred to as an *n-mode state*.

Let $\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n$ be the well-known position and momentum operators of each mode [2]. The *quadrature operator vector* is defined as

$$\hat{\mathbf{R}} := (\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n)^\intercal \tag{3.20}$$

and it satisfies the following relation, known as *canonical commutation relation*:

$$[\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal] = i\,\Omega\,\hat{\mathbb{1}}\,, \tag{3.21}$$

where

$$\Omega := \bigoplus_{i=1}^{n} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{3.22}$$

The characteristic function $\chi_\Theta : \mathbb{R}^{2n} \to \mathbb{C}$ of a bounded linear operator Θ on $L^2(\mathbb{R}^n)$ is given by

$$\chi_\Theta(\mathbf{r}) := \mathrm{Tr}[\Theta \hat{D}_{\mathbf{r}}]\,, \tag{3.23}$$

where the displacement operator is defined as

$$\hat{D}_{\mathbf{r}} := e^{-i\mathbf{r}^\intercal \Omega \hat{\mathbf{R}}}\,. \tag{3.24}$$

Any bounded linear operator Θ on $L^2(\mathbb{R}^n)$ can be expressed in terms of its characteristic function as [2, 81]:

$$\Theta = \mathscr{W}\left(\chi_\Theta(\mathbf{r})\right) := \int_{\mathbb{R}^{2n}} \frac{\mathrm{d}\mathbf{r}}{(2\pi)^n} \chi_\Theta(\mathbf{r}) \hat{D}_{\mathbf{r}}^\dagger\,, \tag{3.25}$$

which is known as the *Fourier-Weyl transform*. This establishes a one-to-one correspondence between bounded operators and their characteristic functions. The first moment $\mathbf{m}(\rho)$ and the covariance matrix $V(\rho)$ of a quantum state ρ are defined as follows:

$$\begin{aligned} \mathbf{m}(\rho) &:= \mathrm{Tr}\left[\hat{\mathbf{R}}\,\rho\right]\,, \\ V(\rho) &:= \mathrm{Tr}\left[\left\{(\hat{\mathbf{R}} - \mathbf{m}(\rho)\,\hat{\mathbb{1}}), (\hat{\mathbf{R}} - \mathbf{m}(\rho)\,\hat{\mathbb{1}})^\intercal\right\}\rho\right]\,, \end{aligned} \tag{3.26}$$

where $(\cdot)^\intercal$ denotes the transpose operation, $\{\hat{A}, \hat{B}\} \coloneqq \hat{A}\hat{B} + \hat{B}\hat{A}$ denotes the anti-commutator. Any covariance matrix $V(\rho)$ satisfies the following matrix inequality, known as *uncertainty relation*:

$$V(\rho) + i\Omega \geq 0\,. \tag{3.27}$$

As a consequence, since Ω is skew-symmetric, any covariance matrix $V(\rho)$ is positive semi-definite on $\mathbb{R}^{2n}$. Conversely, for any symmetric $2n \times 2n$ matrix W such that $W + i\Omega \geq 0$ there exists an n-mode (Gaussian) state ρ with covariance matrix $V(\rho) = W$ [2]. A displacement transformation shifts the first moments without affecting the covariance matrices, i.e.

$$\begin{aligned} \mathbf{m}\Big(\hat{D}_{\mathbf{r}}\rho\hat{D}_{\mathbf{r}}^\dagger\Big) &= \mathbf{m}(\rho) + \mathbf{r}\,, \\ V\Big(\hat{D}_{\mathbf{r}}\rho\hat{D}_{\mathbf{r}}^\dagger\Big) &= V(\rho)\,, \end{aligned} \tag{3.28}$$

as it follows from the fact that

$$\hat{D}_{\mathbf{r}}\hat{\mathbf{R}}\hat{D}_{\mathbf{r}}^\dagger = \hat{\mathbf{R}} + \mathbf{r}\hat{\mathbb{1}}\,. \tag{3.29}$$

Let us proceed with the definition of Gaussian states.

Definition 103 (Gaussian state)**.** *An n-mode state ρ is said to be a Gaussian state if it can be written as a Gibbs state of a quadratic Hamiltonian $\hat{H}$ in the quadrature operator vector:*

$$\hat{H} \coloneqq \frac{1}{2}(\hat{\mathbf{R}} - \mathbf{m})^\intercal H(\hat{\mathbf{R}} - \mathbf{m}) \tag{3.30}$$

for some $2n \times 2n$ symmetric positive-definite matrix H and some vector $\mathbf{m} \in \mathbb{R}^{2n}$. The Gibbs states associated with the Hamiltonian $\hat{H}$ are given by

$$\rho = \left(\frac{e^{-\beta\hat{H}}}{\mathrm{Tr}[e^{-\beta\hat{H}}]}\right)_{\beta\in(0,\infty]}\,, \tag{3.31}$$

where the parameter β is called the 'inverse temperature'.

Remark 104. *Definition 103 also covers pathological cases where both β and specific terms in H diverge (for instance, in tensor products of pure and mixed Gaussian states). To formalise this rigorously, one can define the set of Gaussian states as the trace-norm closure of the set of Gibbs states of quadratic Hamiltonians [82].*

The characteristic function of a Gaussian state ρ is the Fourier transform of a Gaussian probability distribution, evaluated at $\Omega\mathbf{r}$, which can be written in terms of $\mathbf{m}(\rho)$ and $V(\rho)$ as [2]

$$\chi_\rho(\mathbf{r}) = \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal V(\rho)\Omega\mathbf{r} + i(\Omega\mathbf{r})^\intercal\mathbf{m}(\rho)\right)\,. \tag{3.32}$$

Since any quantum state is uniquely identified by its characteristic function, it follows that any Gaussian state is uniquely identified by its first moment and covariance matrix.

Notably, any covariance matrix V can be written in the so-called *Williamson decomposition* as follows [2]:

$$V = SDS^\intercal\,, \tag{3.33}$$

where S is a *symplectic* matrix, i.e. a matrix satisfying $S\Omega S^\intercal = \Omega$, and D is a diagonal matrix of the form $D = \bigoplus_{j=1}^n \begin{pmatrix} d_j & 0 \\ 0 & d_j \end{pmatrix}$, with $d_1, d_2, \ldots, d_n \geq 1$ being the so-called *symplectic eigenvalues*.

A Gaussian state ρ is pure if and only if all the symplectic eigenvalues of its covariance matrix $V(\rho)$ are exactly one [2]. In particular, the Williamson decomposition of the covariance matrix of a pure Gaussian state ψ reads $V(\psi) = SS^\intercal$, for some symplectic matrix S.

Gaussian noise channel

The *Gaussian noise channel* is a Gaussian channel that will turn out to be crucial for our analysis. For any $2n \times 2n$ positive semi-definite matrix K, the *Gaussian noise channel* $\mathcal{N}_K$ is the convex combination of displacement transformations $(\cdot) \mapsto \hat{D}_u(\cdot)\hat{D}_u^\dagger$, where $u \in \mathbb{R}^{2n}$ is distributed according to a Gaussian probability distribution with vanishing mean value and covariance matrix equal to K [2, Chapter 5]. Mathematically, $\mathcal{N}_K$ can be written as

$$\mathcal{N}_K(\Theta) := \int_{\mathbb{R}^{2n}} \mathrm{d}u \; P_K(u)\, \hat{D}_u \Theta \hat{D}_u^\dagger \,, \tag{3.34}$$

where $P_K(u)$ is a Gaussian probability distribution with vanishing mean and covariance matrix equal to K, and $\hat{D}_u := e^{iu^\intercal \Omega \hat{\mathbf{R}}}$ is the displacement operator. Of course, in the case when K is strictly positive, $P_K(u)$ can be written as

$$P_K(u) := \frac{e^{-\frac{1}{2}u^\intercal K^{-1} u}}{(2\pi)^n \sqrt{\det K}} \,. \tag{3.35}$$

Instead, when K has some zero eigenvalues, the definition of $P_K(u)$ involves Dirac deltas as follows. Let $K = \sum_{i=1}^{2n} \lambda_i v_i v_i^\intercal$ be a spectral decomposition of K, where $(v_i)_{i=1,\dots,2n}$ are orthonormal eigenvectors, while $\lambda_1, \dots, \lambda_r > 0$ and $\lambda_{r+1}, \dots, \lambda_{2n} = 0$ are its eigenvalues, with r being the rank of K. Then, $P_K(u)$ can be written as

$$P_K(u) := \frac{e^{-\frac{1}{2}u^\intercal \left(\sum_{i=1}^{r} \lambda_i^{-1} v_i v_i^\intercal\right) u}}{(2\pi)^{(2n-r)/2} \sqrt{\lambda_1 \dots \lambda_r}} \prod_{i=r+1}^{2n} \delta(v_i^\intercal u) \,, \tag{3.36}$$

where $\delta(\cdot)$ denotes the Dirac delta distribution.

It turns out that $\mathcal{N}_K$ is a Gaussian channel that leaves the first moments unchanged, and it acts on the covariance matrix of the input state by adding K, as stated by the following Lemma [2, Chapter 5].

Lemma 105. ([2, Chapter 5]) *Let $K \in \mathbb{R}$ be a $2n \times 2n$ positive semi-definite matrix. The Gaussian noise channel $\mathcal{N}_K$, defined in (3.34), is a Gaussian channel that acts on the first moments and covariance matrices as*

$$\mathbf{m}(\mathcal{N}_K(\rho)) = \mathbf{m}(\rho) \,, \tag{3.37}$$
$$V(\mathcal{N}_K(\rho)) = V(\rho) + K \,, \tag{3.38}$$

for any n-mode input state ρ.

Lemma 105 directly implies the following useful known result [2].

Lemma 106. *Let ρ be a Gaussian state and let us consider the Williamson decomposition of its covariance matrix, $V(\rho) = SDS^\intercal$. Additionally, let ψ be the Gaussian state with the same first moment as ρ and with covariance matrix $SS^\intercal$. Then, ρ can be written as the output of the Gaussian noise channel $\mathcal{N}_{S(D-\mathbb{1})S^\intercal}$ applied to ψ:*

$$\rho = \mathcal{N}_{S(D-\mathbb{1})S^\intercal}(\psi) \,. \tag{3.39}$$

In particular, any mixed Gaussian state with covariance matrix V can be written as a convex combination of pure Gaussian states with all the same covariance matrix W, which satisfies the matrix inequality $W \leq V$.

3.8.2 Proof of our results

In this section, we address the problem of deriving upper bounds on the trace distance between Gaussian states in terms of the norm distance between their first moments and covariance matrices. The section is organised as follows:

- In Subsection 3.8.2, we derive an elegant formula for the second moment of a quadratic observable on a Gaussian state, expressed solely in terms of the covariance matrix and first moment.
- In Subsection 3.8.2, we prove that all odd moments of a Gaussian state with vanishing first moment are zero.
- In Subsection 3.8.2, we introduce the concept of the *derivative of a Gaussian state* and establish some useful properties. Specifically, we show that: (i) this derivative can be represented through a compact formula involving commutators between quadrature operators and the Gaussian state; (ii) we derive a simple upper bound on the trace norm of the derivative of a Gaussian state in terms of the first moment and covariance matrix; and (iii) we find an exact formula for the trace norm of the derivative of a *pure* Gaussian state in terms of its first moment and covariance matrix.
- In Subsection 3.8.2, leveraging the tools developed in the previous subsections, we prove upper bounds on the trace distance between Gaussian states and demonstrate that these bounds are tight.

Exact formula for the second moment of a quadratic observable on a Gaussian state

In this subsection, we derive a closed formula for the second moment of a quadratic observable on a Gaussian state in terms of its covariance matrix and its first moment. This formula was already known in the case of Gaussian states with vanishing first moment [80]. Here, we report an alternative proof.

Lemma 107 (Second moment of a quadratic observable on a Gaussian state)**.** *Let $\rho(V, \mathbf{m})$ be an n-mode Gaussian state with covariance matrix V and first moment $\mathbf{m}$. Let X be an $2n \times 2n$ real symmetric matrix. Then, it holds that*

$$\mathrm{Tr}\Big[(\hat{\mathbf{R}}^\intercal X \hat{\mathbf{R}})^2\, \rho(V, \mathbf{m})\Big] = \left(\frac{1}{2}\,\mathrm{Tr}[VX] + \mathbf{m}^\intercal X \mathbf{m}\right)^2 + \frac{1}{2}\,\mathrm{Tr}[XVXV] + \frac{1}{2}\,\mathrm{Tr}[\Omega X \Omega X] + 2\mathbf{m}^\intercal XVX\mathbf{m}\,. \tag{3.40}$$

Before proving Lemma 107, let us establish some preliminary results.

Lemma 108. *Let ρ be an n-mode state and let X be a $2n \times 2n$ real symmetric matrix. Then, it holds that*

$$\mathrm{Tr}\Big[(\hat{\mathbf{R}}^\intercal X \hat{\mathbf{R}})^2 \rho\Big] = \frac{1}{8} \sum_{i,j,l,k=1}^{2n} X_{ij} X_{kl}\, \mathrm{Tr}\Big[\{\hat{R}_i, \{\hat{R}_j, \{\hat{R}_k, \hat{R}_l\}\}\}\rho\Big] + \frac{1}{2}\,\mathrm{Tr}[\Omega X \Omega X] \tag{3.41}$$

Proof. Note that

$$\mathrm{Tr}\Big[(\hat{\mathbf{R}}^\intercal X \hat{\mathbf{R}})^2 \rho\Big] = \sum_{i,j,l,k=1}^{2n} X_{ij} X_{kl}\, \mathrm{Tr}\Big[\hat{R}_i \hat{R}_j \hat{R}_k \hat{R}_l \rho\Big]\,. \tag{3.42}$$

By exploiting that X is symmetric, we have that

$$\begin{aligned}
&\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\{\hat{R}_i,\{\hat{R}_j,\{\hat{R}_k,\hat{R}_l\}\}\}\rho\Big] \\
&= 2\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\{\hat{R}_i,\{\hat{R}_j,\hat{R}_k\hat{R}_l\}\}\rho\Big] \\
&= 2\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_j\hat{R}_k\hat{R}_l\rho\Big] + 2\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\hat{R}_l\hat{R}_j\rho\Big] \\
&\quad + 2\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_j\hat{R}_k\hat{R}_l\hat{R}_i\rho\Big] + 2\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_k\hat{R}_l\hat{R}_j\hat{R}_i\rho\Big] \\
&= 4\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_j\hat{R}_k\hat{R}_l\rho\Big] + 4\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\hat{R}_l\hat{R}_j\rho\Big]\,.
\end{aligned} \tag{3.43}$$

Moreover, the canonical commutation relation implies that

$$\begin{aligned}
&\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\hat{R}_l\hat{R}_j\rho\Big] \\
&= \operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\hat{R}_j\hat{R}_l\rho\Big] + i\Omega_{lj}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\rho\Big] \\
&= \operatorname{Tr}\Big[\hat{R}_i\hat{R}_j\hat{R}_k\hat{R}_l\rho\Big] + i\Omega_{kj}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_l\rho\Big] + i\Omega_{lj}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\rho\Big]\,.
\end{aligned} \tag{3.44}$$

Using the fact that X is symmetric and Ω is skew-symmetric, we get

$$\begin{aligned}
&i\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\left(\Omega_{kj}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_l\rho\Big] + \Omega_{lj}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\rho\Big]\right) \\
&= i\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\left(\Omega_{jl}\operatorname{Tr}\Big[\hat{R}_k\hat{R}_i\rho\Big] + \Omega_{lj}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\rho\Big]\right) \\
&= i\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\Omega_{jl}\left(\operatorname{Tr}\Big[[\hat{R}_k,\hat{R}_i]\rho\Big]\right) \\
&= -\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\Omega_{jl}\Omega_{ki} \\
&= -\operatorname{Tr}[X\Omega X\Omega]\,.
\end{aligned} \tag{3.45}$$

Consequently, we have that

$$\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_k\hat{R}_l\hat{R}_j\rho\Big] = \sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_j\hat{R}_k\hat{R}_l\rho\Big] - \operatorname{Tr}[X\Omega X\Omega] \tag{3.46}$$

By substituting in (3.43), we finally obtain

$$\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\{\hat{R}_i,\{\hat{R}_j,\{\hat{R}_k,\hat{R}_l\}\}\}\rho\Big] = 8\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\Big[\hat{R}_i\hat{R}_j\hat{R}_k\hat{R}_l\rho\Big] - 4\operatorname{Tr}[X\Omega X\Omega]\,, \tag{3.47}$$

which, after rearranging, proves the desired result. □

Lemma 109. *Let $j \in \{1,2,\dots,2n\}$. The partial derivative of the displacement operator $\hat{D}_{\mathbf{r}} := e^{-i\mathbf{r}^\intercal\Omega\hat{R}}$ with respect to the variable r_j is given by*

$$\frac{\partial}{\partial r_j}\hat{D}_{\mathbf{r}} = -\frac{i}{2}\{(\Omega\hat{\mathbf{R}})_j,\hat{D}_{\mathbf{r}}\}. \tag{3.48}$$

Proof. Since the displacement operator is the following tensor product of single-mode displacement operators

$$\hat{D}_{\mathbf{r}} \coloneqq \hat{D}_{(r_1,r_2)} \otimes \hat{D}_{(r_3,r_4)} \otimes \ldots \otimes \hat{D}_{(r_{2n-1},r_{2n})}\,, \tag{3.49}$$

it suffices to consider the single-mode case $n = 1$. By exploiting that

$$e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}\,. \tag{3.50}$$

for any A, B such that $[A, B] \propto \mathbb{1}$, we obtain that

$$\begin{aligned}\frac{\partial}{\partial r_1}\hat{D}_{\mathbf{r}} &= \frac{\partial}{\partial r_1}\left[e^{-ir_1\hat{R}_2+ir_2\hat{R}_1}\right] \\ &= \frac{\partial}{\partial r_1}\left[e^{-ir_1\hat{R}_2}e^{ir_2\hat{R}_1}e^{-i\frac{1}{2}r_1r_2}\right] \\ &= -i\hat{R}_2\hat{D}_{\mathbf{r}} - i\frac{1}{2}r_2\hat{D}_{\mathbf{r}}\,.\end{aligned} \tag{3.51}$$

Similarly, by using that $e^{A+B} = e^B e^A e^{\frac{1}{2}[A,B]}$ for any A, B such that $[A, B] \propto \mathbb{1}$, we also obtain that

$$\begin{aligned}\frac{\partial}{\partial r_1}\hat{D}_{\mathbf{r}} &= \frac{\partial}{\partial r_1}\left[e^{-ir_1\hat{R}_2+ir_2\hat{R}_1}\right] \\ &= \frac{\partial}{\partial r_1}\left[e^{ir_2\hat{R}_1}e^{-ir_1\hat{R}_2}e^{i\frac{1}{2}r_1r_2}\right] \\ &= -i\hat{D}_{\mathbf{r}}\hat{R}_2 + i\frac{1}{2}r_2\hat{D}_{\mathbf{r}}\,.\end{aligned} \tag{3.52}$$

By adding both the above identities, we obtain that

$$\frac{\partial}{\partial r_1}\hat{D}_{\mathbf{r}} = -\frac{i}{2}\{\hat{R}_2, \hat{D}_{\mathbf{r}}\}\,. \tag{3.53}$$

Analogously, we obtain that

$$\begin{aligned}\frac{\partial}{\partial r_2}\hat{D}_{\mathbf{r}} &= \frac{\partial}{\partial r_2}\left[e^{-ir_1\hat{R}_2+ir_2\hat{R}_1}\right] \\ &= \frac{\partial}{\partial r_2}\left[e^{-ir_1\hat{R}_2}e^{ir_2\hat{R}_1}e^{-i\frac{1}{2}r_1r_2}\right] \\ &= i\hat{D}_{\mathbf{r}}\hat{R}_1 - i\frac{1}{2}r_1\hat{D}_{\mathbf{r}}\end{aligned} \tag{3.54}$$

and that

$$\begin{aligned}\frac{\partial}{\partial r_2}\hat{D}_{\mathbf{r}} &= \frac{\partial}{\partial r_2}\left[e^{-ir_1\hat{R}_2+ir_2\hat{R}_1}\right] \\ &= \frac{\partial}{\partial r_2}\left[e^{ir_2\hat{R}_1}e^{-ir_1\hat{R}_2}e^{i\frac{1}{2}r_1r_2}\right] \\ &= i\hat{R}_1\hat{D}_{\mathbf{r}} + i\frac{1}{2}r_1\hat{D}_{\mathbf{r}}\,.\end{aligned} \tag{3.55}$$

By adding these two expressions, we get

$$\frac{\partial}{\partial r_2}\hat{D}_{\mathbf{r}} = \frac{i}{2}\{\hat{R}_1, \hat{D}_{\mathbf{r}}\}\,. \tag{3.56}$$

Note that (3.53) and (3.56) can be rewritten in a compact form as

$$\frac{\partial}{\partial r_j}\hat{D}_{\mathbf{r}} = -\frac{i}{2}\{(\Omega\hat{\mathbf{R}})_j, \hat{D}_{\mathbf{r}}\}\,. \tag{3.57}$$

This concludes the proof. □

As an example, Lemma 109 is useful to calculate the laplacian $\triangle := \sum_{j=1}^{2n} \frac{\partial^2}{\partial r_j^2}$ of the displacement operator. Indeed, by exploiting such a lemma, one can prove that

$$\triangle \hat{D}_{\mathbf{r}} = -\left(\hat{\mathbf{R}}^\intercal \hat{\mathbf{R}} - \mathbf{r}^\intercal \hat{\mathbf{R}} + \frac{1}{4}\mathbf{r}^\intercal \mathbf{r}\, \mathbb{1}\right) \hat{D}_{\mathbf{r}} \,. \tag{3.58}$$

Moreover, Lemma 109 directly implies a simple expression for the derivative of the characteristic function.

Lemma 110 (Derivative of a characteristic function)**.** *Let Θ be a linear operator on $L^2(\mathbb{R}^n)$ and let $j \in \{1, 2, \dots, 2n\}$. The partial derivative of the characteristic function $\chi_\Theta(\mathbf{r}) := \mathrm{Tr}[\Theta e^{-i\mathbf{r}^\intercal \Omega \hat{R}}]$ with respect to the variable r_j is given by*

$$\frac{\partial}{\partial r_j}\chi_\Theta(\mathbf{r}) = -\frac{i}{2}\chi_{\{\Theta,(\Omega\hat{\mathbf{R}})_j\}}(\boldsymbol{r}) \qquad \forall\, \boldsymbol{r} \in \mathbb{R}^{2n} \,, \tag{3.59}$$

i.e. it is proportional to the characteristic function of the anti-commutator between Θ itself and the quadrature $(\Omega\hat{\boldsymbol{R}})_j$.

Lemma 110 is useful in order to calculate (symmetrised) moments of a quantum state given its characteristic function. This is true because the characteristic function of an operator evaluated in zero is exactly equal to the trace of the operator. As an example, Lemma 110 implies that

$$\mathrm{Tr}\Big[\{(\Omega\hat{\mathbf{R}})_k, \{(\Omega\hat{\mathbf{R}})_l, \Theta\}\}\Big] = -4\,\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_\Theta(0) \,. \tag{3.60}$$

Note that, in general, the moment $\mathrm{Tr}[\Theta \hat{R}_k \hat{R}_l]$ cannot be proportional to the second derivative $\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_\Theta(0)$. This discrepancy arises because $\mathrm{Tr}[\Theta \hat{R}_k \hat{R}_l]$ lacks symmetry under the exchange $k \leftrightarrow l$, whereas the second derivative $\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_\Theta(0)$ is symmetric under such an exchange (as partial derivatives commute). Consequently, a fully symmetrised structure — such as one involving a nested anti-commutator as in (3.60) — is necessary.

Lemma 111. *Let $\rho(V, \mathbf{m})$ be a Gaussian state with covariance matrix V and first moment $\mathbf{m}$. Then, for any $i, j, k, l \in \{1, 2, \dots, 2n\}$, it holds that*

$$\begin{aligned}\mathrm{Tr}[\rho(V,\mathbf{m})\,\{\hat{R}_i, \{\hat{R}_j, \{\hat{R}_k, \hat{R}_l\}\}\}] = {} & 2(V_{ij}V_{kl} + V_{ik}V_{jl} + V_{il}V_{jk}) \\ & + 4\,(V_{kl}m_i m_j + V_{jl}m_i m_k + V_{kj}m_i m_l + V_{il}m_j m_k + V_{ik}m_j m_l + V_{ij}m_k m_l) \\ & + 8 m_i m_j m_k m_l \,.\end{aligned} \tag{3.61}$$

Proof. On the one hand, by performing an explicit calculation of the partial derivatives of the characteristic function

$$\chi_{\rho(V,\mathbf{m})}(\mathbf{r}) = e^{-\frac{1}{4}(\Omega\mathbf{r})^\intercal V \Omega \mathbf{r}} \,, \tag{3.62}$$

we obtain

$$\begin{aligned}\frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{\rho(V,0)}(\mathbf{r})|_{\mathbf{r}=0} = {} & \frac{1}{4}(\tilde{V}_{ij}\tilde{V}_{kl} + \tilde{V}_{ik}\tilde{V}_{jl} + \tilde{V}_{il}\tilde{V}_{jk}) \\ & + \frac{1}{2}\left(\tilde{V}_{kl}\tilde{m}_i\tilde{m}_j + \tilde{V}_{jl}\tilde{m}_i\tilde{m}_k + \tilde{V}_{kj}\tilde{m}_i\tilde{m}_l + \tilde{V}_{il}\tilde{m}_j\tilde{m}_k + \tilde{V}_{ik}\tilde{m}_j\tilde{m}_l + \tilde{V}_{ij}\tilde{m}_k\tilde{m}_l\right) \\ & + \tilde{m}_i\tilde{m}_j\tilde{m}_k\tilde{m}_l \,,\end{aligned} \tag{3.63}$$

where

$$\begin{aligned}\tilde{V} &:= \Omega^\intercal V \Omega \,, \\ \tilde{\mathbf{m}} &:= \Omega \mathbf{m} \,.\end{aligned} \tag{3.64}$$

In particular, it also holds that

$$\begin{aligned}
&\sum_{i,j,k,l=1}^{n} \Omega_{ii'}\Omega_{jj'}\Omega_{kk'}\Omega_{ll'} \frac{\partial}{\partial r_{i'}}\frac{\partial}{\partial r_{j'}}\frac{\partial}{\partial r_{k'}}\frac{\partial}{\partial r_{l'}} \chi_{\rho(V,\mathbf{m})}(\mathbf{r})|_{\mathbf{r}=0} \\
&\quad = \frac{1}{4}(V_{ij}V_{kl} + V_{ik}V_{jl} + V_{il}V_{jk}) \\
&\qquad + \frac{1}{2}\left(V_{kl}m_i m_j + V_{jl}m_i m_k + V_{kj}m_i m_l + V_{il}m_j m_k + V_{ik}m_j m_l + V_{ij}m_k m_l\right) \\
&\qquad + m_i m_j m_k m_l \,.
\end{aligned} \tag{3.65}$$

On the other hand, Lemma 109 implies that

$$\begin{aligned}
\frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{\rho(V,0)}(\mathbf{r}) &= \operatorname{Tr}\left[\rho(V,0)\frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\hat{D}_{\mathbf{r}}\right] \\
&= \frac{1}{16}\operatorname{Tr}\left[\rho(V,0)\left\{(\Omega\hat{\mathbf{R}})_i,\{(\Omega\hat{\mathbf{R}})_j,\{(\Omega\hat{\mathbf{R}})_k,\{(\Omega\hat{\mathbf{R}})_l,\hat{D}_{\mathbf{r}}\}\}\}\right\}\right] .
\end{aligned} \tag{3.66}$$

Consequently, by exploiting that $\hat{D}_0 = \hat{\mathbb{1}}$, we obtain that

$$\frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{\rho(V,0)}(\mathbf{r})|_{\mathbf{r}=0} = \frac{1}{8}\operatorname{Tr}[\rho(V,0)\left\{(\Omega\hat{\mathbf{R}})_i,\{(\Omega\hat{\mathbf{R}})_j,\{(\Omega\hat{\mathbf{R}})_k,(\Omega\hat{\mathbf{R}})_l\}\}\}] \,. \tag{3.67}$$

In particular, we get

$$\sum_{i,j,k,l=1}^{n} \Omega_{ii'}\Omega_{jj'}\Omega_{kk'}\Omega_{ll'} \frac{\partial}{\partial r_{i'}}\frac{\partial}{\partial r_{j'}}\frac{\partial}{\partial r_{k'}}\frac{\partial}{\partial r_{l'}} \chi_{\rho(V,0)}(\mathbf{r})|_{\mathbf{r}=0} = \frac{1}{8}\operatorname{Tr}[\rho(V,0)\left\{\hat{R}_i,\{\hat{R}_j,\{\hat{R}_k,\hat{R}_l\}\}\}] \,. \tag{3.68}$$

By putting (3.65) and (3.68) together, we conclude the proof. □

Now, we are now ready to prove Lemma 107.

Proof of Lemma 107. It holds that

$$\begin{aligned}
\operatorname{Tr}\left[(\hat{\mathbf{R}}^\intercal X\hat{\mathbf{R}})^2\rho(V,\mathbf{m})\right] &\overset{\text{(i)}}{=} \frac{1}{2}\operatorname{Tr}[\Omega X\Omega X] + \frac{1}{8}\sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\operatorname{Tr}\left[\{\hat{R}_i,\{\hat{R}_j,\{\hat{R}_k,\hat{R}_l\}\}\}\rho(V,\mathbf{m})\right] \\
&\overset{\text{(ii)}}{=} \frac{1}{2}\operatorname{Tr}[\Omega X\Omega X] + \sum_{i,j,l,k=1}^{2n} X_{ij}X_{kl}\left(\frac{1}{4}(V_{ij}V_{kl} + V_{ik}V_{jl} + V_{il}V_{jk})\right. \\
&\qquad + \frac{1}{2}\left(V_{kl}m_i m_j + V_{jl}m_i m_k + V_{kj}m_i m_l + V_{il}m_j m_k + V_{ik}m_j m_l + V_{ij}m_k m_l\right) \\
&\qquad \left. + m_i m_j m_k m_l\right) \\
&= \frac{1}{2}\operatorname{Tr}[\Omega X\Omega X] + \frac{1}{4}(\operatorname{Tr}[VX])^2 + \frac{1}{2}\operatorname{Tr}[XVXV] \\
&\quad + \mathbf{m}^\intercal X\mathbf{m}\operatorname{Tr}[VX] + 2\mathbf{m}^\intercal XVX\mathbf{m} + (\mathbf{m}^\intercal X\mathbf{m})^2 \\
&= \left(\frac{1}{2}\operatorname{Tr}[VX] + \mathbf{m}^\intercal X\mathbf{m}\right)^2 + \frac{1}{2}\operatorname{Tr}[XVXV] + \frac{1}{2}\operatorname{Tr}[\Omega X\Omega X] + 2\mathbf{m}^\intercal XVX\mathbf{m} \,.
\end{aligned} \tag{3.69}$$

Here, in (i), we exploited Lemma 108, and in (ii), we applied Lemma 111. □

A Gaussian state with vanishing first moment also has vanishing odd moments

Thanks to the tools regarding developed above, we can easily show the following rather intuitive result.

Lemma 112 (A Gaussian state with vanishing first moment also has vanishing third moments)**.** *Let $\rho(V,0)$ be a Gaussian state with zero first moment and covariance matrix V. Then, for any $j,k,l \in \{1,2,\dots,2n\}$, it holds that*

$$\mathrm{Tr}[\rho(V,0)\,\hat{R}_j\hat{R}_k\hat{R}_l] = 0\,. \tag{3.70}$$

Proof. On the one hand, since the characteristic function of $\rho(V,0)$ is given by

$$\chi_{\rho(V,0)}(\mathbf{r}) = e^{-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}V\Omega\mathbf{r}}\,, \tag{3.71}$$

it holds that

$$\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{\rho(V,0)}(\mathbf{r})|_{\mathbf{r}=0} = 0\,. \tag{3.72}$$

On the other hand, Lemma 109 implies that

$$\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{\rho(V,0)}(\mathbf{r}) = \mathrm{Tr}\left[\rho(V,0)\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\hat{D}_{\mathbf{r}}\right] = \frac{i}{8}\,\mathrm{Tr}\Big[\rho(V,0)\,\{(\Omega\hat{\mathbf{R}})_j,\{(\Omega\hat{\mathbf{R}})_k,\{(\Omega\hat{\mathbf{R}})_l,\hat{D}_{\mathbf{r}}\}\}\}\Big]\,. \tag{3.73}$$

Consequently, by exploiting that $\hat{D}_0 = \hat{\mathbb{1}}$, we obtain that

$$\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{\rho(V,0)}(\mathbf{r})|_{\mathbf{r}=0} = \frac{i}{4}\,\mathrm{Tr}[\rho(V,0)\,\{(\Omega\hat{\mathbf{R}})_j,\{(\Omega\hat{\mathbf{R}})_k,(\Omega\hat{\mathbf{R}})_l\}\}]\,. \tag{3.74}$$

In particular, we get

$$\sum_{j,k,l=1}^{n}\Omega_{jj'}\Omega_{kk'}\Omega_{ll'}\frac{\partial}{\partial r_j}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{\rho(V,0)}(\mathbf{r})|_{\mathbf{r}=0} = \frac{i}{4}\,\mathrm{Tr}[\rho(V,0)\,\{\hat{R}_j,\{\hat{R}_k,\hat{R}_l\}\}]\,. \tag{3.75}$$

By putting (3.72) and (3.75) together, we obtain that

$$\mathrm{Tr}[\rho(V,0)\,\{\hat{R}_j,\{\hat{R}_k,\hat{R}_l\}\}] = 0\,. \tag{3.76}$$

Finally, note that

$$\begin{aligned}\mathrm{Tr}[\rho(V,0)\,\hat{R}_j\hat{R}_k\hat{R}_l] &= \frac{\mathrm{Tr}[\rho(V,0)\,\hat{R}_j\{\hat{R}_k,\hat{R}_l\}]}{2}\\ &= \frac{\mathrm{Tr}[\rho(V,0)\,\{\hat{R}_j\{\hat{R}_k,\hat{R}_l\}\}]}{4}\\ &= 0\,,\end{aligned} \tag{3.77}$$

where the two equalities follow from the canonical commutation relation and from the fact that $\rho(V,0)$ has zero first moment, while the last equality follows from (3.76). □

It is worth noting that the approach used in the lemma above can be easily generalised to show that *any* odd moment of a Gaussian state with zero first moment vanishes. However, in order to prove such a more general result, we use a much simpler approach.

Lemma 113. *A Gaussian state with vanishing first moments has also vanishing odd moments.*

Proof. We need to show that the overlap between a Gaussian state with vanishing first moment and a product of an odd number of quadratures vanishes.

To this end, note that any Gaussian unitary transformation preserves the degree of any polynomial in the quadratures. Hence, thanks to Williamson decomposition, we can assume without loss of generality that the Gaussian state is a tensor product of thermal states. Additionally, taking the trace of each mode, we can further assume without loss of generality the Gaussian state to be a single-mode thermal state. Thus, it suffices to demonstrate that any single-mode thermal state has vanishing odd moments.

Since thermal states are invariant under phase-space inversion and the product of an odd number of quadratures changes sign under such inversion, it follows that the overlap between a single-mode thermal state and a product of an odd number of quadratures must vanish. □

Derivative of a Gaussian state

In this subsection, we first define the *derivative of a Gaussian state* and then prove relevant properties. Throughout the rest of the subsection, we will use the following notation. For any $2n \times 2n$ real positive definite matrix V and any real $2n$-dimensional vector $\mathbf{m}$, we define the operator $\rho(V, \mathbf{m})$ as

$$\rho(V, \mathbf{m}) := \int_{\mathbb{R}^{2n}} \frac{\mathrm{d}\mathbf{r}}{(2\pi)^n} e^{-\frac{1}{4}(\Omega\mathbf{r})^\intercal V \Omega \mathbf{r} - i\mathbf{r}^\intercal \Omega \mathbf{m}} \hat{D}_{\mathbf{r}}^\dagger \,. \tag{3.78}$$

If V satisfies the uncertainty principle $V + i\Omega \geq 0$, the operator $\rho(V, \mathbf{m})$ is exactly equal to the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. This follows from the Fourier-Weyl relation in (3.25) and from the expression of the characteristic function of a Gaussian state in (3.32). For completeness, note that if V does not satisfy the uncertainty principle, the operator $\rho(V, \mathbf{m})$ fails to be positive semi-definite (this can be seen e.g. by contradiction and by observing that the covariance matrix of $\rho(V, \mathbf{m})$ is exactly V), and hence it is not a quantum state.

We begin with the definition of the directional derivative of a Gaussian state.

Definition 114 (Derivative of a Gaussian state)**.** *Let $\rho(V, \mathbf{m})$ be an n-mode Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Let X be a $2n \times 2n$ real symmetric matrix X and let $\mathbf{x}$ be a real $2n$-dimensional vector. We define the directional derivative of the Gaussian state $\rho(V, \mathbf{m})$ along the direction $(X, \mathbf{x})$ as*

$$\partial_{X,\mathbf{x}} \rho(V, \mathbf{m}) := \lim_{\alpha \to 0^+} \frac{\rho(V + \alpha X, \mathbf{m} + \alpha \mathbf{x}) - \rho(V, \mathbf{m})}{\alpha} \,, \tag{3.79}$$

where the limit is taken with respect to the trace norm.

Note that since V is positive definite, the matrix $V + \alpha X$ is positive definite for sufficiently small values of $\alpha > 0$, and thus the operator $\rho(V + \alpha X, \mathbf{m} + \alpha \mathbf{x})$ in (3.79) is well defined. The existence of this limit is guaranteed by the fact that, for any V and $\mathbf{m}$, the state $\rho(V, \mathbf{m})$ is continuously differentiable, as proven in Theorem 135 of Section 3.8.3.

Compact formula for the derivative of a Gaussian state Here, we prove a compact formula for the derivative of a Gaussian state.

Theorem 115 (Compact formula for the derivative of a Gaussian state)**.** *The derivative of the Gaussian state $\rho(V, \mathbf{m})$ along the direction $(X, \mathbf{x})$ can be expressed as*

$$\partial_{X,\mathbf{x}} \rho(V, \mathbf{m}) = -\frac{1}{4} \sum_{i,j=1}^{2n} \tilde{X}_{ij} [\hat{R}_i, [\hat{R}_j, \rho(V, \mathbf{m})]] + i \sum_{j=1}^{2n} \tilde{x}_j [\hat{R}_j, \rho(V, \mathbf{m})], \tag{3.80}$$

where

$$\tilde{X} := \Omega^\intercal X \Omega, \tag{3.81}$$

$$\tilde{\mathbf{x}} := \Omega \mathbf{x}. \tag{3.82}$$

Before proving Theorem 115, let us prove some preliminary results.

Lemma 116. *Let $j \in \{1, 2, \dots, 2n\}$ and $\mathbf{r} \in \mathbb{R}^{2n}$. The commutator between the quadrature operator $\hat{R}_i$ and the displacement operator $\hat{D}_{\mathbf{r}} := e^{-i\mathbf{r}^\intercal \Omega \hat{R}}$ can be expressed as*

$$[\hat{R}_j, \hat{D}_{\mathbf{r}}] = r_j \hat{D}_{\mathbf{r}}\,, \tag{3.83}$$

where r_j is the jth component of the vector $\mathbf{r}$.

Proof. It holds that

$$\begin{aligned}
[\hat{R}_j, \hat{D}_{\mathbf{r}}] &= [\hat{R}_j, e^{-i\mathbf{r}^\intercal \Omega \hat{R}}] \\
&= \sum_{k=1}^{\infty} \frac{1}{k!} \left[\hat{R}_j, \left(-i\mathbf{r}^\intercal \Omega \hat{R}\right)^k\right] \\
&\overset{\text{(i)}}{=} \left[\hat{R}_j, -i\mathbf{r}^\intercal \Omega \hat{R}\right] \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \left(-i\mathbf{r}^\intercal \Omega \hat{R}\right)^{k-1} \\
&= \left[\hat{R}_j, -i\mathbf{r}^\intercal \Omega \hat{R}\right] \hat{D}_{\mathbf{r}}, \\
&\overset{\text{(ii)}}{=} r_j \hat{D}_{\mathbf{r}}\,.
\end{aligned} \tag{3.84}$$

Here, in (i) we exploited that

$$\left[A, B^k\right] = k[A, B]B^{k-1} \tag{3.85}$$

for any A, B satisfying $[A, [A, B]] = [B, [A, B]] = 0$. Moreover, in (ii), by leveraging the canonical commutation relation $[\hat{R}_j, \hat{R}_k] = i\Omega_{jk}$ and the fact that $\Omega\Omega^\intercal = \mathbb{1}$, we observed that

$$\begin{aligned}
\left[\hat{R}_j, -i\mathbf{r}^\intercal \Omega \hat{R}\right] &= -i \sum_{l,k=1}^{2n} r_l \Omega_{lk} \left[\hat{R}_j, \hat{R}_k\right] \\
&= \sum_{l,k=1}^{2n} r_l \Omega_{lk} \Omega_{jk} \hat{\mathbb{1}} \\
&= \sum_{l=1}^{2n} r_l \delta_{lj} \hat{\mathbb{1}} \\
&= r_j \hat{\mathbb{1}}\,.
\end{aligned} \tag{3.86}$$

□

Lemma 117. *Let $i \in \{1, 2, \dots, 2n\}$. Let Θ be a bounded linear operator on $L^2(\mathbb{R}^{2n})$ such that $[\hat{R}_i, \Theta]$ is also bounded. Then, it holds that*

$$\int_{\mathbb{R}^{2n}} \frac{\mathrm{d}\mathbf{r}}{(2\pi)^n} r_i \chi_\Theta(\mathbf{r}) \hat{D}_{\mathbf{r}}^\dagger = [\Theta, \hat{R}_i]\,. \tag{3.87}$$

Proof. Note that the characteristic function of $[\Theta, \hat{R}_i]$ satisfies

$$\chi_{[\Theta, \hat{R}_i]}(\mathbf{r}) = \mathrm{Tr}\left[[\Theta, \hat{R}_i]\hat{D}_{\mathbf{r}}\right] = \mathrm{Tr}\left[\Theta[\hat{R}_i, \hat{D}_{\mathbf{r}}]\right] \overset{\text{(i)}}{=} r_i \,\mathrm{Tr}\left[\Theta \hat{D}_{\mathbf{r}}\right] = r_i \chi_\Theta(\mathbf{r})\,, \tag{3.88}$$

where in (i) we exploited Lemma 116. Hence, the Fourier-Weyl relation in (3.25) concludes the proof. □

Now, we are ready to prove Theorem 115.

Proof of Theorem 115. By exploiting (3.78) and the expression of the characteristic function of a Gaussian state in (3.32), we have that

$$\rho(V+\alpha X,\mathbf{m}+\alpha\mathbf{x})=\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})e^{-\alpha\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r}-i\alpha\mathbf{r}^\intercal\Omega\mathbf{x}}\hat{D}_{\mathbf{r}}^\dagger\,. \tag{3.89}$$

Consequently, it holds that

$$\begin{aligned}
\partial_{X,\mathbf{x}}\,\rho(V,\mathbf{m}) &= \lim_{\alpha\to0^+}\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})\left(\frac{e^{-\alpha\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r}-i\alpha\mathbf{r}^\intercal\Omega\mathbf{x}}-1}{\alpha}\right)\hat{D}_{\mathbf{r}}^\dagger\\
&\overset{\text{(i)}}{=}\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})\left(\lim_{\alpha\to0^+}\frac{e^{-\alpha\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r}-i\alpha\mathbf{r}^\intercal\Omega\mathbf{x}}-1}{\alpha}\right)\hat{D}_{\mathbf{r}}^\dagger\\
&=\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r}-i\mathbf{r}^\intercal\Omega\mathbf{x}\right)\hat{D}_{\mathbf{r}}^\dagger\\
&=-\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})\frac{1}{4}\mathbf{r}^\intercal\tilde{X}\mathbf{r}\,\hat{D}_{\mathbf{r}}^\dagger-i\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})\mathbf{r}^\intercal\tilde{\mathbf{x}}\,\hat{D}_{\mathbf{r}}^\dagger\\
&=-\frac{1}{4}\sum_{i,j=1}^{2n}\tilde{X}_{ij}\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})r_ir_j\hat{D}_{\mathbf{r}}^\dagger-i\sum_{j=1}^{2n}\tilde{x}_j\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})r_j\hat{D}_{\mathbf{r}}^\dagger\,.
\end{aligned} \tag{3.90}$$

Here, the equality (i) is justified by the differentiability of Gaussian states proved in Theorem 135. Additionally, Lemma 117 guarantees that

$$\begin{aligned}
\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})r_j\hat{D}_{\mathbf{r}}^\dagger&=[\rho(V,\mathbf{m}),\hat{R}_j]\,,\\
\int_{\mathbb{R}^{2n}}\frac{\mathrm{d}\mathbf{r}}{(2\pi)^n}\chi_{\rho(V,\mathbf{m})}(\mathbf{r})r_ir_j\hat{D}_{\mathbf{r}}^\dagger&=[[\rho(V,\mathbf{m}),\hat{R}_i],\hat{R}_j]\,.
\end{aligned} \tag{3.91}$$

The assumptions of Lemma 117 are satisfied because $[\rho(V,\mathbf{m}),\hat{R}_i]$ and $[[\rho(V,\mathbf{m}),\hat{R}_i],\hat{R}_j]$ are bounded for any $i,j\in\{1,2,\dots,2n\}$, as a consequence of Proposition 3.18 in the arXiv version of [83]. In particular, we conclude that

$$\begin{aligned}
\partial_{X,\mathbf{x}}\,\rho(V,\mathbf{m})&=-\frac{1}{4}\sum_{i,j=1}^{2n}\tilde{X}_{ij}[[\rho(V,\mathbf{m}),\hat{R}_j],\hat{R}_i]+i\sum_{j=1}^{2n}\tilde{x}_j[\hat{R}_j,\rho(V,\mathbf{m})]\\
&=-\frac{1}{4}\sum_{i,j=1}^{2n}\tilde{X}_{ij}[\hat{R}_i,[\hat{R}_j,\rho(V,\mathbf{m})]]+i\sum_{j=1}^{2n}\tilde{x}_j[\hat{R}_j,\rho(V,\mathbf{m})]\,.
\end{aligned} \tag{3.92}$$

□

Upper bound on the trace norm of the derivative of a Gaussian state Here, we derive a simple upper bound on the trace norm of the derivative of a Gaussian state.

Lemma 118 (Upper bound on the trace norm of the derivative of a Gaussian state)**.** *Let $\rho(V,0)$ be a Gaussian state with covariance matrix V and zero first moment. Then, the trace norm of the derivative of the Gaussian state $\rho(V,0)$ along the direction $(X,0)$ can be upper bounded as*

$$\|\partial_{X,0}\rho(V,0)\|_1\leq\frac{1+\sqrt{3}}{4}\,\mathrm{Tr}[|X|\Omega^\intercal V\Omega]\,. \tag{3.93}$$

Proof. We have that

$$\begin{aligned}\|\partial_{X,0}\rho(V,0)\|_1 &\overset{\text{(i)}}{=} \frac{1}{4}\left\|\sum_{ij}\tilde{X}_{ij}[\hat{R}_i,[\hat{R}_j,\rho]]\right\|_1 \\ &\overset{\text{(ii)}}{=} \frac{1}{4}\left\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\rho+\rho\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}-2\sum_{i,j}\tilde{X}_{ij}\hat{R}_i\rho\hat{R}_j\right\|_1 \\ &\overset{\text{(iii)}}{\leq} \frac{1}{2}\left\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\rho\right\|_1+\frac{1}{2}\left\|\sum_{i,j}\tilde{X}_{ij}\hat{R}_i\rho\hat{R}_j\right\|_1,\end{aligned} \tag{3.94}$$

where in (i) we employed Lemma 115, (ii) is a consequence of the fact X is symmetric, and in (iii) we used triangle inequality and the fact that

$$\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\rho\|_1=\|(\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\rho)^\dagger\|_1=\|\rho\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\|_1\,. \tag{3.95}$$

Let us upper bound each of the two terms in (3.94). The first term can be upper bounded as follows:

$$\begin{aligned}\left\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\rho\right\|_1 &= \left\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\sqrt{\rho}\sqrt{\rho}\right\|_1 \\ &\overset{\text{(iv)}}{\leq} \left\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\sqrt{\rho}\right\|_2\|\sqrt{\rho}\|_2 \\ &= \sqrt{\operatorname{Tr}\left[(\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}})^2\rho\right]} \\ &\overset{\text{(v)}}{=} \frac{\sqrt{(\operatorname{Tr}[\tilde{X}V])^2+2\operatorname{Tr}[\tilde{X}V\tilde{X}V]+2\operatorname{Tr}[\Omega\tilde{X}\Omega\tilde{X}]}}{2} \\ &\overset{\text{(vi)}}{\leq} \frac{\sqrt{\left(\operatorname{Tr}[|X|\Omega^\intercal V\Omega]\right)^2+2\operatorname{Tr}[\tilde{X}V\tilde{X}V]+2\operatorname{Tr}[\Omega\tilde{X}\Omega\tilde{X}]}}{2} \\ &\overset{\text{(vii)}}{\leq} \frac{\sqrt{3}}{2}\operatorname{Tr}[|X|\Omega^\intercal V\Omega]\,.\end{aligned} \tag{3.96}$$

Here, in (iv), we exploited Hölder's inequality. In (v), we exploited Lemma 107. In (vi), we used that

$$(\operatorname{Tr}[\tilde{X}V])^2\leq(\operatorname{Tr}[|\tilde{X}|V])^2=\left(\operatorname{Tr}[|X|\Omega^\intercal V\Omega]\right)^2. \tag{3.97}$$

In (vii), we first considered the real spectral decomposition of the real symmetric matrix $\tilde{X}$, i.e.

$$\tilde{X}=\sum_{j=1}^{2n}\lambda_j\mathbf{e}_j\mathbf{e}_j^\intercal \tag{3.98}$$

where $\lambda_1,\lambda_2,\ldots,\lambda_{2n}\in\mathbb{R}$ are the eigenvalues and $\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_{2n}\in\mathbb{R}^{2n}$ are the real orthonormal

eigenvectors, and second we observed that

$$\begin{aligned}
\mathrm{Tr}[\tilde{X}V\tilde{X}V] + \mathrm{Tr}[\Omega\tilde{X}\Omega\tilde{X}] &= \mathrm{Tr}[\tilde{X}(V+i\Omega)\tilde{X}(V-i\Omega)] \\
&\overset{\text{(viii)}}{=} \mathrm{Tr}\Big[\sqrt{V-i\Omega}\tilde{X}\sqrt{V+i\Omega}\sqrt{V+i\Omega}\tilde{X}\sqrt{V-i\Omega}\Big] \\
&= \Big\|\sqrt{V-i\Omega}\tilde{X}\sqrt{V+i\Omega}\Big\|_2^2 \\
&\leq \left(\sum_{j=1}^{2n} |\lambda_j| \Big\|\sqrt{V-i\Omega}\mathbf{e}_j\mathbf{e}_j^\intercal\sqrt{V+i\Omega}\Big\|_2\right)^2 \\
&\overset{\text{(ix)}}{=} \left(\sum_{j=1}^{2n} |\lambda_j|\sqrt{\mathbf{e}_j^\intercal(V-i\Omega)\mathbf{e}_j\mathbf{e}_j^\intercal(V+i\Omega)\mathbf{e}_j}\right)^2 \\
&\overset{\text{(x)}}{=} \left(\sum_{j=1}^{2n} |\lambda_j|\mathbf{e}_j^\intercal V\mathbf{e}_j\right)^2 \\
&= \Big(\mathrm{Tr}[|\tilde{X}|V]\Big)^2 \\
&= \left(\mathrm{Tr}[|X|\Omega^\intercal V\Omega]\right)^2 .
\end{aligned} \tag{3.99}$$

Here, in (viii) we exploited the uncertainty relation $V + i\Omega \geq 0$, which implies that $V - i\Omega = (V+i\Omega)^\intercal \geq 0$. In (ix), we exploited that $\| \, |v\rangle\langle w| \, \|_2 = \sqrt{\langle v|v\rangle \, \langle w|w\rangle}$. Finally, in (x), we exploited that $\mathbf{e}_j$ is a real vector and Ω is skew-symmetric, implying that $\mathbf{e}_j^\intercal\Omega\mathbf{e}_j = 0$. This completes the proof of $\left\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\rho\right\|_1 \leq \frac{\sqrt{3}}{2}\,\mathrm{Tr}[|X|\Omega^\intercal V\Omega]$.

Now, let us proceed with the bound of the second term in (3.94). Similarly as above, let us write the real spectral decomposition of $\tilde{X}$ as $\tilde{X} = O^\intercal DO$, with $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_{2n})$ and O orthogonal. Then, by defining $\hat{Y}_l \coloneqq \sum_{i=1}^{2n} O_{li}\hat{R}_i$, we have that

$$\begin{aligned}
\left\|\sum_{i,j=1}^{2n} \tilde{X}_{ij}\hat{R}_i\rho\hat{R}_j\right\|_1 &= \left\|\sum_{l=1}^{2n} \lambda_l\hat{Y}_l\rho\hat{Y}_l\right\|_1 \\
&\leq \sum_{l=1}^{2n} |\lambda_l| \left\|\hat{Y}_l\rho\hat{Y}_l\right\|_1 \\
&\overset{\text{(xi)}}{\leq} \sum_{l=1}^{2n} |\lambda_l| \|\hat{Y}_l\sqrt{\rho}\|_2 \|\sqrt{\rho}\hat{Y}_l\|_2 \\
&\leq \sum_{l=1}^{2n} |\lambda_l| \,\mathrm{Tr}[\rho\hat{Y}_l^2] \\
&= \sum_{l=1}^{2n} |\lambda_l| \sum_{i,j=1}^{2n} O_{li}O_{lj} \,\mathrm{Tr}[\rho\hat{R}_i\hat{R}_j] \\
&= \sum_{i,j=1}^{2n} |\tilde{X}|_{ij} \,\mathrm{Tr}[\rho\hat{R}_i\hat{R}_j] \\
&= \frac{1}{2}\sum_{i,j=1}^{2n} |\tilde{X}|_{ij} \,\mathrm{Tr}[\rho\{\hat{R}_i, \hat{R}_j\}] \\
&= \frac{1}{2}\sum_{i,j=1}^{2n} |\tilde{X}|_{ij} V_{ij} \\
&= \frac{1}{2}\,\mathrm{Tr}[V|\tilde{X}|] \\
&= \frac{1}{2}\,\mathrm{Tr}[|X|\Omega^\intercal V\Omega] \,,
\end{aligned} \tag{3.100}$$

where in (xi) we employed Hölder's inequality. Consequently, by using the inequalities (3.94), (3.96), and (3.100), we conclude that

$$\|\partial_{X,0}\rho(V,0)\|_1 \leq \frac{\sqrt{3}+1}{4}\operatorname{Tr}[|X|\Omega^\intercal V\Omega]\,. \tag{3.101}$$

□

Exact formula for the trace norm of the derivative of a pure Gaussian state In this subsection, we determine a closed formula for the trace norm of the derivative of a pure Gaussian state $\rho(V,0)$ in terms of its covariance matrix V.

Theorem 119 (Closed formula for the trace norm of the derivative of a pure Gaussian state)**.** *Let $\rho(V,0)$ be a pure Gaussian state with a (pure) covariance matrix V and zero first moment. Then, the trace norm of the derivative of the Gaussian state $\rho(V,0)$ along the direction $(X,0)$ is given by*

$$\|\partial_{X,0}\rho(V,0)\|_1 = \frac{\sqrt{(\operatorname{Tr}[\tilde{X}V])^2 + 2\operatorname{Tr}[\tilde{X}V\tilde{X}V] + 2\operatorname{Tr}[\Omega\tilde{X}\Omega\tilde{X}]} + \|\sqrt{V+i\Omega}\tilde{X}\sqrt{V+i\Omega}\|_1}{4}, \tag{3.102}$$

where $\tilde{X} := \Omega X\Omega^\intercal$.

Proof. Since $\rho(V,0)$ is pure, we can write $\rho(V,0) = |\psi\rangle\langle\psi|$ for some $|\psi\rangle$. Lemma 115 establishes that

$$\begin{aligned}
\partial_{X,0}\rho(V,0) &= -\frac{1}{4}\sum_{ij}\tilde{X}_{ij}[\hat{R}_i,[\hat{R}_j,|\psi\rangle\langle\psi|]] \\
&\overset{\text{(i)}}{=} -\frac{1}{4}\left(\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\,|\psi\rangle\langle\psi| + |\psi\rangle\langle\psi|\,\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}} - 2\sum_{i,j}\tilde{X}_{ij}\hat{R}_i\,|\psi\rangle\langle\psi|\,\hat{R}_j\right) \\
&= -\frac{1}{4}\left(\Theta_1 - \Theta_2\right)
\end{aligned} \tag{3.103}$$

where in (i) we exploited that $\tilde{X}$ is symmetric, and in (ii) we defined the hermitian operators

$$\begin{aligned}
\Theta_1 &:= \hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\,|\psi\rangle\langle\psi| + |\psi\rangle\langle\psi|\,\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\,, \\
\Theta_2 &:= 2\sum_{i,j}\tilde{X}_{ij}\hat{R}_i\,|\psi\rangle\langle\psi|\,\hat{R}_j\,.
\end{aligned} \tag{3.104}$$

Note that these operators are orthogonal to each other, i.e. $\Theta_1\Theta_2 = \Theta_2\Theta_1 = 0$. This follows from the fact that the first moment of $\rho(V,0)$ vanishes and from Lemma 112, which implies that $\operatorname{Tr}[\psi\hat{R}_i\hat{R}_j\hat{R}_k] = 0$ for all $i,j,k \in \{1,2,\ldots,2n\}$. Consequently, we have that

$$\|\partial_{X,0}\rho(V,0)\|_1 = \frac{\|\Theta_1\|_1}{4} + \frac{\|\Theta_2\|_1}{2}\,, \tag{3.105}$$

where

$$\begin{aligned}
\|\Theta_1\|_1 &= \left\|\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\,|\psi\rangle\langle\psi| + |\psi\rangle\langle\psi|\,\hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\right\|_1 \\
\|\Theta_2\|_1 &= \left\|\sum_{i,j}\tilde{X}_{ij}\hat{R}_i\,|\psi\rangle\langle\psi|\,\hat{R}_j\right\|_1 .
\end{aligned} \tag{3.106}$$

Let us proceed to calculate the term $\|\Theta_1\|_1$ in (3.105). By defining the non-normalised vector $|v\rangle := \hat{\mathbf{R}}^\intercal\tilde{X}\hat{\mathbf{R}}\,|\psi\rangle$, we have that

$$\|\Theta_1\|_1 = \|\,|v\rangle\langle\psi| + |\psi\rangle\langle v|\,\|_1\,. \tag{3.107}$$

The dimension of the support of the operator $|v\rangle\langle\psi| + |\psi\rangle\langle v|$ is two. Hence, its trace norm can be easily calculated and it reads:

$$\|\Theta_1\|_1 = \| \, |v\rangle\langle\psi| + |\psi\rangle\langle v| \, \|_1 = 2\sqrt{\langle v|v\rangle - \mathrm{Im}(\langle\psi|v\rangle)^2} \, . \tag{3.108}$$

Note that $\langle v|v\rangle = \mathrm{Tr}[(\hat{\mathbf{R}}^\intercal \tilde{X} \hat{\mathbf{R}})^2 \psi]$ and that $\mathrm{Im}(\langle\psi|v\rangle) = \mathrm{Im}(\mathrm{Tr}[\hat{\mathbf{R}}^\intercal \tilde{X} \hat{\mathbf{R}} \psi]) = 0$. To show the latter, note that $\hat{\mathbf{R}}^\intercal \tilde{X} \hat{\mathbf{R}}$ is an hermitian operator and thus the expectation value $\mathrm{Tr}[\hat{\mathbf{R}}^\intercal \tilde{X} \hat{\mathbf{R}} \psi]$ is real. Consequently, we have that

$$\begin{aligned} \|\Theta_1\|_1 &= \sqrt{4 \, \mathrm{Tr}[(\hat{\mathbf{R}}^\intercal \tilde{X} \hat{\mathbf{R}})^2 \psi]} \\ &= \sqrt{(\mathrm{Tr}[\tilde{X} V])^2 + 2 \, \mathrm{Tr}[\tilde{X} V \tilde{X} V] + 2 \, \mathrm{Tr}[\Omega \tilde{X} \Omega \tilde{X}]} \, , \end{aligned} \tag{3.109}$$

where in the last line we employed Lemma 107.

Let us now calculate the term $\|\Theta_2\|_1$ in (3.105). To this end, we exploit the simple-to-prove fact that, given (non-orthonormal) vectors $(|\psi_i\rangle)_i$ and a matrix A, the non-zero singular values of the operator $\sum_{ij} A_{ij} \, |\psi_i\rangle\langle\psi_j|$ are the same of the matrix $\sqrt{G} A \sqrt{G}$, where G is the *Gram matrix* defined as $G_{ij} := \langle\psi_i|\psi_j\rangle$. Consequently, we have that

$$\|\Theta_2\|_1 = \left\| \sum_{i,j} \tilde{X}_{ij} \hat{R}_i \, |\psi\rangle\langle\psi| \, \hat{R}_j \right\|_1 = \left\| \sqrt{G} \tilde{X} \sqrt{G} \right\|_1 \, , \tag{3.110}$$

where

$$G_{ij} := \langle\psi| \, \hat{R}_i \hat{R}_j \, |\psi\rangle = \frac{V_{ij} + i\Omega_{ij}}{2}, \, . \tag{3.111}$$

Hence, we have that $\|\Theta_2\|_1 = \frac{1}{2} \|\sqrt{V + i\Omega} \tilde{X} \sqrt{V + i\Omega}\|_1$. Consequently, by substituting into (3.105), we conclude that

$$\|\partial_{X,0} \rho(V, 0)\|_1 = \frac{\sqrt{(\mathrm{Tr}[\tilde{X} V])^2 + 2 \, \mathrm{Tr}[\tilde{X} V \tilde{X} V] + 2 \, \mathrm{Tr}[\Omega \tilde{X} \Omega \tilde{X}]} + \|\sqrt{V + i\Omega} \tilde{X} \sqrt{V + i\Omega}\|_1}{4} \, . \tag{3.112}$$

□

Upper bound on the trace distance between Gaussian states

In this subsection, we derive our main result: an upper bound on the trace distance between Gaussian states in terms of the norm distance of their first moments and covariance matrices.

Theorem 120 (Upper bound on the trace distance between Gaussian states)**.** *Let $\rho(V, \mathbf{m})$ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Similarly, let $\rho(W, \mathbf{t})$ be the Gaussian state with first moment $\mathbf{t}$ and covariance matrix W. Then, the trace distance can be upper bounded as follows:*

$$\frac{1}{2} \|\rho(V, \mathbf{m}) - \rho(W, \mathbf{t})\|_1 \leq \frac{1 + \sqrt{3}}{8} \, \mathrm{Tr}\left[|V - W| \Omega^\intercal \left(\frac{V + W}{2} \right) \Omega \right] + \sqrt{\frac{\min(\|V\|_\infty, \|W\|_\infty)}{2}} \|\mathbf{m} - \mathbf{t}\|_2 \, . \tag{3.113}$$

In particular, Hölder's inequality implies that

$$\begin{aligned} \frac{1}{2} \|\rho(V, \mathbf{m}) - \rho(W, \mathbf{t})\|_1 &\leq \frac{1 + \sqrt{3}}{8} \max(\|V\|_\infty, \|W\|_\infty) \, \|V - W\|_1 + \sqrt{\frac{\min(\|V\|_\infty, \|W\|_\infty)}{2}} \|\mathbf{m} - \mathbf{t}\|_2 \, , \\ \frac{1}{2} \|\rho(V, \mathbf{m}) - \rho(W, \mathbf{t})\|_1 &\leq \frac{1 + \sqrt{3}}{8} \max(\mathrm{Tr} \, V, \mathrm{Tr} \, W) \, \|V - W\|_\infty + \sqrt{\frac{\min(\|V\|_\infty, \|W\|_\infty)}{2}} \|\mathbf{m} - \mathbf{t}\|_2 \, . \end{aligned} \tag{3.114}$$

Proof. We have that

$$\begin{aligned}\|\rho(V,\mathbf{m})-\rho(W,\mathbf{t})\|_1 &\leq \|\rho(V,\mathbf{m})-\rho(W,\mathbf{m})\|_1+\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1\\ &= \|\rho(V,0)-\rho(W,0)\|_1+\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1\end{aligned} \tag{3.115}$$

where in the last line we exploited the invariance of the trace norm under (displacement) unitaries. The first term in (3.115) can be upper bounded as follows:

$$\begin{aligned}\|\rho(W,0)-\rho(V,0)\|_1 &\overset{\text{(i)}}{=} \left\|\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\partial_{W-V,0}\rho\Big(V+\frac{i-1}{m}(W-V),0\Big)\right\|_1\\ &\overset{\text{(ii)}}{\leq} \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\left\|\partial_{W-V,0}\rho\Big(V+\frac{i-1}{m}(W-V),0\Big)\right\|_1\\ &\overset{\text{(iii)}}{\leq} \frac{\sqrt{3}+1}{4}\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\mathrm{Tr}\left[|V-W|\Omega^{\intercal}\left(V+\frac{i-1}{m}(W-V)\right)\Omega\right]\\ &= \frac{\sqrt{3}+1}{4}\mathrm{Tr}\left[|V-W|\Omega^{\intercal}\left(\frac{V+W}{2}\right)\Omega\right].\end{aligned} \tag{3.116}$$

Here, equality (i) is a consequence of Corollary 136 (together with a 'discretization' of the integral), in (ii), we exploited triangle inequality; and in (iii), we employed Lemma 118.

Now, let us upper bound the second term in (3.115). Let us consider the Williamson decomposition $W = SDS^{\intercal}$, as in (3.16). Then, Lemma 105 guarantees that

$$\begin{aligned}\rho(W,\mathbf{m}) &= \mathcal{N}_{S(D-\mathbb{1})S^{\intercal}}\left(\rho(SS^{\intercal},\mathbf{m})\right)\\ \rho(W,\mathbf{t}) &= \mathcal{N}_{S(D-\mathbb{1})S^{\intercal}}\left(\rho(SS^{\intercal},\mathbf{t})\right),\end{aligned} \tag{3.117}$$

where $\mathcal{N}_{S(D-\mathbb{1})S^{\intercal}}$ is the Gaussian noise channel defined in Subsection 3.8.1. Consequently, we have that

$$\begin{aligned}\frac{1}{2}\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1 &= \frac{1}{2}\|\mathcal{N}_{S(D-1)S^{\intercal}}\left(\rho(SS^{\intercal},\mathbf{m})-\rho(SS^{\intercal},\mathbf{t})\right)\|_1\\ &\overset{\text{(iv)}}{\leq} \frac{1}{2}\|\rho(SS^{\intercal},\mathbf{m})-\rho(SS^{\intercal},\mathbf{t})\|_1\\ &\overset{\text{(v)}}{=} \sqrt{1-\mathrm{Tr}[\rho(SS^{\intercal},\mathbf{m})\,\rho(SS^{\intercal},\mathbf{t})]}\\ &\overset{\text{(vi)}}{=} \sqrt{1-e^{-\frac{1}{2}(\mathbf{m}-\mathbf{t})^{\intercal}(SS^{\intercal})^{-1}(\mathbf{m}-\mathbf{t})}}\\ &\overset{\text{(vii)}}{\leq} \sqrt{\frac{(\mathbf{m}-\mathbf{t})^{\intercal}(SS^{\intercal})^{-1}(\mathbf{m}-\mathbf{t})}{2}}\\ &\leq \sqrt{\frac{\|(SS^{\intercal})^{-1}\|_{\infty}}{2}}\|\mathbf{m}-\mathbf{t}\|_2\\ &\overset{\text{(viii)}}{=} \sqrt{\frac{\|SS^{\intercal}\|_{\infty}}{2}}\|\mathbf{m}-\mathbf{t}\|_2\\ &\overset{\text{(ix)}}{\leq} \sqrt{\frac{\|W\|_{\infty}}{2}}\|\mathbf{m}-\mathbf{t}\|_2.\end{aligned} \tag{3.118}$$

Here, in (iv) we applied the monotonicity of the trace norm under quantum channels; in (v) we exploited that $\rho_{SS^{\intercal},\mathbf{m}}$ and $\rho_{SS^{\intercal},\mathbf{t}}$ are pure (Gaussian) states; in (vi), we employed the known formula for the overlap between Gaussian states [2, Eq (4.51)]; in (vii), we exploited that $e^x \geq 1+x$ for all $x \in \mathbb{R}$; in (viii), we exploited that for any symplectic matrix T it holds that

$$\|T^{-1}\|_{\infty} = \|\Omega T^{\intercal}\Omega^{\intercal}\|_{\infty} = \|T^{\intercal}\|_{\infty} = \|T\|_{\infty}, \tag{3.119}$$

where we exploited that $T\Omega T^{\intercal} = \Omega$ and the fact that Ω is orthogonal; finally, in (ix), we used the fact that $SS^{\intercal} \leq SDS^{\intercal} = W$.

By substituting (3.116) and (3.118) into (3.115), we obtain the desired bound. □

Improved upper bound on the trace distance between Gaussian states In this section, we derive an improved upper bound on the trace distance between Gaussian states, which involves the Williamson decomposition of the covariance matrices.

Theorem 121 (Improved upper bound)**.** *Let $\rho(V, \mathbf{m})$ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Similarly, let $\rho(W, \mathbf{t})$ be the Gaussian state with first moment $\mathbf{t}$ and covariance matrix W. Moreover, let $V = S_1 D_1 S_1^\intercal$ and $W = S_2 D_2 S_2^\intercal$ be the Williamson decompositions of V and W, as in (3.16). Then, the following better bound holds:*

$$\frac{1}{2}\left\|\rho(V,\mathbf{m})-\rho(W,\mathbf{t})\right\|_1 \leq \frac{1+\sqrt{3}}{8}\operatorname{Tr}\left[|V-W|\Omega^\intercal\left(\frac{S_1S_1^\intercal+S_2S_2^\intercal}{2}\right)\Omega\right]+\frac{\min(\|S_1\|_\infty,\|S_2\|_\infty)}{\sqrt{2}}\|\mathbf{m}-\mathbf{t}\|_2\,. \tag{3.120}$$

In particular, Hölder's inequality implies that

$$\begin{aligned}\frac{1}{2}\left\|\rho(V,\mathbf{m})-\rho(W,\mathbf{t})\right\|_1 &\leq \frac{1+\sqrt{3}}{8}\max(\|S_1\|_\infty^2,\|S_2\|_\infty^2)\|V-W\|_1+\frac{\min(\|S_1\|_\infty,\|S_2\|_\infty)}{\sqrt{2}}\|\mathbf{m}-\mathbf{t}\|_2\,,\\ \frac{1}{2}\left\|\rho(V,\mathbf{m})-\rho(W,\mathbf{t})\right\|_1 &\leq \frac{1+\sqrt{3}}{8}\max(\|S_1\|_2^2,\|S_2\|_2^2)\|V-W\|_\infty+\frac{\min(\|S_1\|_\infty,\|S_2\|_\infty)}{\sqrt{2}}\|\mathbf{m}-\mathbf{t}\|_2\,.\end{aligned} \tag{3.121}$$

Proof. Similarly to the proof of Theorem 120, we have that

$$\begin{aligned}\frac{1}{2}\left\|\rho(V,\mathbf{m})-\rho(W,\mathbf{t})\right\|_1 &\overset{\text{(i)}}{\leq} \frac{1}{2}\|\rho(V,0)-\rho(W,0)\|_1+\frac{1}{2}\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1\\ &\overset{\text{(ii)}}{\leq} \frac{1}{2}\|\rho(V,0)-\rho(W,0)\|_1+\sqrt{\frac{\|S_2S_2^\intercal\|_\infty}{2}}\|\mathbf{m}-\mathbf{t}\|_2\\ &= \frac{1}{2}\|\rho(V,0)-\rho(W,0)\|_1+\frac{\|S_2\|_\infty}{\sqrt{2}}\|\mathbf{m}-\mathbf{t}\|_2\,.\end{aligned} \tag{3.122}$$

Here, in (i), we exploited (3.115), and in (ii), we employed step (viii) of (3.118). Hence, we just need to prove that

$$\|\rho(V,0)-\rho(W,0)\|_1 \leq \frac{1+\sqrt{3}}{4}\operatorname{Tr}\left[|V-W|\Omega^\intercal\left(\frac{S_1S_1^\intercal+S_2S_2^\intercal}{2}\right)\Omega\right]\,. \tag{3.123}$$

To this end, note that the triangle inequality implies that

$$\|\rho(W,0)-\rho(V,0)\|_1 \leq \lim_{m\to\infty}\sum_{i=1}^{m}\left\|\rho\Big(V+\frac{i}{m}(W-V),0\Big)-\rho\Big(V+\frac{i-1}{m}(W-V),0\Big)\right\|_1\,. \tag{3.124}$$

Moreover, since $V - S_1S_1^\intercal \geq 0$ and $W - S_2S_2^\intercal \geq 0$, it holds that

$$K_i := \left(1-\frac{i-1}{m}\right)(V-S_1S_1^\intercal)+\frac{i-1}{m}(W-S_2S_2^\intercal)\geq 0\,. \tag{3.125}$$

Additionally, note that

$$\begin{aligned}\rho\Big(V+\frac{i-1}{m}(W-V),0\Big) &= \mathcal{N}_{K_i}\left(S_1S_1^\intercal+\frac{i-1}{m}(S_2S_2^\intercal-S_1S_1^\intercal)\right)\,,\\ \rho\Big(V+\frac{i}{m}(W-V),0\Big) &= \mathcal{N}_{K_i}\left(S_1S_1^\intercal+\frac{i-1}{m}(S_2S_2^\intercal-S_1S_1^\intercal)+\frac{W-V}{m}\right)\,,\end{aligned} \tag{3.126}$$

where we denoted as $\mathcal{N}_{K_i}$ the Gaussian noise channel defined in Subsection 3.8.1 and we employed Lemma 105. Consequently, by employing (3.124) and the monotonicity of the trace norm under

quantum channels, we have that

$$\begin{aligned}\|\rho(W,0)-\rho(V,0)\|_1 \leq \lim_{m\to\infty}\sum_{i=1}^{m}\Bigg\|&\rho\Big(S_1S_1^\intercal+\frac{i-1}{m}(S_2S_2^\intercal-S_1S_1^\intercal)+\frac{W-V}{m},0\Big)\\&-\rho\Big(S_1S_1^\intercal+\frac{i-1}{m}(S_2S_2^\intercal-S_1S_1^\intercal),0\Big)\Bigg\|_1 .\end{aligned} \tag{3.127}$$

Hence, we obtain that

$$\begin{aligned}\|\rho(W,0)-\rho(V,0)\|_1 &\overset{\text{(iii)}}{\leq} \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\Big\|\partial_{W-V,0}\rho\Big(S_1S_1^\intercal+\frac{i-1}{m}(S_2S_2^\intercal-S_1S_1^\intercal),0\Big)\Big\|_1\\&\overset{\text{(iv)}}{\leq}\frac{\sqrt{3}+1}{4}\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\operatorname{Tr}\Big[|W-V|\Omega^\intercal\Big(S_1S_1^\intercal+\frac{i-1}{m}(S_2S_2^\intercal-S_1S_1^\intercal)\Big)\Big]\\&=\frac{\sqrt{3}+1}{4}\operatorname{Tr}\Big[|W-V|\Omega^\intercal\left(\frac{S_1S_1^\intercal+S_2S_2^\intercal}{2}\right)\Omega\Big],\end{aligned} \tag{3.128}$$

where in (iii) we employed Theorem 135, while in (iv) we exploited Lemma 118. This, together with (3.122), concludes the proof. □

Tightness of the upper bound In this subsection, we prove that the bounds on the trace distance derived in Theorem 120 are *tight*. The key tool here is to employ the closed formula for the trace norm of the derivative of a pure Gaussian state proved in Theorem 119.

Lemma 122 (Tightness of the upper bounds)**.** *Let $\varepsilon\in(0,1)$ and $a>1$. Let $\rho(V_a,0)$ and $\rho(W_{a,\varepsilon},0)$ be Gaussian states with zero first moments and covariance matrices given by*

$$\begin{aligned}V_a &\coloneqq \begin{pmatrix}a & 0\\ 0 & a^{-1}\end{pmatrix},\\ W_{a,\varepsilon} &\coloneqq \begin{pmatrix}a & 0\\ 0 & a^{-1}+\epsilon\end{pmatrix}.\end{aligned} \tag{3.129}$$

Remarkably, the trace distance satisfies

$$\lim_{\varepsilon\to0^+}\frac{\frac{1}{2}\|\rho(W_{a,\varepsilon},0)-\rho(V_a,0)\|_1}{\frac{1+\sqrt{3}}{8}\max(\|W_{a,\varepsilon}\|_\infty,\|V_a\|_\infty)\,\|W_{a,\varepsilon}-V_a\|_1}=1\,, \tag{3.130}$$

and also

$$\lim_{a\to+\infty}\lim_{\varepsilon\to0^+}\frac{\frac{1}{2}\|\rho(W_{a,\varepsilon},0)-\rho(V_a,0)\|_1}{\frac{1+\sqrt{3}}{8}\max(\operatorname{Tr}W_{a,\varepsilon},\operatorname{Tr}V_a)\,\|W_{a,\varepsilon}-V_a\|_\infty}=1\,, \tag{3.131}$$

establishing the tightness of the upper bounds provided in Theorem 120.

Proof. Note that

$$W_{a,\varepsilon}=V_a+\varepsilon X\,, \tag{3.132}$$

where

$$X\coloneqq\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}. \tag{3.133}$$

Consequently, it holds that

$$\begin{aligned}\lim_{\varepsilon\to0^+}\frac{\|\rho(W_{a,\varepsilon},0)-\rho(V_a,0)\|_1}{\varepsilon}&=\lim_{\varepsilon\to0^+}\frac{\|\rho(V_a+\varepsilon X,0)-\rho(V_a,0)\|_1}{\varepsilon}\\&=\|\partial_{X,0}\rho(V_a,0)\|_1\\&\overset{\text{(i)}}{=}\frac{\sqrt{(\operatorname{Tr}[\tilde{X}V_a])^2+2\operatorname{Tr}[\tilde{X}V_a\tilde{X}V_a]+2\operatorname{Tr}[\Omega\tilde{X}\Omega\tilde{X}]}+\|\sqrt{V_a+i\Omega}\tilde{X}\sqrt{V_a+i\Omega}\|_1}{4},\end{aligned} \tag{3.134}$$

where in (i) we used the closed formula for the derivative of a pure Gaussian state reported in Lemma 119 and we denoted

$$\tilde{X} := \Omega^{\intercal} X \Omega = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} . \tag{3.135}$$

Moreover, we have that

$$\begin{aligned} \mathrm{Tr}[\tilde{X} V_a] &= a \,, \\ \mathrm{Tr}[\tilde{X} V_a \tilde{X} V_a] &= a^2 \,, \\ \mathrm{Tr}[\Omega \tilde{X} \Omega \tilde{X}] &= 0 \,, \\ \| \sqrt{V_a + i\Omega} \tilde{X} \sqrt{V_a + i\Omega} \|_1 &= a \,. \end{aligned} \tag{3.136}$$

As a consequence, we obtain that

$$\lim_{\varepsilon \to 0^+} \frac{\| \rho(W_{a,\varepsilon}, 0) - \rho(V_a, 0) \|_1}{\varepsilon} = \frac{1 + \sqrt{3}}{4} a \tag{3.137}$$

Since

$$\begin{aligned} \| W_{a,\varepsilon} - V_a \|_\infty &= \varepsilon \,, \\ \max(\| W_{a,\varepsilon} \|_\infty, \| V_a \|_\infty) &= \max(a, a^{-1} + \varepsilon) \,, \end{aligned} \tag{3.138}$$

we thus obtain that

$$\lim_{\varepsilon \to 0^+} \frac{\frac{1}{2} \| \rho(W_{a,\varepsilon}, 0) - \rho(V_a, 0) \|_1}{\frac{1+\sqrt{3}}{8} \max(\| W_{a,\varepsilon} \|_\infty, \| V_a \|_\infty) \, \| W_{a,\varepsilon} - V_a \|_1} = 1 \,, \tag{3.139}$$

Additionally, since

$$\begin{aligned} \| W_{a,\varepsilon} - V_a \|_1 &= \varepsilon \,, \\ \max(\mathrm{Tr}\, W_{a,\varepsilon}, \mathrm{Tr}\, V_a) &= a + a^{-1} + \varepsilon \,, \end{aligned} \tag{3.140}$$

we obtain that

$$\lim_{\varepsilon \to 0^+} \frac{\frac{1}{2} \| \rho(W_{a,\varepsilon}, 0) - \rho(V_a, 0) \|_1}{\frac{1+\sqrt{3}}{8} \max(\mathrm{Tr}\, W_{a,\varepsilon}, \mathrm{Tr}\, V_a) \, \| W_{a,\varepsilon} - V_a \|_\infty} = \frac{1}{1 + a^{-2}} \,, \tag{3.141}$$

and hence

$$\lim_{a \to +\infty} \lim_{\varepsilon \to 0^+} \frac{\frac{1}{2} \| \rho(W_{a,\varepsilon}, 0) - \rho(V_a, 0) \|_1}{\frac{1+\sqrt{3}}{8} \max(\mathrm{Tr}\, W_{a,\varepsilon}, \mathrm{Tr}\, V_a) \, \| W_{a,\varepsilon} - V_a \|_\infty} = 1 \,. \tag{3.142}$$

□

3.8.3 Differentiability of Gaussian states

In this section we study the regularity properties of Gaussian states, when they are seen as functions of their statistical moments. In particular, we prove that they are continuously differentiable. The section is organised as follows:

- In Subsection 3.8.3 we introduce all the preliminaries required to prove the main result of this section.
- In Subsection 3.8.3 we give the full proof of the regularity of Gaussian states.

Preliminaries on functional analysis

In this subsection we review some basic facts about functional analysis. In particular, we have to define Fréchet spaces and some properties characterizing them. For the sake of brevity, in the following section the Hilbert space $L^2(\mathbb{R}^n)$ will be called $\mathscr{H}$. As it is well-known, we can deal with vector spaces in infinite dimensions through Banach spaces. In Banach spaces we can define a norm, so that the distance between vectors can be written as a difference in that norm. Normed vector spaces have useful properties, unfortunately there are instances of vector spaces which are not *normable*, i. e. vector spaces whose topology cannot be derived from a norm. One example of those spaces are Fréchet spaces. Fortunately, Fréchet spaces are manageable, because their continuity properties can be derived by a countable number of seminorms. Formally we say that:

Definition 123 (Fréchet space [84])**.** *A Fréchet space is a topological vector space which is complete with respect to a countable family of seminorms* $\{\|\cdot\|_k\}_{k\in\mathbb{N}}$*. Moreover, a complete translationally-invariant metric for the space can be defined by the quantity*

$$d(x,y) \coloneqq \sum_{k\in\mathbb{N}} 2^{-k} \frac{\|x-y\|_k}{1+\|x-y\|_k}\,. \tag{3.143}$$

As we have seen above Fréchet spaces are complete metric spaces, a distance is translationally-invariant if, for any x, y, z vectors, it satisfies

$$d(x+z, y+z) = d(x,y)\,. \tag{3.144}$$

Now, we have to clarify what we mean when we talk about continuity in Fréchet spaces. In particular, we need to define the concept of limit in these sets. As we see in the following Definition, analogously to Banach spaces, where the limit is defined on the space norm, in Fréchet spaces we have to take into account the whole countable family of seminorms.

Definition 124 (Continuity in Fréchet spaces)**.** *We say that a function* $f : A \to B$*, where* A *and* B *are Fréchet spaces with families of seminorms* $\{\|\cdot\|_{A,k}\}_{k\in\mathbb{N}}$ *and* $\{\|\cdot\|_{B,k}\}_{k\in\mathbb{N}}$*, respectively, is continuous at point* $\bar{a} \in A$ *in the Fréchet topology, if* $\exists\, \bar{b} \in B$ *such that* $f(\bar{a}) = \bar{b}$ *and for any sequence* $\{a_n\}_{n\in\mathbb{N}}$ *satisfying the following condition for any* $k \in \mathbb{N}$

$$\lim_{n\to\infty} \|\bar{a} - a_n\|_{A,k} = 0\,, \tag{3.145}$$

it holds that

$$\lim_{n\to\infty} \|\bar{b} - f(a_n)\|_{B,k} = 0\,, \tag{3.146}$$

for any $k \in \mathbb{N}$*. Equivalently, we can define continuity in Fréchet topology by employing the distance introduced in Equation* (3.143) *of Definition 123, and applying the standard notion of continuity in metric spaces. If the above condition is met we write that*

$$\lim_{a\to\bar{a}} f(a) = \bar{b}\,. \tag{3.147}$$

Even though Fréchet space may seem obscure mathematical constructions, a lot of widely used vector spaces are of this kind. One example is the space of infinitely differentiable functions over $\mathbb{R}^n$, here the countable family of seminorms is given by the uniform norms of the derivatives of the function. Another important case is represented by the *regular and rapidly decreasing functions*, also known as Schwartz functions. Here, we define the Schwartz space of functions as:

Definition 125 (Schwartz space [85])**.** *For any* $k \in \mathbb{N}$*, we call Schwartz space* $\mathcal{S}\left(\mathbb{R}^k\right)$ *the set of* C^∞ *complex-valued functions* $f : \mathbb{R}^k \to \mathbb{C}$ *such that*

$$\sup_{\mathbf{r}\in\mathbb{R}^k} |\mathbf{r}^a \partial_b f(\mathbf{r})| < \infty\,, \tag{3.148}$$

for any multi-indices $a, b \in \mathbb{N}^k$*. Here,* $\mathbf{r}^a \coloneqq r_1^{a_1} r_2^{a_2} \dots r_k^{a_k}$ *and* $\partial_b f \coloneqq \frac{\partial^{b_1}}{\partial r_1^{b_1}} \frac{\partial^{b_2}}{\partial r_2^{b_2}} \dots \frac{\partial^{b_k}}{\partial r_k^{b_k}} f$*.*

Up to now, we have discussed sets of complex-valued functions, but in this work we deal with quantum states, so we need to define an analogue of the Schwartz space for infinite-dimensional linear operators. In [83] authors introduce precisely the set that is needed in the following, i. e. the set of Schwartz operators. Besides, they prove many different properties of these operators. Here, we review the definition of this vector space and of the properties required for our work.

Definition 126 (Schwartz operators [83])**.** *For any number of bosonic modes $n \in \mathbb{N}$, we define the set of Schwartz operators $\mathcal{S}(\mathscr{H})$ as the linear operators acting on a dense subspace of $\mathscr{H}$ such that all the following seminorms are bounded:*

$$\left\|\hat{T}\right\|_{a,b} := \sup\left\{\left|\langle \hat{\mathbf{R}}^a\psi|\hat{T}\hat{\mathbf{R}}^b\varphi\rangle\right| : \varphi,\psi \in \mathcal{S}\left(\mathbb{R}^{2n}\right), \|\psi\|_{L^2}, \|\varphi\|_{L^2} \leq 1\right\}\,, \tag{3.149}$$

for any multi-indices $a, b \in \mathbb{N}^{2n}$.

The set $\mathcal{S}(\mathscr{H})$ is a Fréchet space. Moreover, as we are going to see, $\mathcal{S}(\mathscr{H}) \subset \mathcal{T}(\mathscr{H})$, is the set of trace-class operators on $\mathscr{H}$. The seminorms defined above are slightly contrived, so we now want to find a better way to express them.

Proposition 127. *For any linear operator $\hat{T} \in \mathcal{S}(\mathscr{H})$, and for any a, b multi-indices, it holds that*

$$\|\hat{T}\|_{a,b} = \|\hat{\mathbf{R}}^a\hat{T}\hat{\mathbf{R}}^b\|_\infty\,. \tag{3.150}$$

Proof. The proof is a consequence of Lemma 3.2 in the arXiv version of [83]. □

Thanks to Proposition 127, we have a nice way to express the seminorms of the space $\mathcal{S}(\mathscr{H})$. These quantities are expressed in terms of operator norms (or Schatten infinity-norms), since in this work we are studying quantum states, it is interesting to have estimates in terms of the trace-norm. This is precisely the topis of the following Proposition:

Proposition 128 (Lemma 3.6 in the arXiv version of [83])**.** *For any complex separable Hilbert space, the set of the Schwartz operators $\mathcal{S}(\mathscr{H})$ is contained in the set of the trace-class operators $\mathcal{T}(\mathscr{H})$. Moreover, for any a, b multi-indices it is holds that*

$$\|\hat{\mathbf{R}}^a\hat{T}\hat{\mathbf{R}}^b\|_1 \leq \|\hat{E}_n^{-2}\|_1\|\hat{E}_n^2\hat{\mathbf{R}}^a\hat{T}\hat{\mathbf{R}}^b\|_\infty < \infty\,, \tag{3.151}$$

where $\hat{E}_n := \frac{1}{2}\sum_{j=1}^{2n}\hat{R}_j^2$ is the energy operator.

Proposition 129 (Proposition 3.18 in the arXiv version of [83])**.** *For any $n \in \mathbb{N}$ the so Fourier–Weyl transform is a continuous linear homeomorphism between the Schwartz space of functions $\mathcal{S}\left(\mathbb{R}^{2n}\right)$ and the set of Schwartz operators $\mathcal{S}\left(\mathscr{H}\right)$. The map is continuous in the natural Fréchet-space topology as in Definition 124.*

Proposition 129 gives us a way to connect Schwartz operators to Schwartz functions by competely preserving the structure of Fréchet spaces. As we are going to see in the next Section, this result allows us to deal with classical objects (Schwartz functions) in order to prove properties of quantum objects (Schwartz operators).

Proof of differentiability

In this subsection we prove that any quantum Gaussian state $\rho(V, \mathbf{m})$ is continuously differentiable as a function of its statistical moments. First of all we notice that the characteristic function of any Gaussian state, which takes the form $\chi_{\rho(V,\mathbf{m})}(\mathbf{r}) = \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal V\Omega\mathbf{r} + i\mathbf{m}^\intercal\Omega\mathbf{r}\right)$, belongs to the Schwartz space for any V and $\mathbf{m}$, so any Gaussian state is a Schwartz operator. The Fourier–Weyl transform of a Gaussian function is a Schwartz operator even if V is strictly positive but it is not an admissible quantum covariance matrix. In addition, it is also important to remark that any

polynomial multiplied by a Gaussian function is a Schwartz function. First, we explain a sketch of the proof. We want to prove that Gaussian states are continuously differentiable, and we know that Gaussian states can be expressed as the Fourier–Weyl transform of Gaussian functions, so we can first prove that Gaussian functions are continuously differentiable in the Fréchet topology of the Schwartz space. Then, we can apply Proposition 129 in order to lift up the regularity properties of Gaussian states to Gaussian functions.

We begin by stating this simple proposition on the Taylor expansion of Gaussian functions:

Proposition 130. *Given a Gaussian function of the form*

$$\exp\left(-\frac{1}{4}\,(\Omega\mathbf{r})^{\intercal}\,(V+X)\Omega\mathbf{r}+i(\mathbf{m}+\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\,, \tag{3.152}$$

where V is a $2n\times 2n$ strictly positive real symmetric matrix, X is a $2n\times 2n$ real and symmetric matrix, and $\mathbf{m},\,\mathbf{l}\in\mathbb{R}^{2n}$, then for any $\mathbf{r}\in\mathbb{R}^{2n}$ its Taylor polynomial reads

$$\exp\left(-\frac{1}{4}\,(\Omega\mathbf{r})^{\intercal}\,V\Omega\mathbf{r}+i\mathbf{m}^{\intercal}\Omega\mathbf{r}\right)\sum_{k,\ell\in\mathbb{N}}\frac{(-1)^k[(\Omega\mathbf{r})^{\intercal}X\Omega\mathbf{r}]^k i^{\ell}(\mathbf{l}^{\intercal}\Omega\mathbf{r})^{\ell}}{4^k k!\ell!}\,. \tag{3.153}$$

The following Lemma is important because it tells us that the integral remainder of the Taylor formula of a Gaussian at any order is a Schwartz function. This is going to be the first ingredient of our proof.

Lemma 131. *For any $n,k\in\mathbb{N}$, for any $\mathbf{m},\mathbf{l}\in\mathbb{R}^{2n}$, for any polynomial with complex coefficients $\mathscr{P}(\mathbf{r})$, for any, multi-indices a,b, for any $V>0$, and any X symmetric matrix such that $V+X>0$, the function*

$$f_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b,k}(\mathbf{r}):=\int_0^1(1-\alpha)^k\mathscr{P}(\mathbf{r})\exp\left(-\frac{1}{4}\,(\Omega\mathbf{r})^{\intercal}\,(V+\alpha X)\Omega\mathbf{r}+i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\mathrm{d}\alpha \tag{3.154}$$

belongs to $\mathcal{S}(\mathbb{R}^{2n})$.

Proof. In order to prove this Lemma we need to show that for any k, for any polynomial $\mathscr{P}$, for any multi-indices a and b, for any covariance matrix V, for any mean $\mathbf{m}$, and for any variations X and $\mathbf{l}$ the following holds true:

$$\sup_{\mathbf{r}\in\mathbb{R}^{2n}}\left|\mathbf{r}^a\partial_b\int_0^1(1-\alpha)^k\mathscr{P}(\mathbf{r})\exp\left(-\frac{1}{4}\,(\Omega\mathbf{r})^{\intercal}\,(V+\alpha X)\Omega\mathbf{r}+i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\mathrm{d}\alpha\right|<\infty\,. \tag{3.155}$$

So, we need to find a clever upper bound. The key point to prove this lemma is to find a function of α which dominates the following function:

$$(1-\alpha)^k\left|\mathbf{r}^a\partial_b\mathscr{P}(\mathbf{r})\exp\left(-\frac{1}{4}\,(\Omega\mathbf{r})^{\intercal}\,(V+\alpha X)\Omega\mathbf{r}+i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\right|. \tag{3.156}$$

To this end, we begin by considering the least eigenvalue of the positive matrix $V+\alpha X$, calling it $\lambda_{min}(\alpha)$. It is a continuous function of α, and by positivity of the aformentioned matrix we can conclude that

$$\lambda^*_{\min}:=\min_{\alpha\in[0,1]}\lambda_{\min}(\alpha)>0\,. \tag{3.157}$$

Moreover, we know that, for any $\alpha\in[0,1]$, $V+\alpha X\geq\lambda^*_{\min}\mathbb{1}$. The previous statement holds true for any V and X. So we obtain this simple upper bound on the exponential part of (3.156):

$$\exp\left(-\frac{1}{4}\,(\Omega\mathbf{r})^{\intercal}\,(V+\alpha X)\Omega\mathbf{r}+i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\leq\exp\left(-\frac{1}{4}\,(\Omega\mathbf{r})^{\intercal}\,\lambda^*_{\min}\Omega\mathbf{r}+i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\,. \tag{3.158}$$

The above upper bound is independent of α. We now focus on the derivatives of the exponential, in particular we want to dominate the following expression:

$$\left|\partial_b \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right| . \tag{3.159}$$

To do so, we compute the polynomial generated by the derivatives, i.e. we define $\mathcal{Q}_b(\alpha,\mathbf{r})$ through the following formula:

$$\partial_b \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right) = \mathcal{Q}_b(\alpha,\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right) . \tag{3.160}$$

It is easy to notice that $\mathcal{Q}_b(\alpha,\mathbf{r})$ is continuous in α, so we define

$$\mathcal{Q}_b^*(\mathbf{r}) \coloneqq \max_{\alpha\in[0,1]} |\mathcal{Q}_b(\alpha,\mathbf{r})| . \tag{3.161}$$

The above function is non-negative and scales at most polynomially in $\mathbf{r}$. In addition, it is independent of α. Thus, we have

$$\begin{aligned} &\left|\partial_b \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right| \\ &\leq \mathcal{Q}_b^*(\mathbf{r}) \left|\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right| . \end{aligned} \tag{3.162}$$

Therefore, for any $\mathbf{r}\in\mathbb{R}^{2n}$, we are able to give the follwing upper bound:

$$\begin{aligned} &(1-\alpha)^k \left|\mathbf{r}^a \partial_b \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right| \\ &\leq (1-\alpha)^k \left|\mathbf{r}^a \left(\mathscr{P}_b'(\mathbf{r}) + \mathcal{Q}_b^*(\mathbf{r})\right) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal \lambda_{\min}^* \Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right| \\ &\leq (1-\alpha)^k M_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b} . \end{aligned} \tag{3.163}$$

Here we have $\mathscr{P}_b'(\mathbf{r}) \coloneqq \partial_b \mathscr{P}(\mathbf{r})$. Furthermore, we obtain

$$M_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b} \coloneqq \sup_{\mathbf{r}\in\mathbb{R}^{2n}} \left|\mathbf{r}^a \left(\mathscr{P}_b'(\mathbf{r}) + \mathcal{Q}_b^*(\mathbf{r})\right) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal \lambda_{\min}^* \Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right| < \infty . \tag{3.164}$$

The last quantity is finite because we are taking the supremum of the absolute value of a Schwartz function. We are now able to define the non-negative dominating function as

$$\xi_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b,k}(\alpha) \coloneqq (1-\alpha)^k M_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b} . \tag{3.165}$$

It clearly is integrable. Moreover, we have

$$\int_0^1 \xi_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b,k}(\alpha)\mathrm{d}\alpha = \frac{M_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b}}{k+1} < \infty . \tag{3.166}$$

Now, we need to prove that we can exchange the derivatives and the integral. To this aim we initially consider $|b|=1$; so, we have to find a dominating function. Luckily, $\xi_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},0,b,k}(\alpha)$ does the job. Then, we can apply the Leibniz integral rule to exchange the derivative and the integral. For $|b|>1$ we iterate the same argument. After that, for any V, any X, any a and b, any k, and any $\mathbf{m}$ and $\mathbf{l}$, we can estimate the expression in (3.155) as

$$\begin{aligned} &\sup_{\mathbf{r}\in\mathbb{R}^{2n}} \left|\int_0^1 \mathbf{r}^a \partial_b (1-\alpha)^k \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\mathrm{d}\alpha\right| \\ &\leq \sup_{\mathbf{r}\in\mathbb{R}^{2n}} \int_0^1 \left|\mathbf{r}^a \partial_b (1-\alpha)^k \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right|\mathrm{d}\alpha \\ &\leq \int_0^1 (1-\alpha)^k \sup_{\mathbf{r}\in\mathbb{R}^{2n}} \left|\mathbf{r}^a \partial_b \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal\Omega\mathbf{r}\right)\right|\mathrm{d}\alpha \\ &\leq \int_0^1 \xi_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},a,b}(\alpha)\mathrm{d}\alpha < \infty . \end{aligned} \tag{3.167}$$

Hence, we conclude. □

After that, we would like to lift up these regularity properties to the Schwartz operators. We can easily do that thanks to the previous Lemma combined with Proposition 129. In this way we obtain the following Corollary:

Corollary 132. *For any $n, k \in \mathbb{N}$, for any $\mathbf{m}, \mathbf{l} \in \mathbb{R}^{2n}$, for any polynomial with complex coefficients $\mathscr{P}(\mathbf{r})$, for any $V > 0$, and any X symmetric matrix such that $V + X > 0$, it holds that*

$$\left\| \mathscr{W} \left(\int_0^1 (1-\alpha)^k \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal (V + \alpha X) \Omega \mathbf{r} + i(\mathbf{m} + \alpha \mathbf{l})^\intercal \Omega \mathbf{r} \right) \mathrm{d}\alpha \right) \right\|_1 < \infty \,. \tag{3.168}$$

Here, the linear map $\mathscr{W}$ is the Fourier–Weyl transform defined in Section 3.8.1.

Proof. The strategy we employ to prove the above inequality is to prove that the operator

$$\mathscr{W} \left(\int_0^1 (1-\alpha)^k \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal (V + \alpha X) \Omega \mathbf{r} + i(\mathbf{m} + \alpha \mathbf{l})^\intercal \Omega \mathbf{r} \right) \mathrm{d}\alpha \right) \tag{3.169}$$

is a Schwartz operator. Then, the bound follows directly from Proposition 128. Due to Lemma 131 we know that, for any $\mathscr{P}(\mathbf{r})$, V, X, $\mathbf{m}$, $\mathbf{l}$, and k, the function

$$f_{\mathscr{P},V,X,\mathbf{m},\mathbf{l},0,0,k}(\mathbf{r}) = \int_0^1 (1-\alpha)^k \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal (V + \alpha X) \Omega \mathbf{r} + i(\mathbf{m} + \alpha \mathbf{l})^\intercal \Omega \mathbf{r} \right) \mathrm{d}\alpha \tag{3.170}$$

belongs to the Schwartz space. Hence, by invoking Proposition 129, we conclude. □

We have proved that the operator obtained by the Fourier–Weyl transform of any integral remainder of the Taylor expansion of a Gaussian function is a Schwartz operator, therefore it is a trace-class operator. Now, we have to prove some regularity properties of the terms of the Taylor expansion of a generic Gaussian function. To do so, we define the positive cone of the real symmetric matrices as

$$\mathscr{P}_{2n}^+(\mathbb{R}) := \left\{ M \in \mathbb{R}^{2n,2n} : M^\intercal = M,\, M > 0 \right\} . \tag{3.171}$$

Then, we state the following Lemma:

Lemma 133. *For any $n \in \mathbb{N}$, symmetric $V > 0$, $\mathbf{m} \in \mathbb{R}^{2n}$, and any polynomial $\mathscr{P}(\mathbf{r})$ with complex coefficients, the map $\Phi_{\mathscr{P}} : \mathscr{P}_{2n}^+(\mathbb{R}) \times \mathbb{R}^{2n} \to \mathcal{S}(\mathbb{R}^{2n})$, defined as*

$$\Phi_{\mathscr{P}}[V, \mathbf{m}] := \mathcal{P}(\mathbf{r}) \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal V \Omega \mathbf{r} + i \mathbf{m}^\intercal \Omega \mathbf{r} \right) , \tag{3.172}$$

is continuous in the Fréchet topology of the Schwartz space.

Proof. According to Definition 124, in order to prove continuity in the Schwartz space, we need to show that the above map is continuous with respect to all of the seminorms in (125); that is, we need to show that for all $a, b \in \mathbb{N}^k$ it holds that

$$\begin{aligned} \lim_{(X,\mathbf{l}) \to (0,0)} \sup_{\mathbf{r} \in \mathbb{R}^{2n}} \Bigg| \mathbf{r}^a \partial_b \mathscr{P}(\mathbf{r}) \Bigg(& \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal (V + X) \Omega \mathbf{r} + i(\mathbf{m} + \mathbf{l})^\intercal \Omega \mathbf{r} \right) \\ & - \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal V \Omega \mathbf{r} + i \mathbf{m}^\intercal \Omega \mathbf{r} \right) \Bigg) \Bigg| = 0 \,. \end{aligned} \tag{3.173}$$

Here X is always a $2n \times 2n$ real symmetric matrix. In Proposition 130 we write the Taylor expansion of a Gaussian function, then by truncating this expression to the first order and using the integral remainder of the Taylor formula, we obtain

$$\begin{aligned} & \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal (V + X) \Omega \mathbf{r} + i(\mathbf{m} + \mathbf{l})^\intercal \Omega \mathbf{r} \right) \\ & = \exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal V \Omega \mathbf{r} + i \mathbf{m}^\intercal \Omega \mathbf{r} \right) \\ & \quad + \int_0^1 \left[\exp\left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal (V + \alpha X) \Omega \mathbf{r} + i(\mathbf{m} + \alpha \mathbf{l})^\intercal \Omega \mathbf{r} \right) \left(-\frac{1}{4} (\Omega \mathbf{r})^\intercal X \Omega \mathbf{r} + i \mathbf{l}^\intercal \Omega \mathbf{r} \right) \right] \mathrm{d}\alpha \,. \end{aligned} \tag{3.174}$$

Before we can prove the expression (3.173), we need to prove that for any multi-index b the following relation holds

$$\begin{aligned}&\partial_b \mathscr{P}(\mathbf{r}) \int_0^1 \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal \Omega\mathbf{r}\right)\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r} + i\mathbf{l}^\intercal\Omega\mathbf{r}\right)\mathrm{d}\alpha \\ &= \int_0^1 \partial_b \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal \Omega\mathbf{r}\right)\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r} + i\mathbf{l}^\intercal\Omega\mathbf{r}\right)\mathrm{d}\alpha\,.\end{aligned} \tag{3.175}$$

To do so, we recall the first part of the proof of Lemma 131, where the above expression is proved in greater generality. We are now ready to prove the main statement of this Lemma. By exchanging the integral and the derivative and then plugging equation (3.174) into (3.173), we obtain

$$\begin{aligned}&\lim_{(X,\mathbf{l})\to(0,0)} \sup_{\mathbf{r}\in\mathbb{R}^{2n}} \left|\int_0^1 \mathbf{r}^a \partial_b \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal \Omega\mathbf{r}\right)\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r} + i\mathbf{l}^\intercal\Omega\mathbf{r}\right)\mathrm{d}\alpha\right| \\ &\leq \lim_{(X,\mathbf{l})\to(0,0)} \frac{1}{4}\sum_{j,k=1}^{2n} |\tilde{X}_{jk}| \sup_{\mathbf{r}\in\mathbb{R}^{2n}} \left|\int_0^1 \mathbf{r}^a \partial_b \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal \Omega\mathbf{r}\right) r_j r_k \mathrm{d}\alpha\right| \\ &\quad + \lim_{(X,\mathbf{l})\to(0,0)} \sum_{j=1}^{2n} |\tilde{l}_j| \sup_{\mathbf{r}\in\mathbb{R}^{2n}} \left|\int_0^1 \mathbf{r}^a \partial_b \mathscr{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal (V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^\intercal \Omega\mathbf{r}\right) r_j \mathrm{d}\alpha\right|\,.\end{aligned} \tag{3.176}$$

Here $\tilde{X} \coloneqq \Omega^\intercal X \Omega$, and $\tilde{\mathbf{l}} \coloneqq \Omega^\intercal \mathbf{l}$. Since we can choose X arbitrarily small in any given norm, we have that $V + \alpha X > 0$ for any $\alpha \in [0,1]$. Both $\tilde{X}$ and $\tilde{\mathbf{l}}$ vanish as the pair $(X,\mathbf{l})$ goes to zero, while the suprema of the integrals are always bounded thanks to Lemma 131. Hence, we conclude. □

Corollary 134. *For any $n \in \mathbb{N}$, $V > 0$, $\mathbf{m} \in \mathbb{R}^{2n}$ and any polynomial $\mathscr{P}(\mathbf{r})$ with complex coefficients the map $\Psi_{\mathscr{P}} : \mathscr{P}^+_{2n}(\mathbb{R}) \times \mathbb{R}^{2n} \to \mathcal{S}(\mathscr{H})$, defined as*

$$\begin{aligned}\Psi_{\mathscr{P}}[V,\mathbf{m}] &\coloneqq \mathscr{W}\left(\mathcal{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal V\Omega\mathbf{r} + i\mathbf{m}^\intercal\Omega\mathbf{r}\right)\right) \\ &= (2\pi)^{-n}\int_{\mathbb{R}^{2n}} \mathcal{P}(\mathbf{r}) \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal V\Omega\mathbf{r} + i\mathbf{m}^\intercal\Omega\mathbf{r}\right)\hat{D}_{-\mathbf{r}}\mathrm{d}\mathbf{r}\,,\end{aligned} \tag{3.177}$$

is continuous in the Fréchet topology of the Schwartz operators.

Proof. Due to Proposition 129 we know that it is a continuous isomorpohism between the Schwartz space and the set of Schwartz operators. In Lemma 133 we have proved that a polynomial times a Gaussian function is continuous in the topology of the Schwartz space. Hence, we conclude. □

Equipped with these results we can now prove the main result of this section concerning the regularity of the Gaussian states as functions of their statistical moments.

Theorem 135. *For any number of modes $n \in \mathbb{N}$ the map acting as $(V,\mathbf{m}) \mapsto \omega(V,\mathbf{m})$, where*

$$\omega(V,\mathbf{m}) \coloneqq (2\pi)^{-n}\int_{\mathbb{R}^{2n}} \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal V(\Omega\mathbf{r}) + i\mathbf{m}^\intercal\Omega\mathbf{r}\right)\hat{D}_{-\mathbf{r}}\mathrm{d}\mathbf{r}\,, \tag{3.178}$$

is continuously differentiable; that is, it belongs to $C^1\left(\mathscr{P}^+_{2n}(\mathbb{R}) \times \mathbb{R}^{2n} \to \mathscr{T}(\mathscr{H})\right)$. Moreover, for any V and $\mathbf{m}$ the Jacobian operator $\mathcal{J}_{V,\mathbf{m}}$ acts on the variations X and $\mathbf{l}$ as

$$\mathcal{J}_{V,\mathbf{m}}[X,\mathbf{l}] = \mathscr{W}\left(\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal V\Omega\mathbf{r} + i\mathbf{m}^\intercal\Omega\mathbf{r}\right)\left(-\frac{1}{4}(\Omega\mathbf{r})^\intercal X\Omega\mathbf{r} + i\mathbf{l}^\intercal\Omega\mathbf{r}\right)\right) = \partial_{X,\mathbf{l}}\,\rho(V,\mathbf{m})\,. \tag{3.179}$$

Where X is a generic $2n \times 2n$ symmetric matrix, $\mathbf{l} \in \mathbb{R}^{\mathbf{2n}}$, and $\partial_{X,\mathbf{l}}\,\rho(V,\mathbf{m})$ is the directional derivative defined in Equation 3.79.

Proof. First, we do not specify the norm of the set $\mathscr{P}_{2n}^{+}(\mathbb{R}) \times \mathbb{R}^{2n}$, because it is an open subset of a finite-dimensional vector space, so all norms are equivalent; while in the set of trace-class operators $\mathscr{T}(\mathscr{H})$ we use the operator trace-norm $\|\cdot\|_1$. We start by noticing that $\omega(V, \mathbf{m})$ is a Schwartz operator, so for any V and $\mathbf{m}$ it is a trace-class operator. In Proposition 130 we recall the Taylor expansion of the Gaussian function, for any symmetric real matrix X and vector $\mathbf{l}$. We call the first of the expansion as

$$g^{(1)}_{V,X,\mathbf{m},\mathbf{l}}(\mathbf{r}) \coloneqq \exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal} V \Omega\mathbf{r} + i\mathbf{m}^{\intercal}\Omega\mathbf{r}\right)\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal} X \Omega\mathbf{r} + i\mathbf{l}^{\intercal}\Omega\mathbf{r}\right). \tag{3.180}$$

While, we take the second-order integral remainder and we call it

$$\begin{aligned} g^{(2)}_{V,X,\mathbf{m},\mathbf{l}}(\mathbf{r}) \coloneqq \int_0^1 \Bigg[& (1-\alpha)\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}(V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right) \\ & \times\left(\frac{1}{32}\left[(\Omega\mathbf{r})^{\intercal} X \Omega\mathbf{r}\right]^2 - \frac{1}{2}\left(\mathbf{l}^{\intercal}\Omega\mathbf{r}\right)^2 - \frac{i}{4}\left[(\Omega\mathbf{r})^{\intercal} X \Omega\mathbf{r}\right]\left[\mathbf{l}^{\intercal}\Omega\mathbf{r}\right]\right)\Bigg]\mathrm{d}\alpha\,. \end{aligned} \tag{3.181}$$

In order to prove the differentiability of $\omega(V, \mathbf{m})$ we have to evaluate the following limit and prove that it vanishes,

$$\lim_{\|(X,\mathbf{l})\|\to 0} \frac{\left\|\omega(V+X, \mathbf{m}+\mathbf{l}) - \omega(V,\mathbf{m}) - \mathscr{W}\left(g^{(1)}_{V,X,\mathbf{m},\mathbf{l}}\right)\right\|_1}{\|(X,\mathbf{l})\|}, \tag{3.182}$$

where X is always a real $2n \times 2n$ matrix. We know that any polynomial multiplied for a Gaussian function is a Schwartz function, so, if we apply the Fourier–Weyl transform, we always obtain a Schwartz operator, therefore a trace-class operator. In particular, this is true for any term of the Taylor expansion in Proposition 130; therefore $\mathscr{W}\left(g^{(1)}_{V,X,\mathbf{m},\mathbf{l}}\right)$ is continuous. Moreover, by Lemma 134 every term of the expansion is also continuous in the set of Schwartz operators as a function of the statistical moments, in addition Proposition 128 tells us that continuity in the set of Schwartz operators implies continuity of trace-class operators, then every term of the expansion is continuous in trace-norm as a function of the statistical moments $(V, \mathbf{m})$. Finally $g^{(1)}_{V,X,\mathbf{m},\mathbf{l}}$ is linear in $(V, \mathbf{m})$ by construction. We now have to deal with the term $\mathscr{W}\left(g^{(2)}_{V,X,\mathbf{m},\mathbf{l}}\right)$. We want to give an estimate to the following norm:

$$\begin{aligned} & \left\|\mathscr{W}\left(g^{(2)}_{V,X,\mathbf{m},\mathbf{l}}\right)\right\|_1 \\ & \leq \left\|\mathscr{W}\left(\int_0^1 (1-\alpha)\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}(V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\frac{1}{32}\left[(\Omega\mathbf{r})^{\intercal} X \Omega\mathbf{r}\right]^2 \mathrm{d}\alpha\right)\right\|_1 \\ & \quad + \left\|\mathscr{W}\left(\int_0^1 (1-\alpha)\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}(V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\frac{1}{2}\left(\mathbf{l}^{\intercal}\Omega\mathbf{r}\right)^2 \mathrm{d}\alpha\right)\right\|_1 \\ & \quad + \left\|\mathscr{W}\left(\int_0^1 (1-\alpha)\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}(V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)\frac{i}{4}\left[(\Omega\mathbf{r})^{\intercal} X \Omega\mathbf{r}\right]\left[\mathbf{l}^{\intercal}\Omega\mathbf{r}\right]\mathrm{d}\alpha\right)\right\|_1 \\ & \leq \frac{1}{32}\|\tilde{X}\|_\infty^2 \\ & \quad \sum_{j,k,l,m=1}^{2n}\left\|\mathscr{W}\left(\int_0^1 (1-\alpha)\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}(V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)(B\mathbf{r})_j(B\mathbf{r})_k(B\mathbf{r})_l(B\mathbf{r})_m\right)\right\|_1 \\ & \quad + \frac{1}{2}\|\mathbf{l}\|_\infty^2 \sum_{j,k=1}^{2n}\left\|\mathscr{W}\left(\int_0^1 (1-\alpha)\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}(V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)(\Omega\mathbf{r})_j(\Omega\mathbf{r})_k\right)\right\|_1 \\ & \quad + \frac{1}{4}\|\tilde{X}\|_\infty\|\mathbf{l}\|_\infty \\ & \quad \sum_{j,k,l=1}^{2n}\left\|\mathscr{W}\left(\int_0^1 (1-\alpha)\exp\left(-\frac{1}{4}(\Omega\mathbf{r})^{\intercal}(V+\alpha X)\Omega\mathbf{r} + i(\mathbf{m}+\alpha\mathbf{l})^{\intercal}\Omega\mathbf{r}\right)(\Omega\mathbf{r})_j(B\mathbf{r})_k(B\mathbf{r})_l\right)\right\|_1. \end{aligned} \tag{3.183}$$

Here $\tilde{X} = \Omega^\intercal X \Omega$ as in Lemma 133, and B is an orthogonal matrix such that $B\tilde{X}B^\intercal = D$, with D diagonal. We recall that $\|\tilde{X}\|_\infty = \|X\|_\infty$. The three sums in the above expression are all bounded thanks to Corollary 132. Moreover, since we $(X, \mathbf{l})$ belongs to a finite-dimensional vector space, then there exists a constant K such that $\|(X, \mathbf{l})\| \leq K(\|X\|_\infty + \|\mathbf{l}\|_\infty)$. Given all these pieces of information, we can finally evaluate the limit (3.182) as

$$\begin{aligned}
&\lim_{\|(X,\mathbf{l})\|\to 0} \frac{\left\|\omega(V+X, \mathbf{m}+\mathbf{l}) - \omega(V,\mathbf{m}) - \mathscr{W}\left(g^{(1)}_{V,X,\mathbf{m},\mathbf{l}}\right)\right\|_1}{\|(X,\mathbf{l})\|} \\
&= \lim_{\|(X,\mathbf{l})\|\to 0} \frac{\left\|\mathscr{W}\left(g^{(2)}_{V',X,\mathbf{m}',\mathbf{l}}\right)\right\|_1}{\|(X,\mathbf{l})\|} \\
&\leq \lim_{\|(X,\mathbf{l})\|\to 0} \frac{C_1\|X\|^2_\infty + C_2\|\mathbf{l}\|^2_\infty + C_3\|X\|_\infty\|\mathbf{l}\|_\infty}{K(\|X\|_\infty + \|\mathbf{l}\|_\infty)} = 0\,.
\end{aligned} \tag{3.184}$$

Here C_1, C_2, and C_3 are positive constants. Since $\mathscr{W}\left(g^{(1)}_{V,X,\mathbf{m},\mathbf{l}}\right)$ is continuous in trace-norm as a function of $(V, \mathbf{m})$ we have proved that the map $(V, \mathbf{m}) \mapsto \omega(V, \mathbf{m})$ belongs to $C^1\left(\mathscr{P}^+_{2n}(\mathbb{R}) \times \mathbb{R}^{2n} \to \mathscr{T}(\mathscr{H})\right)$. □

The theorem above allows us to write any difference of Gaussian state as the first-order integral remainder.

Corollary 136. *For any V and W quantum covariance matrices, and $\mathbf{m}, \mathbf{t} \in \mathbb{R}^{2n}$, the difference between two Gaussian states reads:*

$$\rho(W, \mathbf{t}) - \rho(V, \mathbf{m}) = \int_0^1 \partial_{X,\mathbf{x}}\rho(V + \alpha X, \mathbf{m} + \alpha\mathbf{x})\mathrm{d}\alpha\,, \tag{3.185}$$

where $X \coloneqq W - V$ and $\mathbf{x} \coloneqq \mathbf{t} - \mathbf{m}$.

Proof. A straightforward consequence of Theorem 135 is that we can write the difference of two Gaussian states as the first-order integral remainder of the Taylor formula [86]. The only caveat we have to address is the fact that the integrand must be integrable. Since any Gaussian state $\rho(V, \mathbf{m})$ in continuously differentiable as a function of the pair $(V, \mathbf{m})$ considering the trace-norm, we have that, for any X and $\mathbf{x}$, any directional derivative of a Gaussian state $\partial_{X,\mathbf{x}}\rho(V), \mathbf{m})$ is continuous in trace-norm. Then, it is also Bochner integrable. Hence, we conclude. □

Finally, we would like to stress that the tools used in the proof suffices to prove that Gaussian states are infinitely differentiable:

Remark 137. *It is possible to extend the proof of Theorem 135 to prove that, for any number of modes $n \in \mathbb{N}$, the map acting as $(V, \mathbf{m}) \mapsto \omega(V, \mathbf{m})$ is infinitely differentiable; that is, it belongs to $C^\infty\left(\mathscr{P}^+_{2n}(\mathbb{R}) \times \mathbb{R}^{2n} \to \mathscr{T}(\mathscr{H})\right)$.*

3.8.4 Alternative approach based on the Gaussian noise channel

In this section, we present an alternative approach to establish stringent upper bounds on the trace distance between Gaussian states in terms of the norm distance of their first moments and covariance matrices. The forthcoming Theorem 138 provides an upper bound that, while less tight than the one established using the derivative approach in Section 3.8.2, is proved using an insightful different method. This approach takes advantage of novel properties of the Gaussian noise channel introduced in Subsection 3.8.1, which may hold independent interest. We now state the bound derived in this section.

Theorem 138. *Let $\rho(V, \mathbf{m})$ and $\rho(W, \mathbf{t})$ be Gaussian states with covariance matrices V, W and first moments $\mathbf{t}, \mathbf{m}$, respectively. Then, it holds that*

$$\frac{1}{2} \left\| \rho(V, \mathbf{m}) - \rho(W, \mathbf{t}) \right\|_1 \leq \frac{1+\sqrt{5}}{8} \max\left\{ \|V\|_\infty, \|W\|_\infty \right\} \|V - W\|_1 + \sqrt{1 - e^{-\frac{1}{2}\min\{\|V\|_\infty, \|W\|_\infty\} \|\mathbf{t}-\mathbf{m}\|_2^2}} \,. \tag{3.186}$$

Proof. It suffices to apply the forthcoming Theorem 139 with the choice $\bar{\phi}(x) := \frac{1+\sqrt{5}}{2} x$. Indeed, using the elementary bounds $\sinh(x/2) \leq \frac{1}{2} x e^{x/2}$ and $1 - e^{-\frac{x^2}{1+4x}} \leq x^2 \leq x^2 e^x$, it is not difficult to show that

$$\phi(x) \leq \frac{1+\sqrt{5}}{2} x \,, \tag{3.187}$$

where the function $\phi(x)$ is defined in the statement of Theorem 139 (note that $\frac{1+\sqrt{5}}{2} x$ is in fact the dominant term in the Taylor expansion of ϕ around 0, i.e. $\phi(x) = \frac{1+\sqrt{5}}{2} x + O\left(x^2\right)$ for $x \to 0$.) □

The forthcoming Theorem 139 is the key result of this section, which directly implies Theorem 138.

Theorem 139. *Let $\bar{\phi} : [0, \infty) \to \mathbb{R}^+$ be a concave function such that $\bar{\phi}(x) \geq \phi(x)$ for all $x \geq 0$, where the function $\phi : [0, \infty) \to [0, 1]$ is defined by*

$$\phi(x) := \frac{1 - e^{-x}}{2} + e^{-x/2} \sqrt{\sinh^2(x/2) + 1 - e^{-\frac{x^2}{1+4x}}} \,. \tag{3.188}$$

Let $\rho(V, \mathbf{m})$ and $\rho(W, \mathbf{t})$ be Gaussian states with covariance matrices V, W and first moments $\mathbf{t}, \mathbf{m}$, respectively. Then, it holds that

$$\frac{1}{2} \left\| \rho(V, \mathbf{m}) - \rho(W, \mathbf{t}) \right\|_1 \leq 2\, \bar{\phi}\left(\frac{1}{8} \max\left\{ \|V\|_\infty, \|W\|_\infty \right\} \|V - W\|_1 \right) + \sqrt{1 - e^{-\frac{1}{2}\min\{\|V\|_\infty, \|W\|_\infty\} \|\mathbf{t}-\mathbf{m}\|_2^2}} \,. \tag{3.189}$$

Remark 140. By looking at a plot of the function ϕ, it seems clear that ϕ itself is concave. However, proving this explicitly seems quite difficult. If this could be done, then in the above Theorem 139 we could take $\bar{\phi} := \phi$. Since $\bar{\phi}$ is bounded, this would give us a bound on the trace distance between any two Gaussian states that is both faithful and bounded. Note that the right-hand side of (3.4), on the contrary, is not bounded. In any case, we can definitely take $\bar{\phi}(x) := \frac{1+\sqrt{5}}{2} x$, which yields Theorem 138 above.

Before proving Theorem 139, let us establish some preliminary results about the *Gaussian noise channel* $\mathcal{N}_K$, defined in subsection 3.8.1. Here, we recall that, given $K \geq 0$, the Gaussian noise channel $\mathcal{N}_K$ is a Gaussian channel that acts on covariance matrices as $V \mapsto V + K$ and leaves unchanged the first moments.

Lemma 141. *Let K be a $2n \times 2n$ real positive semi-definite matrix. Let $\bar{\phi} : [0, \infty) \to \mathbb{R}^+$ be a function defined as in Theorem 139. Then for all $M \geq 1$ it holds that*

$$\sup_{\rho_{\mathrm{G}}:\ \|V(\rho_{\mathrm{G}})\|_\infty \leq M} \frac{1}{2} \left\| \left(\mathcal{N}_K - \mathrm{Id}\right)(\rho_{\mathrm{G}}) \right\|_1 \leq \bar{\phi}\left(\frac{1}{4} M \operatorname{Tr} K \right) \,, \tag{3.190}$$

where the supremum is over all the Gaussian states ρ_G whose covariance matrix $V(\rho_{\mathrm{G}})$ has operator norm at most M. In particular, it holds that

$$\sup_{\rho_{\mathrm{G}}:\ \|V(\rho_{\mathrm{G}})\|_\infty \leq M} \frac{1}{2} \left\| \left(\mathcal{N}_K - \mathrm{Id}\right)(\rho_{\mathrm{G}}) \right\|_1 \leq \frac{1+\sqrt{5}}{8}\, M \operatorname{Tr} K \,. \tag{3.191}$$

Proof. Let us start by upper bounding the quantity $\frac{1}{2}\left\|(\mathcal{N}_K - \mathrm{Id})\left(|0\rangle\langle 0|\right)\right\|_1$. Note that $\mathcal{N}_K(|0\rangle\langle 0|)$ is a Gaussian state with zero first moment and covariance matrix $\mathbb{1} + K$. Let us consider the Williamson decomposition

$$\mathbb{1} + K = SDS^\intercal\,, \qquad D = \bigoplus_j d_j \mathbb{1}_2\,, \tag{3.192}$$

with S symplectic and $d_j \geq 1$. Moreover, let us define

$$\gamma := (\mathbb{1} + K)\#\Big(\Omega(\mathbb{1} + K)^{-1}\Omega^\intercal\Big)\,, \tag{3.193}$$

where we introduced the *matrix geometric mean* defined as $A\#B := \sqrt{A}\sqrt{A^{-1/2}BA^{-1/2}}\sqrt{A}$ for all positive definite matrices A, B [87]. By exploiting that for any invertible matrix X it holds that [87]

$$(XAX^\dagger)\#(XBX^\dagger) = X(A\#B)X^\dagger\,, \tag{3.194}$$

it follows that

$$\gamma = SS^\intercal\,. \tag{3.195}$$

In particular, note that γ is the covariance matrix associated with the pure Gaussian state $U_S|0\rangle$. Introducing a parameter $\mu \geq 0$ to be fixed later, we now write that

$$\begin{aligned}
\frac{1}{2}\left\|(\mathcal{N}_K - \mathrm{Id})\left(|0\rangle\langle 0|\right)\right\|_1 &= \frac{1}{2}\left\|U_S^\dagger \mathcal{N}_K(|0\rangle\langle 0|)U_S - U_S^\dagger|0\rangle\langle 0|U_S\right\|_1 \\
&\leq \frac{1}{2}\left\|U_S^\dagger \mathcal{N}_K(|0\rangle\langle 0|)U_S - \mu|0\rangle\langle 0|\right\|_1 + \frac{1}{2}\left\|\mu|0\rangle\langle 0| - U_S^\dagger|0\rangle\langle 0|U_S\right\|_1 \\
&\overset{\text{(i)}}{=} \frac{1-\mu}{2} + \left(\mu - \prod_j \frac{2}{1+d_j}\right)_+ + \sqrt{\left(\frac{1+\mu}{2}\right)^2 - \mu|\langle 0|U_S|0\rangle|^2} \\
&\overset{\text{(ii)}}{=} \frac{1-\mu}{2} + \left(\mu - \prod_j \frac{2}{1+d_j}\right)_+ + \sqrt{\left(\frac{1+\mu}{2}\right)^2 - \mu\frac{2^n}{\sqrt{\det(\mathbb{1}+\gamma)}}} \\
&\overset{\text{(iii)}}{\leq} \frac{1-\mu}{2} + \left(\mu - e^{-\frac{1}{4}\mathrm{Tr}\,K}\right)_+ + \sqrt{\left(\frac{1+\mu}{2}\right)^2 - \mu\frac{2^n}{\sqrt{\det(\mathbb{1}+\gamma)}}}\,.
\end{aligned} \tag{3.196}$$

The justification of the above manipulations is as follows. In (i) we used the fact that $U_S^\dagger \mathcal{N}_K(|0\rangle\langle 0|)U_S$ is a Gaussian state with zero first moment and covariance matrix equal to $S^{-1}(\mathbb{1}+K)S^{-\intercal} = D$, and it is thus a product of thermal states with mean photon number $(d_j - 1)/2$ [2]; the trace distance from the vacuum state is easy to compute, given that the two density operators commute. In (ii) we substituted the formula [2, Eq (4.51)]. In (iii), instead, we observed that

$$\begin{aligned}
\ln\prod_j \frac{1+d_j}{2} &= \sum_j \ln\left(1 + \frac{1}{2}(d_j - 1)\right) \leq \frac{1}{2}\sum_j (d_j - 1) \\
&= \frac{1}{4}\left(\mathrm{Tr}\,D - 2n\right) \leq \frac{1}{4}\Big(\mathrm{Tr}[\mathbb{1}+K] - 2n\Big) = \frac{1}{4}\,\mathrm{Tr}\,K\,,
\end{aligned} \tag{3.197}$$

where the inequality $\mathrm{Tr}\,D \leq \mathrm{Tr}[\mathbb{1} + K]$ follows from [88, Theorem 5(i)].

We can simplify (3.196) by choosing $\mu = e^{-\frac{1}{4}\mathrm{Tr}\,K}$, which gives immediately

$$\begin{aligned}
\frac{1}{2}\left\|(\mathcal{N}_K - \mathrm{Id})\left(|0\rangle\langle 0|\right)\right\|_1 &\leq \frac{1 - e^{-\frac{1}{4}\mathrm{Tr}\,K}}{2} + \sqrt{\left(\frac{1+e^{-\frac{1}{4}\mathrm{Tr}\,K}}{2}\right)^2 - e^{-\frac{1}{4}\mathrm{Tr}\,K}\frac{2^n}{\sqrt{\det(\mathbb{1}+\gamma)}}} \\
&= \frac{1 - e^{-\frac{1}{4}\mathrm{Tr}\,K}}{2} + \sqrt{\left(\frac{1-e^{-\frac{1}{4}\mathrm{Tr}\,K}}{2}\right)^2 + e^{-\frac{1}{4}\mathrm{Tr}\,K}\left(1 - \frac{2^n}{\sqrt{\det(\mathbb{1}+\gamma)}}\right)}\,.
\end{aligned} \tag{3.198}$$

Numerically, this choice of μ seems to be optimal.

We now set out to manipulate the term containing the determinant of $\mathbb{1}+\gamma$. Since γ is a pure covariance matrix, it has spectrum of the form $\bigcup_j \left\{g_j,\, 1/g_j\right\}$, with $g_j > 0$. Thus,

$$\begin{aligned}
\sqrt{\det(\mathbb{1}+\gamma)} &= \prod_j \sqrt{2+g_j+1/g_j} && (3.199)\\
&\overset{\text{(iv)}}{\leq} \prod_j 2\left(1+\frac{1}{8}\left(g_j+1/g_j-2\right)\right) && (3.200)\\
&= 2^n \exp\left[\sum\nolimits_j \ln\left(1+\frac{1}{8}\left(g_j+1/g_j-2\right)\right)\right] && (3.201)\\
&\overset{\text{(v)}}{\leq} 2^n \exp\left[\frac{1}{8}\sum\nolimits_j \left(g_j+1/g_j-2\right)\right] && (3.202)\\
&= 2^n \exp\left[\frac{1}{8}\operatorname{Tr}\left[\gamma-\mathbb{1}\right]\right] && (3.203)\\
&\overset{\text{(vi)}}{\leq} 2^n \exp\left[\frac{1}{8}\operatorname{Tr}\left[\frac{\mathbb{1}+K+\Omega(\mathbb{1}+K)^{-1}\Omega^{\intercal}}{2}-\mathbb{1}\right]\right] && (3.204)\\
&\overset{\text{(vii)}}{\leq} 2^n \exp\left[\frac{1}{16}\operatorname{Tr}\left[(\mathbb{1}+K)^{-1}-\mathbb{1}+K\right]\right] && (3.205)\\
&= 2^n \exp\left[\frac{1}{16}\operatorname{Tr}\left[\frac{K^2}{\mathbb{1}+K}\right]\right] && (3.206)\\
&\overset{\text{(viii)}}{\leq} 2^n \exp\left[\frac{1}{16}\frac{(\operatorname{Tr} K)^2}{1+\operatorname{Tr} K}\right]. && (3.207)
\end{aligned}$$

The 'miracle' here has happened on the second-to-last line, where the two linear terms have cancelled out. In the above derivation: (iv) follows from the elementary inequality

$$\sqrt{2+x+1/x} \leq 2+\frac{1}{4}\left(x+1/x-2\right), \tag{3.208}$$

valid for all $x > 0$; in (v) we used simply $\ln(1+x) \leq x$; (vi) holds due to the definition of γ given in (3.195), together with the AM-GM inequality $A\#B \leq (A+B)/2$ for matrix means [87]; and in (vii) we used the cyclicity of the trace. To justify (viii), define, for an arbitrary $A \geq 0$ with spectral decomposition $A = \sum_i a_i \left|i\right\rangle\!\left\langle i\right|$, the probability distribution $p_i \coloneqq \frac{a_i}{\operatorname{Tr} A}$; then, we have that

$$\begin{aligned}
\operatorname{Tr}\left[\frac{A^2}{\mathbb{1}+A}\right] &= \sum_i \frac{a_i^2}{1+a_i}\\
&= \sum_i \frac{p_i^2(\operatorname{Tr} A)^2}{1+p_i \operatorname{Tr} A}\\
&= \frac{(\operatorname{Tr} A)^2}{1+\operatorname{Tr} A}\sum_i p_i \frac{p_i(1+\operatorname{Tr} A)}{1+p_i \operatorname{Tr} A} && (3.209)\\
&\leq \frac{(\operatorname{Tr} A)^2}{1+\operatorname{Tr} A}\sum_i p_i\\
&= \frac{(\operatorname{Tr} A)^2}{1+\operatorname{Tr} A}.
\end{aligned}$$

This completes the justification of (3.203).

Plugging (3.203) into (3.198) yields the estimate

$$\frac{1}{2}\left\|\left(\mathcal{N}_K-\mathrm{Id}\right)\left(\left|0\right\rangle\!\left\langle 0\right|\right)\right\|_1 \leq \phi\left(\frac{1}{4}\operatorname{Tr} K\right) \leq \bar{\phi}\left(\frac{1}{4}\operatorname{Tr} K\right). \tag{3.210}$$

Now, for fixed $M \geq 0$, let us consider the following quantity:

$$\sup_{\rho_{\mathrm{G}}:\ \|V(\rho_{\mathrm{G}})\|_\infty \leq M} \frac{1}{2} \left\| \left(\mathcal{N}_K - I \right) (\rho_{\mathrm{G}}) \right\|_1 , \tag{3.211}$$

where the supremum is over all Gaussian states ρ_{G} whose covariance matrix $V(\rho_{\mathrm{G}})$ has operator norm at most M. The supremum is clearly achieved when ρ_{G} has zero first moment, because displacements commute with $\mathcal{N}_K$, and also when ρ_{G} is pure, because any mixed Gaussian state ρ_{G} with covariance matrix $V(\rho_{\mathrm{G}})$ can be written as a convex combination of displaced pure Gaussian states with covariance matrices $\theta \leq V(\rho_{\mathrm{G}})$ (see Lemma 106). Let θ be a generic pure covariance matrix such that $\theta \leq V(\rho_{\mathrm{G}})$, so that $\|\theta\|_\infty \leq M$, and let ψ_{G} be the pure Gaussian state with covariance matrix θ and zero first moment. We write $\theta = ZZ^\intercal$, with Z symplectic, so that $\psi_{\mathrm{G}} = U_Z \left|0\right\rangle\!\left\langle 0\right| U_Z^\dagger$. We now have that

$$\begin{aligned}
\frac{1}{2} \left\| \left(\mathcal{N}_K - I \right) (\psi_G) \right\|_1 &\overset{\text{(ix)}}{=} \frac{1}{2} \left\| U_Z^\dagger \mathcal{N}_K \Big(U_Z \left|0\right\rangle\!\left\langle 0\right| U_Z^\dagger \Big) U_Z - \left|0\right\rangle\!\left\langle 0\right| \right\|_1 && (3.212)\\
&\overset{\text{(x)}}{=} \frac{1}{2} \left\| \mathcal{N}_{Z^{-1} K Z^{-\intercal}} (\left|0\right\rangle\!\left\langle 0\right|) - \left|0\right\rangle\!\left\langle 0\right| \right\|_1 && (3.213)\\
&\leq F\Big(Z^{-1} K Z^{-\intercal} \Big) && (3.214)\\
&\overset{\text{(xi)}}{\leq} \overline{\phi} \left(\frac{1}{4} \operatorname{Tr} Z^{-1} K Z^{-\intercal} \right) && (3.215)\\
&= \overline{\phi} \left(\frac{1}{4} \operatorname{Tr} \theta^{-1} K \right) && (3.216)\\
&\overset{\text{(xii)}}{\leq} \overline{\phi} \left(\frac{1}{4} \left\| \theta^{-1} \right\|_\infty \operatorname{Tr} K \right) && (3.217)\\
&\overset{\text{(xiii)}}{=} \overline{\phi} \left(\frac{1}{4} \left\| \theta \right\|_\infty \operatorname{Tr} K \right) && (3.218)\\
&\overset{\text{(xiv)}}{\leq} \overline{\phi} \left(\frac{1}{4} M \operatorname{Tr} K \right) . && (3.219)
\end{aligned}$$

Here: (ix) is by unitary invariance; (x) holds because the Gaussian channel given by $\rho \mapsto U_Z^\dagger \mathcal{N}_K \Big(U_Z \rho U_Z^\dagger \Big) U_Z$ acts on input covariance matrices as $V \mapsto Z^{-1} \left(Z V Z^\intercal + K \right) Z^{-\intercal} = V + Z^{-1} K Z^{-\intercal}$, does not change the first moment, and hence it coincides with $\mathcal{N}_{Z^{-1} K Z^{-\intercal}}$; in (xi) we employed (3.210), (xii) follows from Hölder's inequality, remembering that $K \geq 0$ is positive semi-definite and observing that $\overline{\phi} \geq \phi$, being concave, must be non-decreasing, because $\lim_{x \to \infty} \phi(x) = 1$; in (xiii) we noticed that $\theta^{-1} = \Omega \theta \Omega^\intercal$, because θ is symplectic; finally, in (xiv) we remembered that $\|\theta\|_\infty \leq M$, leveraging once again the fact that $\overline{\phi}$ is non-decreasing. □

We are now ready to prove Theorem 139.

Proof of Theorem 139. We use a similar idea of [40]. First, given a covariance matrix V, let $\rho(V, 0)$ be the Gaussian state with zero first moment and covariance matrix equal to V. Given two covariance matrices V and W, our crucial idea is to consider a matrix X (to be chosen later) such that

$$\begin{aligned}
X &\geq 0 , \\
X + V - W &\geq 0 ,
\end{aligned} \tag{3.220}$$

so that

$$\mathcal{N}_X (\rho(V, 0)) = \rho(X + V, 0) = \mathcal{N}_{X+V-W} (\rho(W, 0)) . \tag{3.221}$$

Consequently, we have that

$$\begin{aligned}\frac{1}{2}\|\rho(V,0)-\rho(W,0)\|_1 &\leq \frac{1}{2}\|\rho(V,0)-\mathcal{N}_X(\rho(V,0))\|_1+\frac{1}{2}\|\mathcal{N}_{X+V-W}(\rho(W,0))-\rho(W,0)\|_1\\ &\leq \frac{1}{2}\|(\mathcal{N}_X-\mathrm{Id})(\rho(V,0))\|_1+\frac{1}{2}\|(\mathcal{N}_{X+V-W}-\mathrm{Id})(\rho(W,0))\|_1\\ &\overset{\text{(i)}}{\leq} \bar{\phi}\left(\frac{1}{4}\|V\|_\infty \operatorname{Tr} X\right)+\bar{\phi}\left(\frac{1}{4}\|W\|_\infty \operatorname{Tr}[X+V-W]\right)\\ &\overset{\text{(ii)}}{\leq} 2\bar{\phi}\left(\frac{1}{8}\|V\|_\infty \operatorname{Tr} X+\frac{1}{8}\|W\|_\infty \operatorname{Tr}[X+V-W]\right)\\ &\leq 2\bar{\phi}\left(\frac{1}{8}\max(\|V\|_\infty,\|W\|_\infty)\operatorname{Tr}[2X+V-W]\right),\end{aligned} \tag{3.222}$$

where in (i) we exploited Lemma 141 and (ii) follows by concavity of $\bar{\phi}$. Now let us choose X in order to minimise $\operatorname{Tr} X$, satisfying the constraints in (3.220). That is, we choose

$$X := (W-V)_+ = \frac{W-V+|W-V|}{2}. \tag{3.223}$$

Hence, we obtain that

$$\begin{aligned}\frac{1}{2}\|\rho(V,0)-\rho(W,0)\|_1 &\leq 2\bar{\phi}\left(\frac{1}{8}\max(\|V\|_\infty,\|W\|_\infty)\operatorname{Tr}|W-V|\right)\\ &= 2\bar{\phi}\left(\frac{1}{8}\max(\|V\|_\infty,\|W\|_\infty)\|W-V\|_1\right).\end{aligned} \tag{3.224}$$

In the general case of Gaussian states with non-zero first moments, we can apply the following reasoning. Let $\rho(V,\mathbf{m})$ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. For any first moments $\mathbf{m},\mathbf{t}$ and covariance matrices V,W, we have that

$$\begin{aligned}\frac{1}{2}\|\rho(V,\mathbf{m})-\rho(W,\mathbf{t})\|_1 &\leq \frac{1}{2}\|\rho(V,\mathbf{m})-\rho(W,\mathbf{m})\|_1+\frac{1}{2}\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1\\ &= \frac{1}{2}\|\rho_{V,0}-\rho_{W,0}\|_1+\frac{1}{2}\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1\\ &\leq 2\bar{\phi}\left(\frac{1}{8}\max(\|V\|_\infty,\|W\|_\infty)\|W-V\|_1\right)+\frac{1}{2}\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1.\end{aligned} \tag{3.225}$$

Finally, by writing the Williamson decomposition $W=SDS^\intercal$ and applying Lemma 106, we have that

$$\begin{aligned}\frac{1}{2}\|\rho(W,\mathbf{m})-\rho(W,\mathbf{t})\|_1 &\overset{\text{(iii)}}{\leq} \frac{1}{2}\|\rho_{SS^\intercal,\mathbf{m}}-\rho_{SS^\intercal,\mathbf{t}}\|_1\\ &\overset{\text{(iv)}}{=} \sqrt{1-\operatorname{Tr}[\rho_{SS^\intercal,\mathbf{m}}\rho_{SS^\intercal,\mathbf{t}}]}\\ &\overset{\text{(v)}}{=} \sqrt{1-e^{-\frac{1}{2}(\mathbf{m}-\mathbf{t})^\intercal(SS^\intercal)^{-1}(\mathbf{m}-\mathbf{t})}}\\ &\leq \sqrt{1-e^{-\frac{1}{2}\|(SS^\intercal)^{-1}\|_\infty\|\mathbf{m}-\mathbf{t}\|^2}}\\ &\overset{\text{(vi)}}{=} \sqrt{1-e^{-\frac{1}{2}\|SS^\intercal\|_\infty\|\mathbf{m}-\mathbf{t}\|^2}}\\ &\overset{\text{(vii)}}{\leq} \sqrt{1-e^{-\frac{1}{2}\|W\|_\infty\|\mathbf{m}-\mathbf{t}\|^2}}.\end{aligned} \tag{3.226}$$

Here, in (iii), we applied the monotonicity of the trace norm under quantum channels; in (iv), we exploited that $\rho_{SS^\intercal,\mathbf{m}}$ and $\rho_{SS^\intercal,\mathbf{t}}$ are pure states; in (v), we employed the known formula for the overlap between Gaussian states [2, Eq (4.51)]; in (vi), we exploited that for any symplectic matrix T it holds that

$$\|T^{-1}\|_\infty = \|\Omega T^\intercal \Omega^\intercal\|_\infty = \|T^\intercal\|_\infty = \|T\|_\infty, \tag{3.227}$$

where we exploited that $T\Omega T^\intercal=\Omega$ and the fact that Ω is orthogonal; finally, in (vii), we used the fact that $SS^\intercal \leq SDS^\intercal = W$. By combining (3.225) and (3.226), we conclude the proof of (3.189). □

3.8.5 Sample complexity for estimating the covariance matrix using heterodyne measurements

In this section, we analyse an improved sample complexity upper bound for estimating the covariance matrix of a Gaussian state to a given target accuracy, measured with respect to Schatten norms. The strategy relies on the fact that performing heterodyne measurements on a Gaussian state is equivalent to sampling from a classical Gaussian probability distribution whose covariance matrix and mean vector are directly related to the true quantum covariance matrix and mean vector of the underlying Gaussian state.

The heterodyne measurement is a positive operator-valued measure (POVM) defined by the set of operators

$$\left\{ \frac{1}{(2\pi)^n} |\mathbf{r}\rangle\langle\mathbf{r}| \right\}_{\mathbf{r}\in\mathbb{R}^{2n}},$$

where $|\mathbf{r}\rangle := \hat{D}_{\mathbf{r}} |0\rangle$ is a coherent state [2], and $\hat{D}_{\mathbf{r}}$ denotes the displacement operator. This measurement constitutes an experimentally feasible scheme that is routinely implemented in CV quantum experiments [2].

For a quantum state ρ, the probability distribution associated with these POVMs is described by the function $Q_\rho(\mathbf{r}) : \mathbb{R}^{2n} \to \mathbb{R}$, defined as

$$Q_\rho(\mathbf{r}) := \frac{1}{(2\pi)^n} \langle\mathbf{r}| \rho |\mathbf{r}\rangle \ , \tag{3.228}$$

commonly referred to as the Husimi function.

It is well known [2] that for a Gaussian state ρ, the Husimi function is itself a Gaussian probability distribution. It is characterised by the mean vector $\mathbf{m}(\rho)$ and the covariance matrix $\frac{V(\rho)+\mathbb{1}}{2}$, where $V(\rho)$ is the covariance matrix of ρ. Specifically,

$$Q_\rho(\mathbf{r}) = \mathcal{N}\left(\mathbf{m}(\rho), \frac{V(\rho)+\mathbb{1}}{2}\right)(\mathbf{r}) \,, \tag{3.229}$$

where $\mathcal{N}(\mathbf{m}, V)(\mathbf{r})$ denotes a Gaussian probability distribution with mean vector $\mathbf{m}$ and covariance matrix V, explicitly given by

$$\mathcal{N}(\mathbf{m}, V)(\mathbf{r}) := \frac{\exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{m})^\intercal V^{-1}(\mathbf{r}-\mathbf{m})\right)}{(2\pi)^n \sqrt{\det V}} \,. \tag{3.230}$$

We now recall established results in the classical literature [89, 90] concerning explicit bounds for estimating the mean vector and covariance matrix from samples of a multivariate Gaussian distribution. Using Eq. (1.1) from [89], which follows from results such as the Hanson-Wright inequality [90], we state the following lemma for estimating the mean vector:

Lemma 142 (First moment estimation [89])**.** *Let $\varepsilon, \delta > 0$ denote the desired accuracy and failure probability, respectively. Consider a multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^n$ is the mean vector and $\Sigma \in \mathbb{R}^{n\times n}$ is the covariance matrix. Suppose $X_1, X_2, \ldots, X_N$ are independent and identically distributed (i.i.d.) samples drawn from $\mathcal{N}(\mu, \Sigma)$. Define the empirical mean estimator as:*

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Then, with probability at least $1-\delta$, the following bound holds:

$$\|\hat{\mu}_N - \mu\|_2 \leq \sqrt{\frac{\|\Sigma\|_1}{N}} + \sqrt{\frac{2\|\Sigma\|_\infty \log\left(\frac{1}{\delta}\right)}{N}}. \tag{3.231}$$

For estimating the covariance matrix, we use the result of [91, Example 6.3] (also see [92, Page 31]), which provides explicit sample complexity bound under the assumption of a zero-mean Gaussian distribution:

Lemma 143 (Covariance matrix estimation [91])**.** *Let $\varepsilon, \delta > 0$ denote the desired accuracy and failure probability, respectively. Consider a multivariate Gaussian distribution $\mathcal{N}(0, \Sigma)$, where $0 \in \mathbb{R}^n$ is the mean zero vector and $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance matrix. Suppose $X_1, X_2, \ldots, X_N$ are independent and identically distributed (i.i.d.) samples drawn from $\mathcal{N}(0, \Sigma)$. Define the empirical covariance matrix as:*

$$\hat{\Sigma}_N = \frac{1}{N} \sum_{i=1}^{N} X_i X_i^T.$$

Then, with probability at least $1 - \delta$, the following bound holds:

$$\|\hat{\Sigma}_N - \Sigma\|_\infty \leq \|\Sigma\|_\infty \left(2\sqrt{\frac{n}{N}} + 2\sqrt{\frac{2}{N} \ln\left(\frac{2}{\delta}\right)} + \left(\sqrt{\frac{n}{N}} + \sqrt{\frac{2}{N} \log\left(\frac{1}{\delta}\right)} \right)^2 \right). \tag{3.232}$$

Combining the two previous Lemmas, we derive explicit sample complexity bounds for estimating the covariance matrix of a multivariate Gaussian distribution without assuming that the mean vector is zero.

Lemma 144 (Estimation of covariance matrix and mean vector [90])**.** *Let $\varepsilon, \delta > 0$ denote the desired accuracy and failure probability, respectively. Consider a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^n$ is the mean vector and $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance matrix. Let $X_1, X_2, \ldots, X_N$ be i.i.d. samples drawn from $\mathcal{N}(\mu, \Sigma)$. Assume that the sample size N satisfies the condition*

$$N \geq \left(2^4 + \frac{2^{10}}{n} \right) \frac{\|\Sigma\|_\infty^2}{\varepsilon^2} \left(n + \log\left(\frac{2}{\delta}\right) \right). \tag{3.233}$$

Let $\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} X_i$ be the sample mean, and let $\hat{\Sigma}_N$ be the sample covariance matrix defined as

$$\hat{\Sigma}_N = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \hat{\mu}_N)(X_i - \hat{\mu}_N)^T. \tag{3.234}$$

Then, with probability at least $1 - \delta$, the following bounds hold:

$$\|\hat{\Sigma}_N - \Sigma\|_\infty \leq \varepsilon, \tag{3.235}$$

$$\|\hat{\mu}_N - \mu\|_2 \leq \frac{\varepsilon}{2\|\Sigma\|_\infty^{1/2}}. \tag{3.236}$$

Proof. We first consider

$$\hat{W}_N := \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)(X_i - \mu)^T. \tag{3.237}$$

We have that:

$$\|\hat{\Sigma}_N - \Sigma\|_\infty \leq \|\hat{\Sigma}_N - \hat{W}_N\|_\infty + \|\hat{W}_N - \Sigma\|_\infty. \tag{3.238}$$

We now analyse the second term $\|\hat{W}_N - \Sigma\|_\infty$. Note that the random variables $X_i - \mu$ appearing in $\hat{W}_N$ are distributed as $\mathcal{N}(0, \Sigma)$. Consequently, $\hat{W}_N$ serves as an unbiased estimator for the covariance matrix Σ, according to Lemma 143. Specifically, if

$$N \geq \left(C_1 + \frac{C_2}{n} \right) \frac{\|\Sigma\|_\infty^2}{\varepsilon^2} \left(n + \log\left(\frac{2}{\delta}\right) \right) \tag{3.239}$$

where $C_1, C_2 \geq 1$ are constants to be fixed later, then, with probability at least $1 - \delta/2$, we have

$$\begin{aligned}\|\hat{W}_N - \Sigma\|_\infty &\leq \|\Sigma\|_\infty \left(2\sqrt{\frac{n}{N}} + 2\sqrt{\frac{2}{N}\log\left(\frac{2}{\delta}\right)} + \left(\sqrt{\frac{n}{N}} + \sqrt{\frac{2}{N}\log\left(\frac{2}{\delta}\right)}\right)^2\right) \\ &\leq \left(2\sqrt{\frac{1}{C_1}} + 2\sqrt{\frac{2}{C_1 + C_2}} + \left(\sqrt{\frac{1}{C_1}} + \sqrt{\frac{2}{C_1 + C_2}}\right)^2\right)\varepsilon, \end{aligned} \tag{3.240}$$

where we used that $\|\Sigma\|_\infty \geq 1$. Before analysing the first term $\|\hat{\Sigma}_N - \hat{W}_N\|_\infty$, we consider the first moment. Using Lemma 142, we have the following bound that holds with probability $\geq 1 - \delta/2$:

$$\|\hat{\mu}_N - \mu\|_2 \leq \sqrt{\frac{\|\Sigma\|_1}{N}} + \sqrt{\frac{2\|\Sigma\|_\infty \log\left(\frac{2}{\delta}\right)}{N}} \leq \left(\sqrt{\frac{2}{C_1}} + \sqrt{\frac{2}{C_1 + C_2}}\right)\frac{\varepsilon}{\|\Sigma\|_\infty^{1/2}}. \tag{3.241}$$

By the union bound, with probability at least $1 - \delta$, both bounds hold simultaneously for the covariance matrix and the mean vector.

Finally, we consider the difference $\hat{\Sigma}_N - \hat{W}_N$. Note that this can be written as

$$\begin{aligned}\hat{\Sigma}_N - \hat{W}_N &= \frac{1}{1 - \frac{1}{N}}\left(\frac{\hat{W}_N}{N} - (\hat{\mu}_N - \mu)(\hat{\mu}_N - \mu)^T\right) \\ &= \frac{1}{1 - \frac{1}{N}}\left(\frac{\hat{W}_N - \Sigma}{N} + \frac{\Sigma}{N} - (\hat{\mu}_N - \mu)(\hat{\mu}_N - \mu)^T\right). \end{aligned} \tag{3.242}$$

Taking the operator norms on both sides and applying the triangle inequality, we obtain:

$$\begin{aligned}\|\hat{\Sigma}_N - \hat{W}_N\|_\infty &= \frac{1}{1 - \frac{1}{N}}\left(\frac{\|\hat{W}_N - \Sigma\|_\infty}{N} + \frac{\|\Sigma\|_\infty}{N} + \|(\hat{\mu}_N - \mu)(\hat{\mu}_N - \mu)^T\|_\infty\right) \\ &\leq 2\frac{\|\hat{W}_N - \Sigma\|_\infty}{N} + 2\frac{\|\Sigma\|_\infty}{N} + 2\|\hat{\mu}_N - \mu\|_2^2 \\ &\leq 2\left(2\sqrt{\frac{1}{C_1}} + 2\sqrt{\frac{2}{C_1 + C_2}} + \left(\sqrt{\frac{1}{C_1}} + \sqrt{\frac{2}{C_1 + C_2}}\right)^2\right)\frac{\varepsilon^3}{C_1 + C_2} \\ &\quad + 2\frac{\varepsilon^2}{C_1 + C_2} + 2\left(\sqrt{\frac{2}{C_1}} + \sqrt{\frac{2}{C_1 + C_2}}\right)^2\varepsilon^2. \end{aligned} \tag{3.243}$$

where we have applied the bound from Eq. (3.240) and Eq. (3.241).

Combining the bounds in Eqs. (3.238), (3.240), and (3.243), and by choosing $C_1 = 2^4, C_2 = 2^{10}$, we conclude the proof of the lemma. □

As an application of the previous lemma, we derive explicit sample complexity bounds for estimating the covariance matrix of an unknown Gaussian state using heterodyne measurements.

Theorem 145 (Explicit sample complexity bound for estimating the first moment and covariance matrix of a Gaussian state)**.** *Let $\varepsilon, \delta \in (0, 1)$ denote the desired accuracy and failure probability, respectively. Let ρ be an unknown n-mode Gaussian state, with $\mathbf{m}$ and V denoting the first moment and covariance matrix of ρ, respectively. Let N be the number of copies of the unknown state, which we assume to satisfy the condition*

$$N \geq \left(2^8 + \frac{2^{13}}{n}\right)\frac{\|V\|_\infty^2}{\varepsilon^2}\left(2n + \log\left(\frac{2}{\delta}\right)\right).$$

Then, heterodyne measurements performed on each of these N copies of the unknown state ρ suffice to construct a (valid) covariance matrix $\tilde{V}$ and a vector $\hat{\mathbb{m}}$ such that, with probability at least $1-\delta$, the following bounds hold:

$$\|\tilde{V} - V\|_\infty \leq \varepsilon, \tag{3.244}$$

$$\|\tilde{\mathbf{m}} - \mathbf{m}\|_2 \leq \frac{\varepsilon}{4\|V\|_\infty^{1/2}}. \tag{3.245}$$

Proof. By performing heterodyne measurements on the Gaussian state ρ, we sample from the classical probability distribution

$$\mathcal{N}\left[\mathbf{m}(\rho), \frac{V+\mathbb{1}}{2}\right](\mathbf{r}). \tag{3.246}$$

Using Lemma 144, we assume that the number of copies N satisfies

$$N \geq \left(2^8 + \frac{2^{13}}{n}\right)\frac{\|V\|_\infty^2}{\varepsilon^2}\left(2n + \log\left(\frac{2}{\delta}\right)\right) \geq \left(2^4 + \frac{2^{10}}{2n}\right)\frac{\left\|\frac{V+\mathbb{1}}{2}\right\|_\infty^2}{\left(\frac{\varepsilon}{4}\right)^2}\left(2n + \log\left(\frac{2}{\delta}\right)\right), \tag{3.247}$$

which allows us to construct estimators $\hat{\Sigma}$ and $\tilde{\mathbf{m}}$ such that, with probability at least $1-\delta$, the following bounds hold:

$$\|\hat{\Sigma} - \frac{V+\mathbb{1}}{2}\|_\infty \leq \frac{\varepsilon}{4}, \tag{3.248}$$

$$\|\tilde{\mathbf{m}} - \mathbf{m}\|_2 \leq \frac{1}{2\left\|\frac{V+\mathbb{1}}{2}\right\|_\infty^{1/2}} \cdot \frac{\varepsilon}{4} \leq \frac{\varepsilon}{4\|V\|_\infty^{1/2}}. \tag{3.249}$$

Consequently, we define the estimator for the covariance matrix as

$$\hat{V} = 2\hat{\Sigma} - \mathbb{1}. \tag{3.250}$$

From this, we deduce that

$$\|\hat{V} - V\|_\infty \leq \frac{\varepsilon}{2}. \tag{3.251}$$

However, $\hat{V}$ may not satisfy the uncertainty relation, and thus may not be a proper covariance matrix. To ensure that $\tilde{V}$ is a proper covariance matrix, we define

$$\tilde{V} = \hat{V} + \frac{\varepsilon}{2}\mathbb{1}, \tag{3.252}$$

so that by the triangle inequality, we have

$$\|\tilde{V} - V\|_\infty \leq \varepsilon. \tag{3.253}$$

Next, we show that $\tilde{V}$ satisfies the uncertainty relation $\tilde{V} + i\Omega_n \geq 0$. We proceed as follows:

$$\tilde{V} + i\Omega_n = \hat{V} + i\Omega_n + \frac{\varepsilon}{2}\mathbb{1}, \tag{3.254}$$

$$\geq \hat{V} - V + \frac{\varepsilon}{2}\mathbb{1}, \tag{3.255}$$

$$\geq -\|\hat{V} - V\|_\infty \mathbb{1} + \frac{\varepsilon}{2}\mathbb{1}, \tag{3.256}$$

$$\geq 0. \tag{3.257}$$

In the first inequality, we used the fact that $V + i\Omega_n \geq 0$. In the second inequality, we applied the result that for any operator Θ, we have $\Theta \geq -\|\Theta\|_\infty \mathbb{1}$. Therefore, $\tilde{V}$ satisfies the uncertainty relation and is a proper covariance matrix, completing the proof. □

Remark 146. *Note that the sample complexity bound in the previous theorem requires knowledge of an upper bound on the covariance matrix $\|V\|_\infty$ of the unknown Gaussian state ρ in order to predict the number of measurements needed to accurately estimate the first moment and covariance matrix of the unknown state. However, if we assume that the unknown state ρ satisfies an energy constraint, i.e.,*

$$\mathrm{Tr}[\rho\hat{E}] \leq E, \quad \textit{where} \quad \hat{E} \coloneqq \frac{1}{2}\hat{\mathbf{R}}^\intercal\hat{\mathbf{R}},$$

denotes the energy operator [2], then we have the bound

$$\|V\|_\infty \leq 4E.$$

This follows because

$$\|V\|_\infty \leq \mathrm{Tr}\, V = 4\,\mathrm{Tr}[\hat{E}\rho] - 2\|\mathbf{m}\|_2^2 \leq 4E. \tag{3.258}$$

By leveraging the trace-distance bounds introduced earlier, we can derive the following result:

Theorem 147 (Upper bound on the sample complexity of tomography of Gaussian states)**.** *Let $\varepsilon, \delta \in (0,1)$ denote the desired accuracy and failure probability, respectively. Let ρ be an unknown n-mode Gaussian state with an unknown covariance matrix V and first moment $\mathbf{m}$. A number*

$$N \geq \left(2^9 + \frac{2^{14}}{n}\right)\frac{\|V\|_1^2\|V\|_\infty^2}{\varepsilon^2}\left(2n + \log\left(\frac{2}{\delta}\right)\right) \tag{3.259}$$

of copies of ρ suffices to construct a classical description of a Gaussian state $\tilde{\rho}$ such that

$$\Pr\left(\frac{1}{2}\|\tilde{\rho} - \rho\|_1 \leq \varepsilon\right) \geq 1 - \delta. \tag{3.260}$$

Proof. Using Theorem 145, a number of copies

$$N \geq \left(2^8 + \frac{2^{13}}{n}\right)\frac{\|V\|_\infty^2}{\varepsilon'^2}\left(2n + \log\left(\frac{2}{\delta}\right)\right) \tag{3.261}$$

suffices to construct estimates $\tilde{\mathbf{m}}$ and $\tilde{V}$, with the following guarantees (with probability at least $1-\delta$):

$$\begin{aligned} \|\tilde{V} - V\|_\infty &\leq \varepsilon', \\ \|\tilde{\mathbf{m}} - \mathbf{m}\|_2 &\leq \frac{\varepsilon'}{4\|V\|_\infty^{1/2}}. \end{aligned} \tag{3.262}$$

Define $\tilde{\rho}$ as the Gaussian state with first moment $\tilde{\mathbf{m}}$ and covariance matrix $\tilde{V}$. Substituting these bounds into Eq. (3.114), we obtain:

$$\begin{aligned} \frac{1}{2}\|\rho - \tilde{\rho}\|_1 &\leq \frac{1+\sqrt{3}}{8}\max(\|V\|_1, \|\tilde{V}\|_1)\|V - \tilde{V}\|_\infty + \sqrt{\frac{\min(\|V\|_\infty, \|\tilde{V}\|_\infty)}{2}}\|\mathbf{m} - \tilde{\mathbf{m}}\|_2 \\ &\leq \frac{\varepsilon'}{2}\max(\|V\|_1, \|\tilde{V}\|_1) + \frac{\varepsilon'}{4\sqrt{2}} \\ &\leq \frac{\varepsilon'}{2}\|V\|_1 + \frac{\varepsilon'^2}{2} + \frac{\varepsilon'}{4\sqrt{2}}. \end{aligned} \tag{3.263}$$

Setting $\varepsilon' = \frac{\varepsilon}{\sqrt{2}\|V\|_1}$ ensures that $\frac{1}{2}\|\rho - \tilde{\rho}\|_1 \leq \varepsilon$ holds with probability at least $1-\delta$. □

To address the energy-constrained case $\mathrm{Tr}[\rho\hat{E}] \leq E$, note the following bound (as in Eq.(3.258):

$$\|V\|_\infty \leq \|V\|_1 \leq 4E. \tag{3.264}$$

This implies the following corollary:

Corollary 148 (Upper bound on the sample complexity for energy-constrained states)**.** *Let $\varepsilon, \delta \in (0,1)$ denote the desired accuracy and failure probability, respectively. Let ρ be an unknown n-mode Gaussian state satisfying the energy constraint $\mathrm{Tr}[\rho \hat{E}] \leq E$. A number*

$$N \geq \left(2^{15} + \frac{2^{22}}{n}\right) \frac{E^4}{\varepsilon^2} \left(2n + \log\left(\frac{2}{\delta}\right)\right) \tag{3.265}$$

of copies of ρ suffices to construct a classical description of a Gaussian state $\tilde{\rho}$ such that

$$\Pr\left(\frac{1}{2}\|\tilde{\rho} - \rho\|_1 \leq \varepsilon\right) \geq 1 - \delta. \tag{3.266}$$

The constant factors in our sample-complexity bounds are likely overestimated due to the conservative approximations employed in the proofs to ensure simplicity and generality across all parameter regimes. In practice, one can anticipate considerably better constants.

Bibliography

[1] A. S. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction*. Texts and Monographs in Theoretical Physics. De Gruyter, Berlin, Germany, 2nd edition, 2019. ISBN 9783110642407. doi: https://doi.org/10.1515/9783110273403.

[2] A. Serafini. *Quantum continuous variables: A primer of theoretical methods*. CRC Press, Taylor & Francis Group, Boca Raton, USA, 2017. doi: doi.org/10.1201/9781315118727.

[3] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd. Gaussian quantum information. *Rev. Mod. Phys.*, 84(2):621–669, May 2012. doi: 10.1103/RevModPhys.84.621.

[4] N. J. Cerf, G. Leuchs, and E. S. Polzik. *Quantum information with continuous variables of atoms and light*. Imperial College Press, London, UK, 2007. URL http://emu.tind.io/record/298574.

[5] Gerardo Adesso, Sammy Ragy, and Antony R. Lee. Continuous variable quantum information: Gaussian states and beyond. *Open Systems amp; Information Dynamics*, 21(01n02):1440001, March 2014. ISSN 1793-7191. doi: 10.1142/s1230161214400010. URL http://dx.doi.org/10.1142/S1230161214400010.

[6] S. Lloyd and S. L. Braunstein. Quantum computation over continuous variables. *Phys. Rev. Lett.*, 82:1784–1787, 1999. doi: 10.1103/PhysRevLett.82.1784.

[7] D. Gottesman, A. Kitaev, and J. Preskill. Encoding a qubit in an oscillator. *Phys. Rev. A*, 64:012310, 2001. doi: 10.1103/PhysRevA.64.012310.

[8] N. C. Menicucci, P. van Loock, M. Gu, C. Weedbrook, T. C. Ralph, and M. A. Nielsen. Universal quantum computation with continuous-variable cluster states. *Phys. Rev. Lett.*, 97: 110501, 2006. doi: 10.1103/PhysRevLett.97.110501.

[9] Mazyar Mirrahimi, Zaki Leghtas, Victor V Albert, Steven Touzard, Robert J Schoelkopf, Liang Jiang, and Michel H Devoret. Dynamically protected cat-qubits: a new paradigm for universal quantum computation. *New J. Phys.*, 16(4):045014, April 2014. ISSN 1367-2630. doi: 10.1088/1367-2630/16/4/045014.

[10] Nissim Ofek, Andrei Petrenko, Reinier Heeres, Philip Reinhold, Zaki Leghtas, Brian Vlastakis, Yehan Liu, Luigi Frunzio, Steven Girvin, Liang Jiang, Mazyar Mirrahimi, M. Devoret, and R. Schoelkopf. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature*, 536, 07 2016. doi: 10.1038/nature18949.

[11] Marios H. Michael, Matti Silveri, R. T. Brierley, Victor V. Albert, Juha Salmilehto, Liang Jiang, and S. M. Girvin. New class of quantum error-correcting codes for a bosonic mode. *Phys. Rev. X*, 6:031006, Jul 2016. doi: 10.1103/PhysRevX.6.031006.

[12] Jérémie Guillaud and Mazyar Mirrahimi. Repetition cat qubits for fault-tolerant quantum computation. *Phys. Rev. X*, 9(4):041053, December 2019. ISSN 2160-3308. doi: 10.1103/physrevx.9.041053.

[13] Koen Alexander, Andrea Bahgat, Avishai Benyamini, Dylan Black, Damien Bonneau, Stanley Burgos, Ben Burridge, Geoff Campbell, Gabriel Catalano, Alex Ceballos, Chia-Ming Chang, CJ Chung, Fariba Danesh, Tom Dauer, Michael Davis, Eric Dudley, Ping Er-Xuan, Josep Fargas, Alessandro Farsi, Colleen Fenrich, Jonathan Frazer, Masaya Fukami, Yogeeswaran Ganesan, Gary Gibson, Mercedes Gimeno-Segovia, Sebastian Goeldi, Patrick Goley, Ryan Haislmaier, Sami Halimi, Paul Hansen, Sam Hardy, Jason Horng, Matthew House, Hong Hu, Mehdi Jadidi, Henrik Johansson, Thomas Jones, Vimal Kamineni, Nicholas Kelez, Ravi Koustuban, George Kovall, Peter Krogen, Nikhil Kumar, Yong Liang, Nicholas LiCausi, Dan Llewellyn, Kimberly Lokovic, Michael Lovelady, Vitor Manfrinato, Ann Melnichuk, Mario Souza, Gabriel Mendoza, Brad Moores, Shaunak Mukherjee, Joseph Munns, Francois-Xavier Musalem, Faraz Najafi, Jeremy L. O'Brien, J. Elliott Ortmann, Sunil Pai, Bryan Park, Hsuan-Tung Peng, Nicholas Penthorn, Brennan Peterson, Matt Poush, Geoff J. Pryde, Tarun Ramprasad, Gareth Ray, Angelita Rodriguez, Brian Roxworthy, Terry Rudolph, Dylan J. Saunders, Pete Shadbolt, Deesha Shah, Hyungki Shin, Jake Smith, Ben Sohn, Young-Ik Sohn, Gyeongho Son, Chris Sparrow, Matteo Staffaroni, Camille Stavrakas, Vijay Sukumaran, Davide Tamborini, Mark G. Thompson, Khanh Tran, Mark Triplet, Maryann Tung, Alexey Vert, Mihai D. Vidrighin, Ilya Vorobeichik, Peter Weigel, Mathhew Wingert, Jamie Wooding, and Xinran Zhou. A manufacturable platform for photonic quantum computing, 2024.

[14] Lukas Brenner, Libor Caha, Xavier Coiteux-Roy, and Robert Koenig. Factoring an integer with three oscillators and a qubit, 2024. URL https://arxiv.org/abs/2412.13164.

[15] M. M. Wolf, D. Pérez-García, and G. Giedke. Quantum capacities of bosonic channels. *Phys. Rev. Lett.*, 98:130501, 2007. doi: 10.1103/PhysRevLett.98.130501.

[16] M. Takeoka, S. Guha, and Mark M. Wilde. Fundamental rate-loss tradeoff for optical quantum key distribution. *Nature Comm.*, 5(1):5235, 2014. doi: 10.1038/ncomms6235.

[17] Stefanno Pirandola, Riccardo Laurenza, Carlo Ottaviani, and Leonardo Banchi. Fundamental limits of repeaterless quantum communications. *Nature Comm.*, 8(1):15043, 2017. doi: 10.1038/ncomms15043.

[18] F. A. Mele, Ludovico Lami, and V. Giovannetti. Restoring quantum communication efficiency over high loss optical fibers. *Phys. Rev. Lett.*, 129:180501, 2022. doi: 10.1103/PhysRevLett.129.180501.

[19] Francesco Anna Mele, Ludovico Lami, and Vittorio Giovannetti. Maximum tolerable excess noise in CV-QKD and improved lower bound on two-way capacities, 2023.

[20] Francesco Anna Mele, Farzin Salek, Vittorio Giovannetti, and Ludovico Lami. Quantum communication on the bosonic loss-dephasing channel, 2024.

[21] F. A. Mele, Ludovico Lami, and V. Giovannetti. Quantum optical communication in the presence of strong attenuation noise. *Phys. Rev. A*, 106:042437, 2022. doi: 10.1103/PhysRevA.106.042437.

[22] F. A. Mele, G. De Palma, M. Fanizza, V. Giovannetti, and L. Lami. Optical fibres with memory effects and their quantum communication capacities, 2023.

[23] The LIGO Scientific Collaboration. Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light. *Nature Phot.*, 7(8):613–619, 2013. doi: 10.1038/nphoton.2013.177.

[24] Junhua Zhang, Mark Um, Dingshun Lv, Jing-Ning Zhang, Lu-Ming Duan, and Kihwan Kim. Noon states of nine quantized vibrations in two radial modes of a trapped ion. *Phys. Rev. Lett.*, 121:160502, Oct 2018. doi: 10.1103/PhysRevLett.121.160502.

[25] V. Meyer, M. A. Rowe, D. Kielpinski, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. Experimental demonstration of entanglement-enhanced rotation angle estimation using trapped ions. *Phys. Rev. Lett.*, 86:5870–5873, Jun 2001. doi: 10.1103/PhysRevLett.86.5870.

[26] K. McCormick, J. Keller, S. Burd, D. Wineland, A. Wilson, and D. Leibfried. Quantum-enhanced sensing of a single-ion mechanical oscillator. *Nature*, 572:1, 08 2019. doi: 10.1038/s41586-019-1421-y.

[27] Matthias Ohliger, Vincent Nesme, David Gross, Yi-Kai Liu, and Jens Eisert. Continuous variable quantum compressed sensing, 2011.

[28] Lars S. Madsen, Fabian Laudenbach, Mohsen Falamarzi. Askarani, Fabien Rortais, Trevor Vincent, Jacob F. F. Bulmer, Filippo M. Miatto, Leonhard Neuhaus, Lukas G. Helt, Matthew J. Collins, Adriana E. Lita, Thomas Gerrits, Sae Woo Nam, Varun D. Vaidya, Matteo Menotti, Ish Dhand, Zachary Vernon, Nicolás Quesada, and Jonathan Lavoie. Quantum computational advantage with a programmable photonic processor. *Nature*, 606(7912):75–81, 2022. doi: 10.1038/s41586-022-04725-x.

[29] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, Peng Hu, Xiao-Yan Yang, Wei-Jun Zhang, Hao Li, Yuxuan Li, Xiao Jiang, Lin Gan, Guangwen Yang, Lixing You, Zhen Wang, Li Li, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, December 2020. ISSN 1095-9203. doi: 10.1126/science.abe8770.

[30] D. Hangleiter and J. Eisert. Computational advantage of quantum random sampling. *Rev. Mod. Phys.*, 95:035001, 2023. doi: 10.1103/RevModPhys.95.035001.

[31] F. H. B. Somhorst, R. van der Meer, M. Correa Anguita, R. Schadow, H. J. Snijders, M. de Goede, B. Kassenberg, P. Venderbosch, C. Taballione, J. P. Epping, H. H. van den Vlekkert, J. F. F. Bulmer, J. Lugani, I. A. Walmsley, P. W. H. Pinkse, J. Eisert, N. Walk, and J. J. Renema. Quantum photo-thermodynamics on a programmable photonic quantum processor. *Nature Comm.*, 14:3895, 2023. doi: 10.1038/s41467-023-38413-9.

[32] C. W. Helstrom. *Quantum detection and estimation theory*. Academic press, New York, USA, 1976. doi: https://doi.org/10.1007/BF01007479.

[33] A. S. Holevo. Investigations in the general theory of statistical decisions. *Trudy Mat. Inst. Steklov*, 124:3–140, 1976. doi: 10.1070/RM9856. (English translation: Proc. Steklov Inst. Math. 124:1–140, 1978).

[34] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010. doi: https://doi.org/10.1017/CBO9780511976667.

[35] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2nd edition, 2017. ISBN 9781316813300. doi: https://doi.org/10.1017/CBO9781139525343.

[36] Sumeet Khatri and Mark M. Wilde. Principles of quantum communication theory: A modern approach, 2020.

[37] J. Watrous. *The theory of quantum information*. Cambridge University Press, 2018. doi: 10.1017/9781316848142.

[38] A. S. Holevo. On estimates of trace-norm distance between quantum Gaussian states, 2024. URL https://arxiv.org/abs/2408.11400.

[39] A. S. Holevo. On estimates of the bures distance between bosonic gaussian states, 2024. URL https://arxiv.org/abs/2412.04875.

[40] F. A. Mele, A. A. Mele, L. Bittel, J. Eisert, V. Giovannetti, L. Lami, L. Leone, and S. F. E. Oliviero. Learning quantum states of continuous variable systems, 2024. URL https://arxiv.org/abs/2405.01431.

[41] Lennart Bittel, Francesco Anna Mele, Antonio Anna Mele, Salvatore Tirone, and Ludovico Lami. Optimal estimates of trace distance between bosonic gaussian states and applications to learning, 2024. URL https://arxiv.org/abs/2411.02368.

[42] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, page 363–390. IEEE, November 2023. doi: 10.1109/focs57990.2023.00028. URL http://dx.doi.org/10.1109/FOCS57990.2023.00028.

[43] L. Banchi, S.L. Braunstein, and S. Pirandola. Quantum fidelity for arbitrary Gaussian states. *Phys. Rev. Lett.*, 115(26):260501, 2015. doi: 10.1103/physrevlett.115.260501.

[44] A. Anshu and S. Arunachalam. A survey on the complexity of learning quantum states. *Nat. Rev. Phys.*, 6(1):59–69, Dec 2023. ISSN 2522-5820. doi: 10.1038/s42254-023-00662-4.

[45] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nat Commun*, 6(1):8498, 2015. ISSN 2041-1723. doi: 10.1038/ncomms9498.

[46] Srilekha Gandhari, Victor V. Albert, Thomas Gerrits, Jacob M. Taylor, and Michael J. Gullans. Precision bounds on continuous-variable state tomography using classical shadows, 2023.

[47] S. Becker, N. Datta, L. Lami, and C. Rouzé. Classical shadow tomography for continuous variables quantum systems. *IEEE Trans. Inf. Theory*, 70(5):3427–3452, 2024. doi: 10.1109/TIT.2024.3357972.

[48] C. Oh, S. Chen, Y. Wong, S. Zhou, H.-Y. Huang, J. A. H. Nielsen, Z.-H. Liu, J. S. Neergaard-Nielsen, U. L. A., L. Jiang, and J. Preskill. Entanglement-enabled advantage for learning a bosonic random displacement channel, 2024.

[49] Omar Fawzi, Aadil Oufkir, and Robert Salzmann. Optimal fidelity estimation from binary measurements for discrete and continuous variable systems, 2024. URL https://arxiv.org/abs/2409.04189.

[50] Tim Möbus, Andreas Bluhm, Matthias C. Caro, Albert H. Werner, and Cambyse Rouzé. Dissipation-enabled bosonic hamiltonian learning via new information-propagation bounds, 2023. URL https://arxiv.org/abs/2307.15026.

[51] Varun Upreti and Ulysse Chabaud. An efficient quantum state verification framework and its application to bosonic systems, 2024. URL https://arxiv.org/abs/2411.04688.

[52] Marco Fanizza, Cambyse Rouzé, and Daniel Stilck França. Efficient hamiltonian, structure and trace distance learning of gaussian states, 2024. URL https://arxiv.org/abs/2411.03163.

[53] Rosanna Nichols, Pietro Liuzzo-Scorpo, Paul A. Knott, and Gerardo Adesso. Multiparameter gaussian quantum metrology. *Phys. Rev. A*, 98:012114, Jul 2018. doi: 10.1103/PhysRevA.98.012114. URL https://link.aps.org/doi/10.1103/PhysRevA.98.012114.

[54] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, Jan 2006. doi: 10.1103/PhysRevLett.96.010401. URL https://link.aps.org/doi/10.1103/PhysRevLett.96.010401.

[55] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nature Photonics*, 5(4):222–229, March 2011. ISSN 1749-4893. doi: 10.1038/nphoton.2011.35. URL http://dx.doi.org/10.1038/nphoton.2011.35.

[56] Alex Monras. Phase space formalism for quantum estimation of gaussian states, 2013. URL https://arxiv.org/abs/1303.3682.

[57] Yang Gao and Hwang Lee. Bounds on quantum multiple-parameter estimation with gaussian state. *The European Physical Journal D*, 68(11), November 2014. ISSN 1434-6079. doi: 10.1140/epjd/e2014-50560-1. URL http://dx.doi.org/10.1140/epjd/e2014-50560-1.

[58] Alex Monras and Fabrizio Illuminati. Information geometry of gaussian channels. *Physical Review A*, 81(6), June 2010. ISSN 1094-1622. doi: 10.1103/physreva.81.062326. URL http://dx.doi.org/10.1103/PhysRevA.81.062326.

[59] L. Bittel, A. A. Mele, J. Eisert, and L. Leone. Optimal trace-distance bounds for free-fermionic states: Testing and improved tomography. *arXiv*, 2024. URL https://arxiv.org/abs/2409.17953.

[60] Zixin Huang and Mark M. Wilde. Information geometry of bosonic gaussian thermal states, 2024. URL https://arxiv.org/abs/2411.18268.

[61] See the Supplemental Material for details on the derivation of the derivative of a Gaussian state, the proof of the main results, and the applications of our results to quantum state tomography.

[62] S. S. Barsov and V. V. Ul'yanov. Estimates of the proximity of gaussian measures. *Soviet Mathematics, Doklady*, 34(3):462–466, 1987. Available at https://www.researchgate.net/publication/268861237_Estimates_of_the_proximity_of_Gaussian_measures.

[63] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional Gaussians with the same mean, 2023.

[64] Jamil Arbas, Hassan Ashtiani, and Christopher Liaw. Polynomial time and private learning of unbounded Gaussian mixture models, 2023.

[65] R. O'Donnell and J. Wright. Efficient quantum tomography, 2015.

[66] J. Haah, A.W. Harrow, J. Zhengfeng, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Th.*, 2017. ISSN 1557-9654. doi: 10.1109/tit.2017.2719044.

[67] R. Kueng, H. Rauhut, and U. Terstiege. Low rank matrix recovery from rank one measurements, 2014.

[68] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning?, 2023.

[69] M. Guta, J. Kahn, R. Kueng, and J.A. Tropp. Fast state tomography with optimal error bounds, 2018.

[70] Marco Fanizza, Niklas Galke, Josep Lumbreras, Cambyse Rouzé, and Andreas Winter. Learning finitely correlated states: stability of the spectral reconstruction, 2023.

[71] H. Huang, Y. Liu, M. Broughton, I. Kim, A. Anshu, Z. Landau, and J. R. McClean. Learning shallow quantum circuits. In *Proc. 56th Annu. ACM Symp. Theory Comput.*, pages 1343–1351, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703836. doi: 10.1145/3618260.3649722.

[72] S. Arunachalam, S. Bravyi, A. Dutt, and T. J. Yoder. Optimal algorithms for learning quantum phase states, 2023.

[73] A. Montanaro. Learning stabilizer states by bell sampling, 2017.

[74] S. Grewal, V. Iyer, W. Kretschmer, and D. Liang. Efficient learning of quantum states prepared with few non-Clifford gates, 2023.

[75] L. Leone, S.F.E. Oliviero, and A. Hamma. Learning t-doped stabilizer states. *Quantum*, 8: 1361, May 2024. ISSN 2521-327X. doi: 10.22331/q-2024-05-27-1361. URL http://dx.doi.org/10.22331/q-2024-05-27-1361.

[76] D. Hangleiter and M.J. Gullans. Bell sampling from quantum circuits, 2024.

[77] S. Aaronson and S. Grewal. Efficient tomography of non-interacting fermion states, 2023.

[78] Antonio Anna Mele and Yaroslav Herasymenko. Efficient learning of quantum states prepared with few fermionic non-gaussian gates. *PRX Quantum*, 6(1), January 2025. ISSN 2691-3399. doi: 10.1103/prxquantum.6.010319. URL http://dx.doi.org/10.1103/PRXQuantum.6.010319.

[79] Zixin Huang and Mark M. Wilde. Information geometry of bosonic gaussian thermal states, 2024. URL https://arxiv.org/abs/2411.18268.

[80] Mark M. Wilde, Marco Tomamichel, Seth Lloyd, and Mario Berta. Gaussian hypothesis testing and quantum illumination. *Physical Review Letters*, 119(12), September 2017. ISSN 1079-7114. doi: 10.1103/physrevlett.119.120501. URL http://dx.doi.org/10.1103/PhysRevLett.119.120501.

[81] H. J. Groenewold. On the principles of elementary quantum mechanics. *Physica*, 12(7): 405–460, 1946. ISSN 0031-8914. doi: https://doi.org/10.1016/S0031-8914(46)80059-4.

[82] Ludovico Lami, B. Regula, X. Wang, R. Nichols, Andreas Winter, and Gerardo Adesso. Gaussian quantum resource theories. *Phys. Rev. A*, 98:022335, 2018. doi: 10.1103/PhysRevA.98.022335.

[83] M. Keyl, J. Kiukas, and R. F. Werner. Schwartz operators. *Rev. Math. Phys.*, 28(03):1630001, 2016. doi: 10.1142/S0129055X16300016.

[84] L. Narici and E. Beckenstein. *Topological Vector Spaces*. Chapman and Hall/CRC, Jul 2010. ISBN 9780429147890. doi: 10.1201/9781584888673.

[85] L. Hörmander. *The Analysis of Linear Partial Differential Operators I, (Distribution theory and Fourier Analysis)*. Springer-Verlag, 1990. ISBN 3-540-52343-X. doi: https://doi.org/10.1007/978-3-642-61497-2.

[86] E. Zeidler. *Applied Functional Analysis*. Springer, New York, 1995. ISBN 9781461208211. doi: 10.1007/978-1-4612-0821-1.

[87] R. Bhatia. *Positive definite matrices*. Princeton university press, 2009. doi: https://doi.org/10.1515/9781400827787.

[88] Rajendra Bhatia and Tanvi Jain. On symplectic eigenvalues of positive definite matrices. *Journal of Mathematical Physics*, 56(11), November 2015. ISSN 1089-7658. doi: 10.1063/1.4935852. URL http://dx.doi.org/10.1063/1.4935852.

[89] Gábor Lugosi and Shahar Mendelson. Sub-gaussian estimators of the mean of a random vector, 2017. URL https://arxiv.org/abs/1702.00482.

[90] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi: 10.1017/9781108231596.

[91] Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. doi: https://doi.org/10.1017/9781108627771.

[92] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices, 2011. URL https://arxiv.org/abs/1011.3027.

Chapter 4

Review on bosonic trace distance bounds

Let us give a brief literature review of the bosonic trace-distance bounds derived in our recent works [1–5]. These bounds constitute a useful tool for addressing problems in quantum learning theory with continuous-variable systems.

A central question in this line of research is how to bound the trace distance between two continuous-variable quantum states ρ and σ in terms of the distance between their first moments, $\|\mathbf{m}(\rho) - \mathbf{m}(\sigma)\|$, and their covariance matrices, $\|V(\rho) - V(\sigma)\|$. This question is of fundamental importance because, although the trace distance plays a key role in all the areas of quantum information science [6], it is generally hard to compute for continuous-variable systems. In contrast, first moments and covariance matrices are typically much easier to estimate, both analytically and experimentally. The answer to this question depends critically on the assumptions made about the states ρ and σ; various bounds have been derived under different assumptions regarding the Gaussianity or the purity of the states:

- **Upper bound on the trace distance between Gaussian states**: It is well known that Gaussian states are uniquely determined by their first moments and covariance matrices. This naturally raises the fundamental question: *If two Gaussian states have first moments and covariance matrices that are ε-close, how close are the states in trace distance?* This problem was first introduced in [1], where partial results were obtained. Later, Holevo [7] and Fanizza et al. [8] made further progress on this problem. Then, we fully solved this question in [3, 4], by establishing the following tight bound:

  Let ρ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Similarly, let σ be the Gaussian state with first moment $\mathbf{t}$ and covariance matrix W. Then, the following the trace norm between ρ and σ satisfies the following tight upper bounds [4]:

$$\frac{1}{2}\|\rho(V,\mathbf{m}) - \rho(W,\mathbf{t})\|_1 \leq \frac{1+\sqrt{3}}{8}\operatorname{Tr}\left[(V^{-1}+W^{-1})|V-W|\right] + \frac{1}{2}\left\|V^{-1/2}(\mathbf{m}-\mathbf{t})\right\|_2 . \tag{4.1}$$

- **Lower bound on the trace distance between Gaussian states**: In [1] we proved the following lower bound on the trace distance between Gaussian states:

  Let ρ be the Gaussian state with first moment $\mathbf{m}$ and covariance matrix V. Similarly, let σ be the Gaussian state with first moment $\mathbf{t}$ and covariance matrix W. Then, the trace distance between ρ and σ can be lower bounded as [1, Eq. S265]:

$$\begin{aligned} \|\rho-\sigma\|_1 &\geq \frac{1}{100}\min\left\{1, \frac{\|\mathbf{m}-\mathbf{t}\|}{\sqrt{4\min(\|V\|_\infty, \|W\|_\infty)+1}}\right\} , \\ \|\rho-\sigma\|_1 &\geq \frac{1}{100}\min\left\{1, \frac{\|V-W\|_2}{4\min(\|V\|_\infty, \|W\|_\infty)+1}\right\} . \end{aligned} \tag{4.2}$$

- **Upper bound on the trace distance between a non-Gaussian state and a pure Gaussian state**: In [1] we established the following upper bound on the trace distance between a non-Gaussian state and a pure Gaussian state:

  Let ψ be a pure Gaussian state and let ρ be a (possibly non-Gaussian) state. Then, the trace distance between ψ and ρ can be upper bounded as [1]

  $$\frac{1}{2}\|\rho-\psi\|_1 \leq \sqrt{\max\Big(\mathrm{Tr}\Big[\hat{E}\rho\Big],\mathrm{Tr}\Big[\hat{E}\sigma\Big]\Big)}\sqrt{2\|\boldsymbol{m}(\rho)-\boldsymbol{m}(\psi)\|^2+\|V(\rho)-V(\psi)\|_\infty}\,, \tag{4.3}$$

  where $\hat{E}$ denotes the energy operator.

- **Lower bound on the trace distance between non-Gaussian states**: In [2] we establish lower bounds on the trace distance between two possibly non-Gaussian states ρ and σ:

  $$\begin{aligned}\|\rho-\sigma\|_1 &\geq \frac{\|\mathbf{m}(\sigma)-\mathbf{m}(\rho)\|^2}{16\max\Big(\mathrm{Tr}\Big[\hat{E}\rho\Big],\mathrm{Tr}\Big[\hat{E}\sigma\Big]\Big)}\,,\\ \|\rho-\sigma\|_1 &\geq \frac{\|V(\rho)-V(\sigma)\|_\infty^2}{1549\,\max(\mathrm{Tr}[\hat{E}^2\rho],\mathrm{Tr}[\hat{E}^2\sigma])}\,.\end{aligned} \tag{4.4}$$

  It is worth noting that this trace distance lower bound scales quadratically with the norm distance between the covariance matrices, in contrast to the lower bound for Gaussian states in 4.2, which exhibits linear scaling. Nevertheless, in [2], we show that this quadratic behaviour is tight.

Another related result is presented in [5], where an **algorithm to compute the trace distance between two Gaussian states up to a fixed precision** has been developed. The construction of this algorithm relies on two key steps:

- First, Ref. [5] derives a simple bound on the probability $P_{>M}$ that a given Gaussian state has more than M photons, establishing that such a probability decays exponentially in M. The result is as follows [5]:

  Let ρ be an n-mode Gaussian state, and let $P_{>M}$ denote the probability that ρ has more than M photons. According to Born's rule, this probability is defined by

  $$P_{>M} \coloneqq \mathrm{Tr}[(\mathbb{1}-\Pi_M)\rho]\,, \tag{4.5}$$

  where Π_M is the projector onto the subspace spanned by all n-mode Fock states with total photon number less than or equal to M:

  $$\Pi_M \coloneqq \sum_{\mathbf{k}\in\mathbb{N}^n:\sum_{i=1}^n k_i\leq M} |\mathbf{k}\rangle\langle\mathbf{k}|\,, \tag{4.6}$$

  with $|\boldsymbol{k}\rangle = |k_1\rangle\otimes|k_2\rangle\otimes\ldots\otimes|k_n\rangle$ denoting an n-mode Fock state. Then, the following bound holds:

  $$P_{>M} \leq 2^n e^{1/2}\, e^{-M/(4\,\mathrm{Tr}[\rho\hat{N}]+2)}\,, \tag{4.7}$$

  where $\hat{N}$ is the photon number operator.

- Second, as an application of this bound, Ref. [5] establishes an upper bound on the trace distance between the Gaussian state ρ and its finite-dimensional truncation ρ_M, defined as

  $$\rho_M \coloneqq \frac{\Pi_M\rho\Pi_M}{\mathrm{Tr}[\Pi_M\rho\Pi_M]}\,, \tag{4.8}$$

  where Π_M is the projector given in Eq. (4.6). The trace distance between ρ and ρ_M is shown to vanish exponentially in M as follows:

  $$\frac{1}{2}\,\|\rho-\rho_M\|_1 \leq 2^{n/2}e^{1/4}\, e^{-M/\left(8\,\mathrm{Tr}[\rho\hat{N}]+4\right)}\,. \tag{4.9}$$

These two results play a key role in constructing the aforementioned algorithm that computes the trace distance between two Gaussian states up to a fixed precision [5].

Bibliography

[1] Francesco A. Mele, Antonio A. Mele, Lennart Bittel, Jens Eisert, Vittorio Giovannetti, Ludovico Lami, Lorenzo Leone, and Salvatore F. E. Oliviero. Learning quantum states of continuous-variable systems. *Nature Physics*, 21(12):2002–2008, November 2025. ISSN 1745-2481. doi: 10.1038/s41567-025-03086-2. URL http://dx.doi.org/10.1038/s41567-025-03086-2.

[2] Francesco Anna Mele, Salvatore Francesco Emanuele Oliviero, Varun Upreti, and Ulysse Chabaud. The symplectic rank of non-gaussian quantum states. *arXiv preprint arXiv:2504.19319*, 2025. URL https://arxiv.org/abs/2504.19319.

[3] Lennart Bittel, Francesco Anna Mele, Antonio Anna Mele, Salvatore Tirone, and Ludovico Lami. Optimal estimates of trace distance between bosonic gaussian states and applications to learning, 2024. URL https://arxiv.org/abs/2411.02368.

[4] Lennart Bittel, Francesco A. Mele, Jens Eisert, and Antonio A. Mele. Energy-independent tomography of gaussian states, 2025. URL https://arxiv.org/abs/2508.14979.

[5] Francesco Anna Mele, Giovanni Barbarino, Vittorio Giovannetti, and Marco Fanizza. Achievable rates in non-asymptotic bosonic quantum communication. 2025. URL https://arxiv.org/abs/2502.05524.

[6] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. Publications of the Scuola Normale Superiore. Scuola Normale Superiore, Pisa, Italy, 2011.

[7] A. S. Holevo. On estimates of trace-norm distance between quantum Gaussian states. *arXiv*, 2024. URL https://arxiv.org/abs/2408.11400.

[8] Marco Fanizza, Cambyse Rouze, and Daniel Stilck França. Efficient Hamiltonian, structure and trace distance learning of Gaussian states. *arXiv*, 2024. URL https://arxiv.org/abs/2411.03163.

Part II

Quantum Shannon Theory with continuous variable systems

Chapter 5

Maximum tolerable excess noise in continuous-variable quantum key distribution and improved lower bound on two-way capacities

In this second part of the thesis, we turn our attention from quantum learning theory to *quantum Shannon theory* with continuous-variable (CV) systems, where the goal is to quantify the ultimate rates at which entanglement and secret keys can be distributed through noisy bosonic channels. A central figure of merit in this setting is provided by the *two-way capacities* of a CV quantum channel, which capture the optimal rate at which entanglement and secret keys can be generated by two distant parties connected by a lossy and noisy transmission line when unlimited two-way classical communication is allowed, but *no quantum repeaters* are available. Since repeater-based architectures are technologically demanding, understanding what can be achieved in their absence is both practically relevant and conceptually central.

In this chapter, we address this question for the most important class of CV quantum channels, namely phase-insensitive single-mode bosonic Gaussian channels (thermal attenuators, thermal amplifiers, and additive Gaussian-noise channels), which accurately describe the dominant noise mechanisms in optical fibres and free-space links. We determine the maximum excess noise compatible with continuous-variable QKD (CV-QKD), thereby providing a sharp benchmark that separates a regime where point-to-point CV-QKD is impossible—and repeaters become necessary—from a regime where end-to-end QKD can be achieved without quantum repeaters. Below this threshold, we present an explicit entanglement distribution and distillation protocol, obtained by optimising a concrete construction that combines recurrence and hashing procedures. The achievable rates of this protocol yield state-of-the-art lower bounds on the energy-constrained and unconstrained two-way quantum capacity and secret-key capacity in the high-noise regime, including the parameter range where coherent-information-based bounds become negative. In particular, our results give the first nonzero lower bound on the two-way quantum capacity in this regime and make explicit that entanglement distribution remains possible whenever the channel is not entanglement breaking.

[This chapter is based on F. A. Mele, L. Lami, V. Giovannetti, Maximum tolerable excess noise in continuous-variable quantum key distribution and improved lower bound on two-way capacities, Nature Photonics 19, 329–334 (2025)]

5.1 Introduction

Quantum key distribution (QKD) stands as the gold standard for unconditionally secure communication. Since its inception in 1984 [1], QKD has transitioned from a theoretical concept

to a commercially viable technology. Continuous-variable systems [2], such as electromagnetic modes, offer a powerful approach to QKD, known as CV-QKD [3–5]. Unlike its discrete-variable counterpart [4], CV-QKD is expected to be seamlessly integrated into existing optical fibre-based Internet infrastructure in the near future [4, 6–10], making it highly practical for real-world applications. A significant challenge in this field, essential for the development of a large-scale quantum internet [4, 11], is extending CV-QKD over long distances without intermediate nodes [4]. In recent years, numerous experiments have been setting new distance records, achieving CV-QKD across optical fibres exceeding one hundred kilometers in length [6–10]. This progress prompts a critical question: "What is the maximum achievable distance of CV-QKD according to the laws of quantum physics?" Equivalently, "What is the maximum tolerable excess noise in CV-QKD?" This has been identified as a crucial open problem in the field [12, Section 7].

In quantum Shannon theory [13, 14], the capacities of quantum channels determine the ultimate limits of quantum communication which are achievable across the channel. Specifically, the *secret-key capacity* $K(\Phi)$ and the *two-way quantum capacity* $Q_2(\Phi)$ of a quantum channel Φ, collectively called the *two-way capacities*, are the maximum rates of secret-key bits and entanglement bits (or *ebits*), respectively, achievable across Φ when the sender (Alice) and the receiver (Bob) are assisted by two-way classical communication [13, 14]. These capacities quantify the optimal performance for QKD and entanglement distribution. Importantly, QKD across Φ is achievable if and only if $K(\Phi) > 0$, while entanglement distribution across Φ is achievable if and only if $Q_2(\Phi) > 0$. Since in practice only a finite amount of energy can be utilised in a communication protocol, it is common to consider the *energy-constrained* two-way capacities [15]. These are denoted as $K(\Phi, N_s)$ and $Q_2(\Phi, N_s)$, where N_s represents the maximum average photon number per input signal to the channel Φ. Since an ebit can be converted into a secret-key bit [16], these capacities satisfy $K(\Phi, N_s) \geq Q_2(\Phi, N_s)$.

Within the framework of quantum optical communication, a realistic model to describe optical fibres is the *thermal attenuator* $\mathscr{E}_{\lambda,\nu}$ [2, 12, 17]. It is a single-mode quantum channel characterised by two parameters: the transmissivity $\lambda \in [0, 1]$ of the fibre and the added thermal noise $\nu \in [0, \infty)$ [2]. We provide a detailed definition of the thermal attenuator in the Methods. Computing the two-way capacities of the thermal attenuator is essential to determine the ultimate performances of QKD or entanglement distribution protocols that can be achieved on optical networks without the use of intermediate nodes or quantum repeaters [4, 18, 19]. The two-way capacities of the thermal attenuator have been determined for all λ only when $\nu = 0$ and there is no energy constraint [17], assumptions that may not be entirely physically realistic, depending on the setting. Except for this very special case, the two-way capacities of the thermal attenuator are still unknown, despite the many upper [15, 17, 20–24] and lower bounds [12, 22, 25–27] that have been established. In particular, in a large parameter region all lower bounds (prior to our work) vanish, while the upper bounds do not. This entails that the precise noise threshold above which entanglement distribution or key distribution become impossible had not been determined in the prior literature. This leads to the mathematical formulation of the aforementioned open problem [12, Section 7]: *for a given ν, determine the minimum transmissivity λ for which the secret-key capacity of the thermal attenuator is strictly positive, i.e. for which CV-QKD is achievable.*

5.2 Results

5.2.1 Maximum tolerable excess noise in CV-QKD

In this section, we solve the above problem [12, Section 7], establishing the following simple formula for the minimum admissible transmissivity $\lambda_{\min}$ in CV-QKD:

$$\lambda_{\min} = \frac{\nu}{\nu + 1}\,. \tag{5.1}$$

Note that the transmissivity λ of an optical fibre decreases exponentially with its length L as [28, 29]

$$\lambda = 10^{-\gamma \frac{L}{10\,\mathrm{km}}}\,, \tag{5.2}$$

where the attenuation coefficient γ typically satisfies $\gamma \simeq 0.2$, with the lowest recorded value being $\gamma \simeq 0.14$ [28, 29]. We can thus determine the maximum achievable distance $L_{\max}$ in CV-QKD:

$$L_{\max} = \frac{10\,\mathrm{km}}{\gamma} \log_{10}\left(1+\nu^{-1}\right)\,. \tag{5.3}$$

The parameter ν is related to the wavelength λ_{wave} employed for communication through the Bose–Einstein distribution:

$$\nu = \frac{1}{\exp\left(\frac{hc}{\lambda_{\mathrm{wave}} k_B T}\right) - 1}\,, \tag{5.4}$$

where T is the room temperature, h is the Planck constant, c is the speed of light, and k_B is the Boltzmann constant. Hence, the maximum achievable distance in CV-QKD can be expressed as a function of γ, T, and λ_{wave} as follows:

$$\begin{aligned} L_{\max} &= \frac{10\log_{10} e}{\gamma}\left(\frac{hc}{\lambda_{\mathrm{wave}} k_B T}\right) \\ &\simeq \left(\frac{298\,\mathrm{K}}{T}\right)\left(\frac{1.5\,\mu\mathrm{m}}{\lambda_{\mathrm{wave}}}\right)\frac{139\,\mathrm{km}}{\gamma}\,. \end{aligned} \tag{5.5}$$

This serves as the ultimate benchmark for CV-QKD: if the fibre length exceeds $L_{\max}$, quantum mechanics rules out the possibility of achieving CV-QKD without intermediate nodes; conversely, if the length is smaller than $L_{\max}$, CV-QKD without intermediate nodes is achievable.

To illustrate the significance of (5.5), consider its application to the current Internet infrastructure. Set the attenuation coefficient γ to the lowest recorded value of $\gamma \simeq 0.14$, the temperature T to the standard room temperature of $T = 298\,\mathrm{K}$, and the wavelength λ_{wave} to the telecom wavelength of $\lambda_{\mathrm{wave}} \simeq 1.5\,\mu\mathrm{m}$, which is used in the current Internet infrastructure as well as in state-of-the-art CV-QKD experiments [6–10]. By using (5.5), the maximum achievable distance is estimated as

$$L_{\max} \simeq 990\,\mathrm{km}\,. \tag{5.6}$$

This result is informative: it indicates that achieving CV-QKD over optical fibres longer than approximately 990 km necessitates the use of intermediate (trusted, and thus more costly) nodes. Conversely, it also establishes that there exists a protocol enabling point-to-point CV-QKD over distances up to approximately 990 km. Strikingly, this estimated maximum distance of approximately 990 km is not far from the distances achieved by state-of-the-art CV-QKD experiments, which are around 200 km [6–10]. This ultimately establishes that to go much beyond the current distance records, the use of intermediate nodes is essential.

The above result can be generalised to encompass all Gaussian channels [2], including, besides the thermal attenuator $\mathscr{E}_{\lambda,\nu}$, also the *thermal amplifier* $\Phi_{g,\nu}$, the *additive Gaussian noise* Λ_ξ, and many others. The channels $\mathscr{E}_{\lambda,\nu}$, $\Phi_{g,\nu}$, and Λ_ξ are collectively called "phase-insensitive bosonic Gaussian channels" (piBGCs). The thermal amplifier $\Phi_{g,\nu}$ is characterised by two parameters, the gain $g \in [0,\infty)$ and the added thermal noise ν. The additive Gaussian noise Λ_ξ is characterised by a single parameter $\xi \in [0,\infty)$, representing the added *classical* noise [2]. Detailed definitions of these channels are provided in the Methods. In the forthcoming Theorem 1 we characterise the parameter region where the two-way capacities of the piBGCs are strictly positive, i.e. where they can be used for entanglement distribution and QKD. Specifically, we show that this is possible if and only if the channel is not *entanglement breaking* [14]. This provides a simple and concise solution to the problem posed in [12, Section 7]; remarkably, our proof is based solely on Gaussian quantum information techniques [2].

Theorem 1. *Let Φ be a single-mode Gaussian channel and let $N_s > 0$. The energy-constrained two-way quantum capacity $Q_2(\Phi, N_s)$ and secret-key capacity $K(\Phi, N_s)$ are strictly positive if and only if Φ is not entanglement breaking.*

In particular, for all $\nu \geq 0$, $\lambda \in [0,1]$, $g \geq 1$, and $\xi \geq 0$, the energy-constrained two-way quantum capacity of the thermal attenuator $\mathscr{E}_{\lambda,\nu}$, thermal amplifier $\Phi_{g,\nu}$, and additive Gaussian noise Λ_ξ satisfy:

$$\begin{aligned} Q_2(\mathscr{E}_{\lambda,\nu}, N_s) > 0 &\Leftrightarrow \lambda \in \left(\frac{\nu}{\nu+1}, 1\right] , \\ Q_2(\Phi_{g,\nu}, N_s) > 0 &\Leftrightarrow g \in \left[1, \frac{\nu+1}{\nu}\right) , \\ Q_2(\Lambda_\xi, N_s) > 0 &\Leftrightarrow \xi \in [0, 1) . \end{aligned} \tag{5.7}$$

The same holds for the secret-key capacity $K(\cdot, N_s)$ as well as for the unconstrained capacities.

Proof. Since any entanglement-breaking channel has vanishing two-way capacities [14], it suffices to consider the case where Φ is not entanglement breaking. Assume that Alice prepares many copies of the two-mode squeezed vacuum state $|\Psi_{N_s}\rangle_{AA'}$ with mean local photon number N_s, i.e.

$$|\Psi_{N_s}\rangle := \frac{1}{\sqrt{N_s+1}} \sum_{n=0}^{\infty} \left(\frac{N_s}{N_s+1}\right)^{n/2} |n\rangle_A \otimes |n\rangle_{A'} , \tag{5.8}$$

and sends the systems A' through the Gaussian channel $\Phi_{A'\to B}$. Now Alice and Bob share many copies of the two-mode Gaussian state

$$C_{N_s} := \left(\mathrm{Id}_A \otimes \Phi_{A'\to B}\right)(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}) , \tag{5.9}$$

which is a *generalised Choi state* of Φ. By definition, a generalised Choi state of a quantum channel $\mathscr{N}_{A'\to B}$ is a bipartite state of the form $\left(\mathrm{Id}_A \otimes \mathscr{N}_{A'\to B}\right)(|\Psi\rangle\langle\Psi|_{AA'})$, where the input state $|\Psi\rangle_{AA'}$ is such that its reduced state $\mathrm{Tr}_A[|\Psi\rangle\langle\Psi|_{AA'}]$ is invertible. In Lemma S9 of the Supplementary Information we prove that all the generalised Choi states of a non-entanglement-breaking channel are entangled. In particular, we deduce that C_{N_s} is entangled. Additionally, by exploiting the fact that a two-mode Gaussian state is entangled if and only if it is not *PPT* [2, 30, 31], it thus follows that C_{N_s} is not PPT. Finally, since any two-mode Gaussian state that it is not PPT is also *distillable* [32] — i.e. it can be converted into ebits with a strictly positive rate — we conclude that $K(\Phi, N_s) \geq Q_2(\Phi, N_s) > 0$.

If Φ is a piBGC, an alternative, more explicit proof of the fact that C_{N_s} is entangled uses the following entanglement criterion [2, 31]: a two-mode Gaussian state is entangled if and only if its covariance matrix $V := \begin{pmatrix} V_A & V_{AB} \\ V_{AB}^\intercal & V_B \end{pmatrix}$ satisfies the condition

$$1 + \det V + 2\det V_{AB} < \det V_A + \det V_B . \tag{5.10}$$

In Theorem S12 of the Supplementary Information we show that the covariance matrix of C_{N_s} satisfies such a condition.

The parameter regions in (5.7) are precisely those where $\mathscr{E}_{\lambda,\nu}$, $\Phi_{g,\nu}$, and Λ_ξ are not entanglement breaking [17, 33, 34].

□

We remark that Theorem 1 establishes the maximum tolerable noise not only in key distribution but also in entanglement distribution. Thus, the condition $\lambda > \frac{\nu}{\nu+1}$ serves as a necessary and sufficient condition not only for achieving CV-QKD without relying on intermediate nodes but also for achieving entanglement distribution without the use of (possibly expensive) quantum repeaters [18, 19].

As another example of application of Eq. (5.1), using [35, Eq. (1)–(2)] and the values in [35, Table I], one can see that entanglement distribution between Earth and the Moon, with lenses of aperture $a_R = w_0 = 5\,\mathrm{cm}$, are only possible at wavelengths below 0.17 mm, due to the cosmic microwave background at 2.725 K.

5.2.2 Improved lower bound on two-way capacities

While (5.7) establishes the parameter ranges for which the piBGCs have positive capacities, it does not give us any explicit estimate on those capacities. Here we do precisely that, finding a new lower bound on the two-way capacities of the piBGCs that constitutes a significant improvement upon the state-of-the-art lower bounds [12, 22, 25–27, 36] in a large parameter region. Our result is fully constructive, as it is proved by designing and analysing a concrete entanglement distribution protocol.

The best known lower bounds on the two-way capacities of the piBGCs, prior to our work, are the (reverse) *coherent information* lower bounds [22, 25]. These bounds are derived by evaluating the ebit rate of a two-step entanglement distribution protocol: first, Alice sends halves of the two-mode squeezed vacuum state through the channel; second, Alice and Bob apply the hashing protocol [37, 38] to distil entanglement. Our protocol improves on both steps. It draws inspiration from techniques used for distilling entanglement from two-qubit Werner states [39, 40], where the ebit rate can be increased by introducing a *recurrence stage* [38–40] before the hashing protocol. In essence, our protocol involves sending halves of a suitably encoded finite-dimensional entangled state into the channel; projecting the channel's output into an appropriate finite-dimensional subspace; applying recurrence protocols [38–41]; and finally executing an improved version of the hashing protocol [42]. While some steps of our protocol are obtained by suitably combining existing finite-dimensional protocols, the conceptual and technical novelty of our approach is to find a way to effectively apply them to the continuous-variable setting at hand.

Let us present our entanglement distribution protocol across a piBGC Φ. It is composed of six steps named S1–S6, and it depends on three parameters over which we will optimise numerically: $M \in \mathbb{N}^+$, $c \in (0,1)$, $k \in \mathbb{N}$. These parameters have the following intuitive interpretations: M represents the maximum number of photons in the input state of the protocol, c indicates the level of coherence in that state, and k corresponds to the number of iterations of a specific subroutine within the protocol.

Entanglement distribution protocol:

S1: Alice prepares many copies of the state

$$|\Psi_{M,c}\rangle_{AA'} \coloneqq c\,|e_0\rangle_A \otimes |0\rangle_{A'} + \sqrt{1-c^2}\,|e_1\rangle_A \otimes |M\rangle_{A'}\,, \tag{5.11}$$

and sends the subsystem A' to Bob through the channel Φ. Here, $|0\rangle_{A'}$ and $|M\rangle_{A'}$ denote the vacuum and the Mth Fock state, while $|e_0\rangle_A$ and $|e_1\rangle_A$ represent orthogonal states of Alice's register A, which may be any quantum system (optical or not). As detailed in Section III.C of the Supplementary Information, an experimental realisation of the state in (5.11) may involve *NOON states* [43].

Now Alice and Bob share many pairs of the state $\big(\mathrm{Id}_A \otimes \Phi\big)(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|)$. If there is an energy constraint N_s, the parameters c and M have to satisfy $(1-c^2)M \leq N_s$.

S2: Bob performs the local POVM $\{\Pi_M, \mathbb{1} - \Pi_M\}$ on each pair, where $\Pi_M \coloneqq |0\rangle\langle 0| + |M\rangle\langle M|$. If Bob finds the outcome associated with Π_M, then the pair is kept, otherwise it is discarded. Hence, by re-mapping $|0\rangle \to |e_0\rangle$ and $|M\rangle \to |e_1\rangle$, each pair is now in an effective two-qubit state.

S3: Alice and Bob apply the Pauli-based twirling, reported in (5.16) in the Methods, in order to transform each of the remaining pairs in a Bell-diagonal state.

S4: Alice and Bob run k times the following sub-routine, dubbed *P1-or-P2* [41].

P1-or-P2 sub-routine:

- **Step 4.1**: At this point of the protocol, the two-qubit state ρ_{AB} of each pair is of the form

$$\rho_{AB} = \sum_{i,j=0}^{1} p_{ij}\, \mathbb{1}_A \otimes X^j Z^i \, |\psi_{00}\rangle\langle\psi_{00}| \, (\mathbb{1}_A \otimes X^j Z^i)^\dagger \,, \tag{5.12}$$

  where $|\psi_{00}\rangle$ denotes the ebit state (see (5.13) of the Methods), X, Z denote the well-known Pauli operators, and $\{p_{ij}\}_{i,j\in\{0,1\}}$ is a probability distribution.

  Alice and Bob collect all pairs in groups of two. For a given group, call A_1B_1 and A_2B_2 the four qubits involved. If $p_{10} < p_{01}$, they apply the CNOT gate on A_1A_2 and B_1B_2, respectively, where A_1, B_1 are the control qubits and A_2, B_2 are the target qubits. Otherwise, they apply first the Hadamard gate on each qubit, then the CNOT gate as in the above case, and finally the Hadamard gate on A_1 and B_1.

- **Step 4.2**: For each group of two pairs, Alice and Bob perform a projective measurement in the computational basis of A_2 and B_2, thereby discarding these systems. If the outcomes are different, they discard also the pair A_1B_1.

The sub-routine tends to increase the value of p_{00}, bringing the state ρ_{AB} closer to the ebit state $|\psi_{00}\rangle$. The condition $p_{10} < p_{01}$ in Step 4.1 means that the X error in (5.12) is more prominent than the Z error. If that is the case, the sub-routine reduces the X error at the expense of the Z error. However, by selectively applying the two procedures both errors end up being corrected [41].

S5: Depending on the state ρ of the remaining pairs (see Section III.A of the Supplementary Information for details), both Alice and Bob apply on each qubit one of the following unitaries: the qubit rotation around the y-axis of an angle $\pi/2$, the Hadamard gate, or the identity.

S6: In the end, Alice and Bob run the improved version of the hashing protocol introduced in [42] in order to generate ebits.

Our lower bound on the two-way quantum capacity $Q_2(\Phi)$, calculated in Theorem S14 of the Supplementary Information, is the supremum over $M \in \mathbb{N}^+$, $c \in (0,1)$, and $k \in \mathbb{N}$ of the ebit rate of the above protocol. Since the secret-key capacity K is always larger than Q_2, the resulting expression is also a lower bound on $K(\Phi)$. Our lower bound on the energy-constrained two-way capacities with energy constraint N_s is obtained by optimising with the additional condition $(1-c^2)M \leq N_s$.

We plot our bounds on the two-way capacities of the thermal attenuator $\mathscr{E}_{\lambda,\nu}$ in Fig. 1(a), of the thermal amplifier $\Phi_{g,\nu}$ in Fig. 1(b), and of the additive Gaussian noise Λ_ξ in Fig. 1(c). These plots demonstrate that our bound is strictly tighter than all known lower bounds on both the two-way quantum and the secret-key capacity [12, 22, 25, 27, 36] in a large parameter region. Additionally, as shown in Section V of the Supplementary Information, the energy-constrained version of our bound outperforms the tightest known lower bound on the energy-constrained two-way capacities [26] in a substantial parameter region. Indeed, our numerical analysis reveals that the optimal value of M in (5.11) is never greater than three. This implies that our protocol is highly energy efficient, using only states with up to three photons to distribute entanglement.

Notably, unlike all known lower bounds [12, 22, 25–27, 36], which vanish in a large parameter region, we observe numerically that our lower bound is *faithful*: it remains strictly positive if and only if the tightest known upper bound [17] is also strictly positive. Note that the proof of our Theorem 1 provides an alternative entanglement distribution protocol which is mathematically guaranteed to be faithful. However, its ebit rate is much lower than that of the protocol presented in this section.

In Section IV of the Supplementary Information, we introduce an additional entanglement distribution protocol that combines and optimises the multi-rail protocol from [44] and the qudit P1-or-P2 protocol from [41]. The ebit rate of this protocol is calculated in Theorem S14 of the

Supplementary Information and constitutes an additional lower bound on the two-way capacities of the piBGCs. Importantly, this additional lower bound turns out to be tighter than our previously discussed lower bound on $Q_2(\mathscr{E}_{\lambda,\nu})$ in the low excess noise regime $\nu \lesssim 1$, as shown in Fig. 6 of the Supplementary Information. Note that in Fig. 1 we considered a higher excess noise regime where $\nu = 10$.

Finally, let us briefly discuss the experimental feasibility of the entanglement distribution protocol introduced in this section. This protocol encounters the same experimental challenges as all known entanglement distillation protocols, as it primarily consists of the fundamental building blocks of any entanglement distillation protocol, namely recurrence and hashing protocols [38–42]. Significant experimental progress has been made in recent years regarding entanglement distillation [45–47], leaving hope that these experimental challenges will be mitigated in the near future. See Section III.C of the Supplementary Information for a detailed discussion of the experimental feasibility of our protocol.

Discussion

In this work, we have determined a simple necessary and sufficient condition for achieving continuous-variable quantum-key distribution without intermediate nodes [4] and entanglement distribution without quantum repeaters [18, 19]: point-to-point key distribution and entanglement distribution across an optical link of transmissivity λ and added thermal noise ν are achievable if and only if $\lambda > \frac{\nu}{\nu+1}$. This solves completely the problem of determining the ultimate limitations to CV-QKD imposed by the laws of quantum physics [12]. By leveraging this result, we have established the maximum achievable distance of CV-QKD, demonstrating that the current Internet infrastructure can in principle support CV-QKD only up to distances of approximately 990 km, which is not so far from the distance records of approximately 200 km achieved by state-of-the-art experiments [6–10]. Consequently, surpassing these distance records by an order of magnitude at standard telecom wavelengths will necessarily require the use of intermediate nodes.

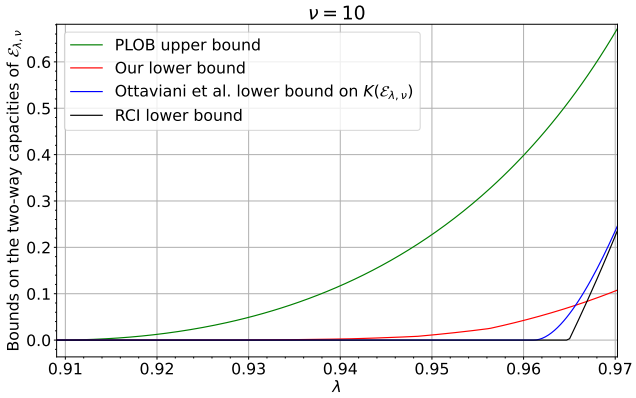
Moreover, we have derived the best lower bound to date on the two-way quantum and secret-key capacities of all phase-insensitive bosonic Gaussian channels in the regime of large noise — in a large parameter region, ours is in fact the *only* non-zero lower bound. Our new bound constitutes a significant improvement upon state-of-the-art lower bounds [12, 25–27, 36]. We remark that the last improvement on the (unconstrained) two-way quantum capacity prior to our work dates back to 2009 [25]. It was an open question whether the latter could equal the true two-way quantum capacities: our work provides a negative answer to this question, showing that entanglement distribution is possible in a much broader parameter region than previously known. Our results are fully explicit: to prove our lower bound, we constructed a general yet relatively simple entanglement distribution (and hence also key distribution) protocol that works whenever the underlying channel is not entanglement breaking.

In conclusion, we have discovered new protocols and fundamental limitations to quantum communication across optical links, establishing the ultimate noise threshold at which entanglement and secret keys can be distributed. This is likely to bear a significant impact on the design of practical QKD and entanglement distribution protocols on optical networks.

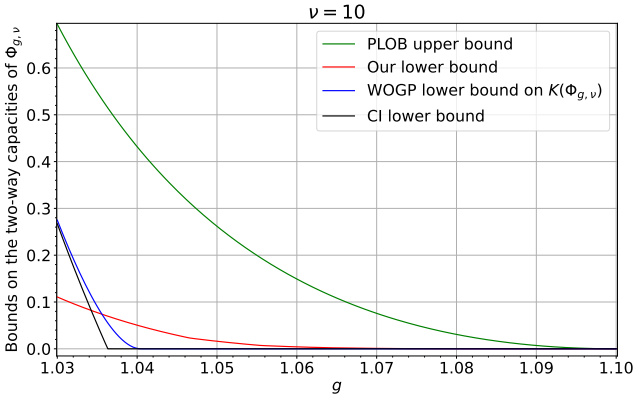
Methods

Let $\mathfrak{S}(\mathcal{H})$ denote the space of density operators on a Hilbert space $\mathcal{H}$. Let $\mathcal{H}_2$ be a single-qubit Hilbert space with orthonormal basis $\{|0\rangle, |1\rangle\}$. For all $i, j \in \{0, 1\}$, the state $|\psi_{ij}\rangle_{AB} \in \mathcal{H}_2^{(A)} \otimes \mathcal{H}_2^{(B)}$ defined as

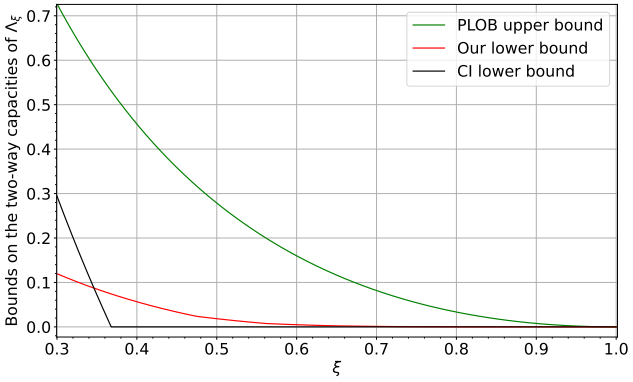
$$|\psi_{ij}\rangle_{AB} := \frac{1}{\sqrt{2}} \sum_{m=0}^{1} (-1)^{im} |m\rangle_A \otimes |m \oplus j\rangle_B \ , \tag{5.13}$$



(a)



(b)



(c)

Figure 5.1: Bounds on the two-way capacities of the thermal attenuator, thermal amplifier, and additive Gaussian noise. **(a).** Bounds on the two-way quantum capacity Q_2 and secret-key capacity K of the thermal attenuator $\mathcal{E}_{\lambda,\nu}$ plotted with respect to λ for $\nu = 10$. The red line is our lower bound; the black line is the reverse coherent information lower bound reported in (5.27); the blue line is the lower bound on $K(\mathcal{E}_{\lambda,\nu})$ discovered by Ottaviani et al. [27] (this is not a lower bound on Q_2, since in general it only holds that $Q_2 \leq K$); and the green line is the upper bound discovered by Pirandola et al. [17] reported in (5.26). **(b).** Bounds on the two-way quantum capacity Q_2 and secret-key capacity K of the thermal amplifier $\Phi_{g,\nu}$ plotted with respect to g for $\nu = 10$. The red line is our lower bound; the black line is the coherent information lower bound reported in (5.28); the blue line is the lower bound on $K(\Phi_{g,\nu})$ discovered by Wang et al. [36]; and the green line is the upper bound discovered by Pirandola et al. [17] reported in (5.26). **(c).** Bounds on the two-way quantum capacity Q_2 and secret-key capacity K of the additive Gaussian noise Λ_ξ plotted with respect to ξ. The red line is our lower bound; the black line is the coherent information lower bound reported in (5.29); and the green line is the upper bound discovered by Pirandola et al. [17] reported in (5.26).

where $\oplus$ denotes the modulo 2 addition, is called a Bell state (or maximally entangled state). We will also refer to $|\psi_{00}\rangle$ as an entanglement bit, or *ebit*. Any Bell state can be written in terms of the ebit as

$$|\psi_{ij}\rangle_{AB} = \mathbb{1}_A \otimes X^j Z^i \, |\psi_{00}\rangle_{AB} \ , \tag{5.14}$$

where $\mathbb{1}_A$ denotes the identity operator on $\mathcal{H}_2^{(A)}$ and X, Z denote the well-known Pauli operators on $\mathcal{H}_2^{(B)}$. The Pauli-based twirling

$$\mathscr{T} : \mathfrak{S}(\mathcal{H}_2^{(A)} \otimes \mathcal{H}_2^{(B)}) \to \mathfrak{S}(\mathcal{H}_2^{(A)} \otimes \mathcal{H}_2^{(B)}) \tag{5.15}$$

is defined as

$$\mathscr{T}(\rho_{AB}) = \frac{1}{4} \sum_{i,j=0}^{1} X_A^j Z_A^i \otimes X_B^j Z_B^i \, \rho_{AB} \, (X_A^j Z_A^i \otimes X_B^j Z_B^i)^\dagger \tag{5.16}$$

for all $\rho_{AB} \in \mathfrak{S}(\mathcal{H}_2^{(A)} \otimes \mathcal{H}_2^{(B)})$ and it maps any input state in a Bell-diagonal state:

$$\mathscr{T}(\rho_{AB}) = \sum_{i,j=0}^{1} \langle \psi_{ij}| \, \rho_{AB} \, |\psi_{ij}\rangle \ |\psi_{ij}\rangle\langle\psi_{ij}|_{AB} \, . \tag{5.17}$$

Physically realisable transformations between two quantum systems with Hilbert spaces $\mathcal{H}$ (input) and $\mathcal{H}'$ (output) are modelled by quantum channels, i.e. completely positive and trace preserving maps $\Phi : \mathfrak{S}(\mathcal{H}) \to \mathfrak{S}(\mathcal{H}')$. The two-way quantum capacity $Q_2(\Phi)$ and secret-key capacity $K(\Phi)$ of a quantum channel Φ is the maximum achievable rate of qubits and secret-key bits, respectively, that can be reliably transmitted through Φ by assuming that the sender Alice and the receiver Bob have free access to a public, noiseless, two-way classical communication line. The rate of qubits (resp. secret-key bits) is defined as the ratio between the number of reliably transmitted qubits (resp. secret-key bits) and the number of uses of Φ. A rigorous definition of the two-way capacities can be found in [14, Chapters 14 and 15]. For any Φ, the two-way capacities satisfy

$$Q_2(\Phi) \leq K(\Phi) \, , \tag{5.18}$$

since an ebit can generate a secret-key bit, thanks to the 'E91' protocol [16].

In practice, Alice has access to a limited budget N_s of energy to produce each input signal, as measured by a Hamiltonian H on $\mathcal{H}$. Fixed $N_s > 0$, the energy-constrained two-way capacities $Q_2(\Phi, N_s)$ and $K(\Phi, N_s)$ are defined in the same way as the two-way capacities defined above, apart from the fact that the maximisation of the rate is restricted to the strategies such that the average expected value of H on all input signals is required to be at most N_s. In addition note that the generalisation of (5.18) to the energy-constrained case holds, i.e.

$$Q_2(\Phi, N_s) \leq K(\Phi, N_s) \, , \tag{5.19}$$

that any energy-constrained capacity is upper bounded by the corresponding unconstrained capacity, and that it tends to it in the limit $N_s \to \infty$.

The goal of an entanglement distillation protocol is to turn a large number n of copies of a bipartite entangled state ρ_{AB} shared between Alice and Bob into a number m of ebits by local operations and classical communication. The yield is defined by the ratio m/n. The distillable entanglement $E_d(\rho_{AB})$ of ρ_{AB} is defined as the maximum yields over all the possible entanglement distillation protocols [38] [14, Chapter 8]. The state ρ_{AB} is said to be *distillable* if $E_d(\rho_{AB}) > 0$. The coherent information (resp. reverse coherent information) of ρ_{AB} is defined by

$$I_c(\rho_{AB}) \coloneqq S(\rho_B) - S(\rho_{AB}) \tag{5.20}$$

(resp. $I_{rc}(\rho_{AB}) \coloneqq S(\rho_A) - S(\rho_{AB})$), where $\rho_B \coloneqq \mathrm{Tr}_A \, \rho_{AB}$ and analogously for ρ_A, and moreover

$$S(\sigma) \coloneqq -\mathrm{Tr}[\sigma \log_2 \sigma] \tag{5.21}$$

is the von Neumann entropy. The yield $I_c(\rho_{AB})$ (resp. $I_{rc}(\rho_{AB})$) is achievable by an entanglement distillation protocol [37] that only exploits one-way forward (resp. backward) classical communication. In particular, the following inequality, known as *hashing inequality*, holds:

$$E_d(\rho_{AB}) \geq \max\{I_c(\rho_{AB}), I_{rc}(\rho_{AB})\}. \tag{5.22}$$

Let us briefly link the notions of distillable entanglement E_d and two-way quantum capacity $Q_2(\Phi, N_s)$. Suppose that Alice produces n copies of a state $\rho_{AA'}$ such that $\operatorname{Tr}\rho_{A'}H_{A'} \leq N_s$. Then, she can use the channel n times to send all subsystems A' to Bob. Then, Alice and Bob share n copies of $\mathrm{Id}_A \otimes \Phi(\rho_{AA'})$, which can now be used to generate $\simeq n\, E_d\left(\mathrm{Id}_A \otimes \Phi(\rho_{AA'})\right)$ ebits by means of a suitable entanglement distillation protocol. Consequently, it holds that

$$Q_2(\Phi, N_s) \geq E_d\left(\mathrm{Id}_A \otimes \Phi(\rho_{AA'})\right). \tag{5.23}$$

In the context of entanglement distillation, the goal of a *recurrence protocol* is to transform a certain number of copies of the state ρ_{AB} into fewer copies of another state ρ'_{AB} such that $\langle\psi_{00}|\,\rho'_{AB}\,|\psi_{00}\rangle > \langle\psi_{00}|\,\rho_{AB}\,|\psi_{00}\rangle$ [38–40]. Examples of recurrence protocols for qubits can be found in [40, 48, 49], and their generalisations to the case of qudits in [50–52]. In the present paper we will exploit the recently introduced P1-or-P2 recurrence protocol [41]. Since an infinite number of iterations of a recurrence protocol is generally needed to generate a Bell state $|\psi_{00}\rangle$, the yield of a recurrence protocol is zero. To achieve a nonzero yield, one may adopt a suitable number of iterations of a recurrence protocol and then apply the hashing or breeding protocol [39, 40]. The latter protocols, which exploit only one-way classical communication, achieves the yield of the hashing inequality in (5.22). Improvements of the hashing and breeding protocols, which exploit two-way classical communication and work on bipartite-qubit systems that are diagonal in the Bell basis, have been provided in [42].

Let $\mathcal{H}_S, \mathcal{H}_E := L^2(\mathbb{R})$ be single modes of electromagnetic radiation with definite frequency and polarisation, and let a and b be the corresponding annihilation operators. Now, let us define the piBGCs. For all $\lambda \in [0,1]$, $g \geq 1$, $\nu \geq 0$, $\xi \geq 0$, the thermal attenuator $\mathscr{E}_{\lambda,\nu}$, the thermal amplifier $\Phi_{g,\nu}$, and the additive Gaussian noise Λ_ξ are quantum channels on $\mathcal{H}_S$ defined by

$$\begin{aligned}
\mathscr{E}_{\lambda,\nu}(\rho) &:= \operatorname{Tr}_E\left[U_\lambda^{SE}\left(\rho^S \otimes \tau_\nu^E\right)U_\lambda^{SE\dagger}\right], \\
\Phi_{g,\nu}(\rho) &:= \operatorname{Tr}_E\left[U_g^{SE}\left(\rho^S \otimes \tau_\nu^E\right)U_g^{SE\dagger}\right], \\
\Lambda_\xi(\rho) &:= \frac{1}{\pi\xi}\int_{\mathbb{C}} \mathrm{d}^2 z\, e^{-\frac{|z|^2}{\xi}} D(z)\,\rho\, D(z)^\dagger,
\end{aligned} \tag{5.24}$$

where $\tau_\nu^E := \frac{1}{\nu+1}\sum_{n=0}^{\infty}\left(\frac{\nu}{\nu+1}\right)^n |n\rangle\langle n|_E$ is the thermal state with $\{|n\rangle_E\}_{n\in\mathbb{N}}$ being the Fock states on $\mathcal{H}_E$, U_λ^{SE} is the beam splitter unitary of transmissivity λ, U_g^{SE} is the two-mode squeezing unitary of gain g, and $D(z)$ is the displacement operator:

$$\begin{aligned}
U_\lambda^{SE} &:= \exp\left[\arccos\sqrt{\lambda}\left(a^\dagger b - a\, b^\dagger\right)\right], \\
U_g^{SE} &:= \exp\left[\operatorname{arccosh}\sqrt{g}\left(a^\dagger b^\dagger - a\, b\right)\right], \\
D(z) &:= \exp\left[z a^\dagger - z^* a\right].
\end{aligned} \tag{5.25}$$

In a communication scenario, the piBGCs are understood to map Alice's single-mode systems A' to Bob's single-mode systems B. The Hamiltonian on $\mathcal{H}_S$ is the photon number operator $a^\dagger a$ and, by definition, the energy of an input signal initialised in a state ρ is equal to its mean photon number $\operatorname{Tr}[\rho\, a^\dagger a]$. The tightest known upper bounds on the two-way capacities of these channels,

shown by Pirandola et al. [17], are

$$
\begin{aligned}
K(\mathscr{E}_{\lambda,\nu}) &\leq \begin{cases} -h(\nu) - \log_2[(1-\lambda)\lambda^\nu], & \text{if } \lambda \in (\frac{\nu}{\nu+1}, 1], \\ 0, & \text{otherwise} \end{cases} \\
K(\Phi_{g,\nu}) &\leq \begin{cases} -h(\nu) + \log_2\left(\frac{g^{\nu+1}}{g-1}\right), & \text{if } g \in [1, 1+\frac{1}{\nu}), \\ 0, & \text{otherwise} \end{cases} \\
K(\Lambda_\xi) &\leq \begin{cases} \frac{\xi-1}{\ln 2} - \log_2(\xi), & \text{if } \xi \in [0,1), \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
\tag{5.26}
$$

where $h(\nu) \coloneqq (\nu+1)\log_2(\nu+1) - \nu \log_2 \nu$ (see [23] for a strong-converse extension of the formulas above). These upper bounds vanish if and only if the piBGCs are entanglement breaking [17, 33, 34]. The tightest known lower bounds (before our work) on Q_2 are [22, 25]

$$
Q_2(\mathscr{E}_{\lambda,\nu}) \geq \max\{0, -h(\nu) - \log_2(1-\lambda)\}, \tag{5.27}
$$

$$
Q_2(\Phi_{g,\nu}) \geq \max\left\{0, -h(\nu) + \log_2\left(\frac{g}{g-1}\right)\right\}, \tag{5.28}
$$

$$
Q_2(\Lambda_\xi) \geq \max\{0, -\log_2(e\,\xi)\}. \tag{5.29}
$$

These lower bounds can be proved first by applying (6.33) with the choice $\rho_{AA'} = |\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}$, where

$$
|\Psi_{N_s}\rangle \coloneqq \frac{1}{\sqrt{N_s+1}} \sum_{n=0}^{\infty} \left(\frac{N_s}{N_s+1}\right)^{n/2} |n\rangle \otimes |n\rangle \tag{5.30}
$$

is the two-mode squeezed vacuum state with local mean photon number equal to N_s, second by applying the hashing inequality in (5.22), and finally by taking the limit $N_s \to \infty$. Specifically, the lower bound in (5.27) is achieved by the reverse coherent information, while that in (5.28)–(5.29) is achieved by the coherent information. Although the right-hand sides of (5.27)–(5.29) also lower bound the secret-key capacity K, improved estimates of $K(\mathscr{E}_{\lambda,\nu})$ and $K(\Phi_{g,\nu})$ have been put forth by Ottaviani et al. [27] (see also [12, Sec. VII]) and by Wong et al. [36], respectively.

Lower bounds on the energy-constrained two-way capacities with energy constraint N_s of piBGCs are the coherent information and the reverse coherent information evaluated on the state obtained by sending the subsystem A' of $|\Psi_{N_s}\rangle_{AA'}$ through the channel. For sufficiently small values of N_s, improved lower bounds have been found by Noh et al. [26]. The best known upper bound on the energy-constrained two-way capacity of the thermal attenuator is — depending on the parameters λ, ν, and N_s — the unconstrained upper bound discovered by Pirandola et al. [17] reported in (5.26) or the bound found by Davis et al. [15] (which is equal to the bound found in [21] for $\nu = 0$). Upper bounds on the energy-constrained two-way capacities of the thermal amplifier and additive Gaussian noise are the unconstrained upper bound discovered by Pirandola et al. [17] reported in (5.26) and the bounds which can be obtained by exploiting the results of [15, 20].

5.3 Supplementary material

5.3.1 Notation and preliminaries

Let $\mathfrak{S}(\mathcal{H})$ be the set of quantum states on a Hilbert space $\mathcal{H}$. The trace norm of a bounded linear operator Θ is defined by $\|\Theta\|_1 \coloneqq \operatorname{Tr}\sqrt{\Theta^\dagger\Theta}$. The von Neumann entropy of a quantum state ρ is denoted by $S(\rho) \coloneqq -\operatorname{Tr}[\rho \log_2 \rho]$. Let $\mathcal{H}_2$ be a bi-dimensional Hilbert space and let $\{|0\rangle, |1\rangle\}$ be an orthonormal basis. For all $i, j \in \{0,1\}$, the state $|\psi_{ij}\rangle_{AB} \in \mathcal{H}_2^{(A)} \otimes \mathcal{H}_2^{(B)}$ is defined as

$$
|\psi_{ij}\rangle_{AB} \coloneqq \frac{1}{\sqrt{2}} \sum_{m=0}^{1} (-1)^{im} |m\rangle_A \otimes |m \oplus j\rangle_B, \tag{5.31}
$$

and is called a Bell state (or maximally entangled state), where $\oplus$ denotes the modulo 2 addition.

Gaussian quantum information

Let us briefly review the formalism of Gaussian quantum information [2]. We consider m-modes of harmonic oscillators S_1, S_2, ..., S_m, which are associated with the Hilbert space $L^2(\mathbb{R}^m)$ of square integrable functions. Each of these modes represents a single-mode of electromagnetic radiation with definite frequency and polarisation. For all $j = 1, 2, \ldots, m$ the annihilation operator a_j of the mode S_i is defined as $a_j := \frac{\hat{x}_j + i\hat{p}_j}{\sqrt{2}}$, where $\hat{x}_j$ and $\hat{p}_j$ are the well-known position and momentum operators of S_j. The operator $a_j^\dagger a_j$ is called the photon number of the mode S_j. The nth Fock state of the mode S_j is denoted by $|n\rangle_{S_j}$. By defining the so-called quadrature vector $\hat{\mathbf{R}} := (\hat{x}_1, \hat{p}_1, ..., \hat{x}_m, \hat{p}_m)^\intercal$, one can write the canonical commutation relations as $[\hat{\mathbf{R}}, \hat{\mathbf{R}}^\intercal] = i\,\Omega_m$, where $\Omega_m := \mathbb{1}_m \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\mathbb{1}_m$ is the $m \times m$ identity matrix. The characteristic function $\chi_\rho : \mathbb{R}^{2m} \to \mathbb{C}$ of a state $\rho \in \mathfrak{S}(L^2(\mathbb{R}^m))$ is defined as $\chi_\rho(\mathbf{r}) = \mathrm{Tr}[\rho D_{-\mathbf{r}}]$, where for all $\mathbf{r} \in \mathbb{R}^{2m}$ the displacement operator $D_\mathbf{r}$ is defined as

$$D_\mathbf{r} := e^{i\mathbf{r}^\intercal \Omega_m \hat{\mathbf{R}}}\,. \tag{5.32}$$

Any state ρ can be written in terms of its characteristic function as

$$\rho = \int_{\mathbb{R}^{2m}} \frac{\mathrm{d}^{2m}\mathbf{r}}{(2\pi)^m} \chi_\rho(\mathbf{r}) D_\mathbf{r} \tag{5.33}$$

and hence quantum states and characteristic functions are in one-to-one correspondence. The first moment and the covariance matrix of a quantum state ρ are defined as

$$\mathbf{m}(\rho) = \mathrm{Tr}\left[\hat{\mathbf{R}}\,\rho\right]\,, \tag{5.34}$$

$$V(\rho) = \mathrm{Tr}\left[\left\{(\hat{\mathbf{R}} - \mathbf{m}(\rho)), (\hat{\mathbf{R}} - \mathbf{m}(\rho))^\intercal\right\}\rho\right]\,, \tag{5.35}$$

respectively, where $\{A, B\} := AB + BA$ is the anti-commutator. Note that the covariance matrix is defined with respect an ordering of the modes in the definition of the quadrature vector: here such an ordering is $(S_1, S_2, \ldots, S_m)$. A state ρ is said to be Gaussian if there exists a $2m \times 2m$ real positive definite matrix H_ρ and a vector $\mathbf{m}_\rho \in \mathbb{R}^{2m}$ such that ρ can be written as a ground or a thermal state of the Hamiltonian $\frac{1}{2}(\hat{\mathbf{R}} - \mathbf{m}_\rho)^\intercal H (\hat{\mathbf{R}} - \mathbf{m}_\rho)$, i.e.

$$\rho = \frac{e^{-\frac{1}{2}(\hat{\mathbf{R}} - \mathbf{m}_\rho)^\intercal H_\rho (\hat{\mathbf{R}} - \mathbf{m}_\rho)}}{\mathrm{Tr}\left[e^{-\frac{1}{2}(\hat{\mathbf{R}} - \mathbf{m}_\rho)^\intercal H_\rho (\hat{\mathbf{R}} - \mathbf{m}_\rho)}\right]}\,. \tag{5.36}$$

It can be shown that $\mathbf{m}(\rho) = \mathbf{m}_\rho$ and $V(\rho) = V_\rho$, where $V_\rho := \coth\left(\frac{i\,\Omega_m H_\rho}{2}\right) i\,\Omega_m$. The characteristic function of a Gaussian state ρ is a Gaussian function in $\mathbf{r}$ which can be written in terms of $\mathbf{m}(\rho)$ and $V(\rho)$ as

$$\chi_\rho(\mathbf{r}) = \exp\left(-\frac{1}{4}(\Omega_m \mathbf{r})^\intercal V(\rho) \Omega_m \mathbf{r} + i(\Omega_m \mathbf{r})^\intercal \mathbf{m}(\rho)\right)\,. \tag{5.37}$$

An example of Gaussian state is the thermal state $\tau_{N_s} := \frac{1}{N_s+1} \sum_{n=0}^{\infty} \left(\frac{N_s}{N_s+1}\right)^n |n\rangle\langle n|$, where the parameter $N_s \geq 0$ is its mean photon number ($N_s = \mathrm{Tr}[a^\dagger a\, \tau_{N_s}]$), which satisfies

$$\begin{aligned} \mathbf{m}(\tau_{N_s}) &= (0, 0)^\intercal\,, \\ V(\tau_{N_s}) &= (2N_S + 1)\mathbb{1}_2\,. \end{aligned} \tag{5.38}$$

Another example of Gaussian state is the two-mode squeezed vacuum state $|\Psi_{N_s}\rangle_{S_1S_2}$, which for all $N_s \geq 0$ it is defined as

$$|\Psi_{N_s}\rangle_{S_1S_2} \coloneqq \frac{1}{\sqrt{N_s+1}} \sum_{n=0}^{\infty} \left(\frac{N_s}{N_s+1}\right)^{n/2} |n\rangle_{S_1} |n\rangle_{S_2} \,, \tag{5.39}$$

where N_s denotes the mean photon number of the mode S_1 (or, equivalently, of the mode S_2), i.e.

$$N_s = \mathrm{Tr}_{S_2}[a_1^\dagger a_1 \; |\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{S_1S_2}] \,. \tag{5.40}$$

The first moment and covariance matrix of $|\Psi_{N_s}\rangle_{S_1S_2}$ are

$$\begin{aligned} &\mathbf{m}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|) = (0,0,0,0)^{\mathrm{T}} \,, \\ &V(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|) = \begin{pmatrix} (2N_s+1)\mathbb{1}_2 & 2\sqrt{N_s(N_s+1)}\sigma_z \\ 2\sqrt{N_s(N_s+1)}\sigma_z & (2N_s+1)\mathbb{1}_2 \end{pmatrix} , \end{aligned} \tag{5.41}$$

where $\mathbb{1}_2 \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\sigma_z \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

A quantum channel is said to be Gaussian if it maps Gaussian states into Gaussian states. Later we will focus on three important examples of Gaussian quantum channels: the thermal attenuator, the thermal amplifier, and the additive Gaussian noise. Before concluding this brief recap of Gaussian quantum information, let us state a lemma which will be useful in the following. The forthcoming Lemma 2 provides a necessary and sufficient condition on the covariance matrix to assess whether a two-mode Gaussian state is entangled [2, 31]. This condition is based on the fact that a two-mode Gaussian states is separable (not entangled) if and only if it is PPT [2, 31].

Lemma 2 [2, 31]**.** *Let $\rho \in \mathfrak{S}(\mathcal{H}_{S_1} \otimes \mathcal{H}_{S_2})$ be a two-mode Gaussian state. Let us write its covariance matrix $V(\rho)$ with respect the ordering (S_1, S_2) as*

$$V(\rho) = \begin{pmatrix} V_{S_1} & V_{S_1S_2} \\ V_{S_1S_2}^{\intercal} & V_{S_2} \end{pmatrix} \tag{5.42}$$

and define the function $f : \mathfrak{S}(\mathcal{H}_{S_1} \otimes \mathcal{H}_{S_2}) \to \mathbb{R}$ as $f(\rho) \coloneqq 1 + \det(V(\rho)) + 2\det(V_{S_1S_2}) - \det(V_{S_1}) - \det(V_{S_2})$. The state ρ is entangled if and only if $f(\rho) < 0$.

The forthcoming Lemma 3 gives necessary and sufficient condition on a Gaussian quantum channel to be entanglement breaking [13, Chapter 4.6].

Lemma 3. *[34] Let $\Phi : \mathfrak{S}(L^2(\mathbb{R}^m)) \to \mathfrak{S}(L^2(\mathbb{R}^m))$ be a Gaussian quantum channel. Let $K, \beta \in \mathbb{R}^{2m\times 2m}$ and $l \in \mathbb{R}^{2m}$ such that for all $\rho \in \mathfrak{S}(L^2(\mathbb{R}^m))$ it holds that*

$$\begin{aligned} &\mathbf{m}\left(\Phi(\rho)\right) = K\,\mathbf{m}(\rho) \,, \\ &V\left(\Phi(\rho)\right) = K^{\intercal} V(\rho) K + \beta \,. \end{aligned} \tag{5.43}$$

Then, Φ is entanglement breaking if and only if β admits the following decomposition:

$$\beta = \alpha + \gamma \,, \quad \text{where } \alpha, \gamma \in \mathbb{R}^{2m\times 2m} \text{ with } \alpha \geq i\,\Omega_m \text{ and } \gamma \geq iK^{\intercal}\,\Omega_m K \,. \tag{5.44}$$

Two-way capacities of a quantum channel

The two-way quantum capacity $Q_2(\Phi)$ and the secret-key capacity $K(\Phi)$ of a quantum channel Φ are the maximum achievable rate of qubits and secret-key bits, respectively, that can be reliably transmitted through Φ by assuming that the sender Alice and the receiver Bob have free access to a public, noiseless, two-way classical communication line. The rate of qubits (resp. secret-key

bits) is defined as the ratio between the number of reliably transmitted qubits (resp. secret-key bits) and the number of uses of Φ [14, Chapters 14 and 15]. An ebit is a Bell state $|\psi_{00}\rangle_{AB}$ shared between Alice and Bob. For any Φ, the two-way capacities satisfy

$$Q_2(\Phi) \leq K(\Phi)\,. \tag{5.45}$$

Indeed, by recalling that Alice and Bob can freely send an infinite amount of bits to each other, an ebit can generate a secret-key bit, thanks to E91 protocol [16], and hence $Q_2(\Phi) \leq K(\Phi)$. The two-way quantum capacity $Q_2(\Phi)$ and the secret-key capacity $K(\Phi)$ are collectively called the *two-way capacities of* Φ.

In practice, Alice has access to a limited budget (N_s) of energy to produce each input signal. Here, by definition, the energy of a signal initialised in a state ρ is equal to its mean photon number $\mathrm{Tr}[\rho\, a^\dagger a]$. Fixed $N_s > 0$, the energy-constrained (EC) two-way capacities $Q_2(\Phi, N_s)$ and $K(\Phi, N_s)$ are defined as above but the maximisation of the rate is restricted to the strategies such that the average photon number less or equal to N_s. In other words, N_s is the maximum allowed average photon number of the input signals to the channel Φ. In addition note that the generalisation of 5.45 to the EC case holds, i.e.

$$Q_2(\Phi, N_s) \leq K(\Phi, N_s)\,, \tag{5.46}$$

and that any EC capacity is upper bounded by the corresponding unconstrained capacity and tends to it in the limit $N_s \to \infty$.

Entanglement distillation

The goal of an entanglement distillation protocol is to turn a large number n of copies of a bipartite entangled state $\rho_{AB} \in \mathfrak{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ shared between Alice and Bob into a smaller number m of ebits by LOCCs (local operations and classical communication). The yield of an entanglement distillation protocol is defined by the ratio m/n. The two-way distillable entanglement $E_d(\rho_{AB})$ of ρ_{AB} is defined as the maximum yields over all the possible entanglement distillation protocols [38] [14, Chapter 8]. The state ρ_{AB} is said to be *distillable* if $E_d(\rho_{AB}) > 0$. The coherent information of ρ_{AB} is defined by

$$I_{\mathrm{c}}(\rho_{AB}) \coloneqq S(\mathrm{Tr}_A\, \rho_{AB}) - S(\rho_{AB}) \tag{5.47}$$

and it is a yield achievable by an entanglement distillation protocol which requires classical communication only from Alice to Bob [37]. By exchanging the roles of Alice and Bob in such an entanglement distillation protocol, the reverse coherent information of ρ_{AB}, which is defined by

$$I_{\mathrm{rc}}(\rho_{AB}) \coloneqq S(\mathrm{Tr}_B\, \rho_{AB}) - S(\rho_{AB})\,, \tag{5.48}$$

is a yield achievable by an entanglement distillation protocol which only requires classical communication only from Bob to Alice [37]. In particular, the following inequality, known as *hashing inequality*, holds:

$$E_d(\rho_{AB}) \geq \max\{I_{\mathrm{c}}(\rho_{AB})\,, I_{\mathrm{rc}}(\rho_{AB})\}\,. \tag{5.49}$$

Let us briefly link the notions of distillable entanglement E_d and two-way quantum capacity $Q_2(\Phi, N_s)$. Suppose that Alice produces n copies of a state $\rho_{AA'}$ such that the mean photon number of the half A' is less or equal to N_s. Then, she uses n times the channel Φ to send the halves A', which satisfy the energy constraint, to Bob. Hence, n copies of $\mathrm{Id}_A \otimes \Phi(\rho_{AA'})$ are shared between Alice and Bob and can be used to generate ebits by means of an entanglement distillation protocol. Consequently, it holds that

$$Q_2(\Phi, N_s) \geq E_d\left(\mathrm{Id}_A \otimes \Phi(\rho_{AA'})\right) \tag{5.50}$$

for all $N_s \geq 0$ and all $\rho_{A'A}$ satisfying $\mathrm{Tr}[a^\dagger a\, \rho_{A'A}] \leq N_s$, where a denotes the annihilation operator on A'.

If a bipartite state ρ_{AB} is such that the hashing inequality is trivial (i.e. the right-hand side of (5.49) is negative), in order to obtain a non-trivial lower bound on $E_d(\rho_{AB})$, one can adopt a sufficiently large number of iterations of a *recurrence protocol* on ρ_{AB} prior to apply the hashing inequality. In the context of entanglement distillation, the goal of a recurrence protocol is to transform a certain number of copies of the state ρ_{AB} into fewer copies of another state ρ'_{AB} such that $\langle\psi_{00}|\,\rho'_{AB}\,|\psi_{00}\rangle > \langle\psi_{00}|\,\rho_{AB}\,|\psi_{00}\rangle$ [38–40]. Examples of recurrence protocols for qubits can be found in [40, 48, 49], and their generalisations to the case of qudits in [50–52]. In the present paper we will exploit the recently introduced P1-or-P2 recurrence protocol [41]. To achieve a nonzero yield, one may adopt a suitable number of iterations of a recurrence protocol and then apply the hashing or breeding protocol [39, 40]. The latter protocols, which exploit only one-way classical communication, achieves the yield of the hashing inequality in (5.49). Improvements of the hashing and breeding protocols, which exploit two-way classical communication, have been provided in [42]: the two-way distillable entanglement of a convex combination of Bell states $\rho_{AB} := \sum_{ij=0}^{1} \alpha_{ij}\,|\psi_{ij}\rangle\langle\psi_{ij}|$ is lower bounded by

$$Y(\alpha_{00},\alpha_{01},\alpha_{10},\alpha_{11}) := \max\left(0, 1 - H(\{\alpha_{ij}\}) + \tfrac{1}{2}(\alpha_{00}+\alpha_{10})(\alpha_{11}+\alpha_{01})\left[H_2\left(\frac{\alpha_{00}}{\alpha_{00}+\alpha_{10}}\right) + H_2\left(\frac{\alpha_{11}}{\alpha_{01}+\alpha_{11}}\right)\right]\right), \tag{5.51}$$

with $H(\{\alpha_{ij}\}) := -\sum_{m,n=0}^{1} \alpha_{mn} \log_2 \alpha_{mn}$ being the Shannon entropy and $H_2(x) := -x\log_2 x - (1-x)\log_2(1-x)$ for all $x \in [0,1]$ being the binary entropy. The yield in (5.51) is larger than the yield achieved by the hashing protocol, which is $I_{\mathrm{c}}\left(\sum_{ij=0}^{1} \alpha_{ij}\,|\psi_{ij}\rangle\langle\psi_{ij}|\right) = 1 - H(\{\alpha_{ij}\})$. Protocols with larger yiels than (5.51) may be obtained by exploiting the numerical methods introduced in [53].

Now, let us briefly review the definition, the relevant properties, and the known bounds on the two-way capacities of phase-insensitive bosonic Gaussian channels, namely thermal attenuator, thermal amplifier, and additive Gaussian noise.

Thermal attenuator

Let $\mathcal{H}_S$ and $\mathcal{H}_E$ be single-mode systems and let a and b denote their annihilation operators, respectively. For all $\lambda \in [0,1]$ and $\nu \geq 0$, a thermal attenuator $\mathscr{E}_{\lambda,\nu} : \mathfrak{S}(\mathcal{H}_S) \to \mathfrak{S}(\mathcal{H}_S)$ is a quantum channel defined by

$$\mathscr{E}_{\lambda,\nu}(\rho) := \mathrm{Tr}_E\left[U_\lambda^{SE}\left(\rho^S \otimes \tau_\nu^E\right)U_\lambda^{SE\dagger}\right], \tag{5.52}$$

where U_λ^{SE} denotes the unitary operator associated with a beam splitter of transmissivity λ, i.e.

$$U_\lambda^{SE} := \exp\left[\arccos\sqrt{\lambda}\left(a^\dagger b - a\,b^\dagger\right)\right], \tag{5.53}$$

and $\tau_\nu \in \mathfrak{S}(\mathcal{H}_E)$ denotes the thermal state with mean photon number equal to ν. The beam splitter unitary can be expressed via the following disentangling formula [54, Appendix 5]

$$U_\lambda^{SE} = e^{-\sqrt{\frac{1-\lambda}{\lambda}}ab^\dagger} e^{\frac{1}{2}\ln\lambda\left(a^\dagger a - b^\dagger b\right)} e^{\sqrt{\frac{1-\lambda}{\lambda}}a^\dagger b}. \tag{5.54}$$

By writing the quadrature vector $\hat{\mathbf{R}}$ with respect the ordering (S,E), it can be shown that

$$\left(U_\lambda^{SE}\right)^\dagger \hat{\mathbf{R}}\, U_\lambda^{SE} = S_\lambda\, \hat{\mathbf{R}}, \tag{5.55}$$

where

$$S_\lambda := \begin{pmatrix} \sqrt{\lambda}\,\mathbb{1}_2 & \sqrt{1-\lambda}\,\mathbb{1}_2 \\ -\sqrt{1-\lambda}\,\mathbb{1}_2 & \sqrt{\lambda}\,\mathbb{1}_2 \end{pmatrix}. \tag{5.56}$$

This implies that for all $\sigma_{SE} \in \mathfrak{S}(\mathcal{H}_S \otimes H_E)$ it holds that

$$\begin{aligned} \mathbf{m}\left(U_\lambda^{SE}\sigma_{SE}\left(U_\lambda^{SE}\right)^\dagger\right) &= S_\lambda\,\mathbf{m}(\sigma_{SE}), \\ V\left(U_\lambda^{SE}\sigma_{SE}\left(U_\lambda^{SE}\right)^\dagger\right) &= S_\lambda\, V(\sigma_{SE})\, S_\lambda^\intercal. \end{aligned} \tag{5.57}$$

In terms of the annihilation operators a and b, the transformation in (5.55) reads

$$\begin{aligned}
\left(U_\lambda^{SE}\right)^\dagger a\, U_\lambda^{SE} &= \sqrt{\lambda}\, a + \sqrt{1-\lambda}\, b\,, \\
U_\lambda^{SE} a \left(U_\lambda^{SE}\right)^\dagger &= \sqrt{\lambda}\, a - \sqrt{1-\lambda}\, b\,, \\
\left(U_\lambda^{SE}\right)^\dagger b\, U_\lambda^{SE} &= -\sqrt{1-\lambda}\, a + \sqrt{\lambda}\, b\,, \\
U_\lambda^{SE} b \left(U_\lambda^{SE}\right)^\dagger &= \sqrt{1-\lambda}\, a + \sqrt{\lambda}\, b\,.
\end{aligned} \tag{5.58}$$

It can be shown that for any single-mode state ρ it holds that

$$\begin{aligned}
\mathbf{m}\left(\mathscr{E}_{\lambda,\nu}(\rho)\right) &= \sqrt{\lambda}\, \mathbf{m}(\rho)\,, \\
V\left(\mathscr{E}_{\lambda,\nu}(\rho)\right) &= \lambda\, V(\rho) + (1-\lambda)(2\nu+1)\mathbb{1}_2\,,
\end{aligned} \tag{5.59}$$

and, in terms of the characteristic function, for all $\mathbf{r} \in \mathbb{R}^2$ it holds that

$$\chi_{\mathscr{E}_{\lambda,\nu}(\rho)}(\mathbf{r}) = \chi_\rho(\sqrt{\lambda}\mathbf{r}) e^{-\frac{1}{4}(1-\lambda)(2\nu+1)|\mathbf{r}|^2}\,. \tag{5.60}$$

By exploiting (7.45) and the fact that quantum states and characteristic functions are in one-to-one correspondence, for all $\lambda_1, \lambda_2 \in [0,1]$ and $\nu \geq 0$ the following composition rule holds:

$$\mathscr{E}_{\lambda_1,\nu} \circ \mathscr{E}_{\lambda_2,\nu} = \mathscr{E}_{\lambda_1\lambda_2,\nu}\,. \tag{5.61}$$

In Theorem 6 we will provide a simple Kraus representation of the thermal attenuator.

Bounds on two-way capacities of the thermal attenuator

The best known upper bound on the two-way capacities of the thermal attenuator, shown by Pirandola-Laurenza-Ottaviani-Banchi (PLOB) [17], is

$$K(\mathscr{E}_{\lambda,\nu}) \leq \begin{cases} -h(\nu) - \log_2[(1-\lambda)\lambda^\nu], & \text{if } \lambda \in (\frac{\nu}{\nu+1}, 1], \\ 0, & \text{otherwise} \end{cases} \tag{5.62}$$

where

$$h(\nu) \coloneqq (\nu+1)\log_2(\nu+1) - \nu \log_2 \nu \tag{5.63}$$

is the so-called bosonic entropy. The parameter region in which such an upper bound vanishes coincides with the parameter region in which the thermal attenuator $\mathscr{E}_{\lambda,\nu}$ is entanglement breaking, i.e. $\nu \geq 0$ and $\lambda \in [0, \frac{\nu}{\nu+1}]$ [33, 34]. The best known lower bound (before our work) on $Q_2(\mathscr{E}_{\lambda,\nu})$ is given by [25]

$$Q_2(\mathscr{E}_{\lambda,\nu}) \geq \max\{0, -h(\nu) - \log_2(1-\lambda)\}\,. \tag{5.64}$$

Although this is also a lower bound on $K(\mathscr{E}_{\lambda,\nu})$, it is not the best among those currently known. Indeed, an improved lower bound on $K(\mathscr{E}_{\lambda,\nu})$ has been shown by Ottaviani et al. [27]. In the energy-constrained case, the best known lower bound (before our work) on the EC two-way capacities of the thermal attenuator has been found by Noh-Pirandola-Jiang (NPJ) [26], while the best known upper bound is — depending on the parameters λ, ν, and N_s — the bound found by Davis-Shirokov-Wilde (DSW) [15] or the PLOB bound in (5.62).

The lower bound in (5.64) on the two-way capacities of the thermal attenuator can be proved first by applying (5.50) with the choice $\rho_{AA'} = |\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}$, where $|\Psi_{N_s}\rangle$ is the two-mode squeezed vacuum states with local mean photon number equal to N_s defined in (5.39), second by applying the hashing inequality in (5.49), and finally by proving that the reverse coherent information satisfies

$$\lim_{N_s\to\infty} I_{\mathrm{rc}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) = -h(\nu) - \log_2(1-\lambda)\,. \tag{5.65}$$

Analogously, the coherent information

$$I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right)$$

and the reverse coherent information

$$I_{\mathrm{rc}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right)$$

are lower bounds on the EC two-way capacities of the thermal attenuator $\mathscr{E}_{\lambda,\nu}$ with energy constraint equal to N_s:

$$\begin{aligned} Q_2(\mathscr{E}_{\lambda,\nu}, N_s) &\geq \max\left\{I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right), I_{\mathrm{rc}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right)\right\} \\ &= \begin{cases} I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right), & \text{if } N_s \leq \nu, \\ I_{\mathrm{rc}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right), & \text{otherwise.} \end{cases} \end{aligned} \tag{5.66}$$

It holds that [17, 22, 25, 26]

$$\begin{aligned} I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) &= h\left(\lambda N_s + (1-\lambda)\nu\right) - h\left(\frac{D + (1-\lambda)(N_s-\nu) - 1}{2}\right) - h\left(\frac{D - (1-\lambda)(N_s-\nu) - 1}{2}\right), \\ I_{\mathrm{rc}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) &= h(N_s) - h\left(\frac{D + (1-\lambda)(N_s-\nu) - 1}{2}\right) - h\left(\frac{D - (1-\lambda)(N_s-\nu) - 1}{2}\right), \end{aligned} \tag{5.67}$$

where $D := \sqrt{\left((1+\lambda)N_s + (1-\lambda)\nu + 1\right)^2 - 4\lambda N_s(N_s+1)}$. The NPJ lower bound, proved by mixing forward (coherent information) and backward (reverse coherent information) strategies, is [26]

$$Q_2(\mathscr{E}_{\lambda,\nu}, N_s) \geq \sup_{\substack{x\in[0,1],\, N_1,N_2\geq 0 \\ xN_1+(1-x)N_2=N_s}} \left[x\, I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_1}\rangle\langle\Psi_{N_1}|)\right) + (1-x) I_{\mathrm{rc}}\left(\mathrm{Id}_A \otimes \mathscr{E}_{\lambda,\nu}(|\Psi_{N_2}\rangle\langle\Psi_{N_2}|)\right)\right]. \tag{5.68}$$

Fixed λ and ν, if the energy constraint N_s is sufficiently large, the NPJ lower bound is equal to the reverse coherent information bound (i.e. the optimal values of the supremum problem in 5.68 are $x = 0$, $N_1 = 0$, and $N_2 = N_s$).

Thermal amplifier

Let $\mathcal{H}_S$ and $\mathcal{H}_E$ be single-mode systems and let a and b denote their annihilation operators, respectively. For all $g \geq 1$ and $\nu \geq 0$, a thermal amplifier $\Phi_{g,\nu} : \mathfrak{S}(\mathcal{H}_S) \to \mathfrak{S}(\mathcal{H}_S)$ is a quantum channel defined by

$$\Phi_{g,\nu}(\rho) := \mathrm{Tr}_E\left[U_g^{SE}\left(\rho^S \otimes \tau_\nu^E\right)U_g^{SE\dagger}\right], \tag{5.69}$$

where U_g^{SE} denotes the unitary operator associated with two-mode squeezing of parameter g, i.e.

$$U_g^{SE} := \exp\left[\operatorname{arccosh}\sqrt{g}\left(a^\dagger b^\dagger - a\, b\right)\right]. \tag{5.70}$$

The two-mode squeezing unitary can be expressed via the following disentangling formula [54, Appendix 5]

$$U_g^{SE} = e^{\sqrt{\frac{g-1}{g}}a^\dagger b^\dagger} e^{\frac{1}{2}\ln\left(\frac{1}{g}\right)\left(a^\dagger a - b^\dagger b + 1\right)} e^{-\sqrt{\frac{g-1}{g}}ab}. \tag{5.71}$$

By writing the quadrature vector $\hat{\mathbf{R}}$ with respect the ordering (S, E), it can be shown that

$$\left(U_g^{SE}\right)^\dagger \hat{\mathbf{R}}\, U_g^{SE} = S_g\, \hat{\mathbf{R}}, \tag{5.72}$$

where

$$S_g := \begin{pmatrix} \sqrt{g}\,\mathbb{1}_2 & \sqrt{g-1}\,\sigma_z \\ \sqrt{g-1}\,\sigma_z & \sqrt{g}\,\mathbb{1}_2 \end{pmatrix}. \tag{5.73}$$

This implies that for all $\sigma_{SE} \in \mathfrak{S}(\mathcal{H}_S \otimes H_E)$ it holds that

$$\begin{aligned} \mathbf{m}\left(U_g^{SE}\sigma_{SE}\left(U_g^{SE}\right)^\dagger\right) &= S_g\,\mathbf{m}(\sigma_{SE})\,, \\ V\left(U_g^{SE}\sigma_{SE}\left(U_g^{SE}\right)^\dagger\right) &= S_g\,V(\sigma_{SE})\,S_g^\intercal\,. \end{aligned} \tag{5.74}$$

In terms of the annihilation operators a and b, the transformation in (5.72) reads

$$\begin{aligned} \left(U_g^{SE}\right)^\dagger a\,U_g^{SE} &= \sqrt{g}\,a + \sqrt{g-1}\,b^\dagger\,, \\ U_g^{SE}a\,\left(U_g^{SE}\right)^\dagger &= \sqrt{g}\,a - \sqrt{g-1}\,b^\dagger\,, \\ \left(U_g^{SE}\right)^\dagger b\,U_g^{SE} &= \sqrt{g-1}\,a^\dagger + \sqrt{g}\,b\,, \\ U_g^{SE}b\,\left(U_g^{SE}\right)^\dagger &= -\sqrt{g-1}\,a^\dagger + \sqrt{g}\,b\,. \end{aligned} \tag{5.75}$$

It can be shown that for any single-mode state ρ it holds that

$$\begin{aligned} \mathbf{m}\left(\Phi_{g,\nu}(\rho)\right) &= \sqrt{g}\,\mathbf{m}(\rho)\,, \\ V\left(\Phi_{g,\nu}(\rho)\right) &= g\,V(\rho) + (g-1)(2\nu+1)\mathbb{1}_2\,, \end{aligned} \tag{5.76}$$

and, in terms of the characteristic function, for all $\mathbf{r} \in \mathbb{R}^2$ it holds that

$$\chi_{\Phi_{g,\nu}(\rho)}(\mathbf{r}) = \chi_\rho(\sqrt{g}\mathbf{r})e^{-\frac{1}{4}(g-1)(2\nu+1)|\mathbf{r}|^2}\,. \tag{5.77}$$

By exploiting (5.77) and the fact that quantum states and characteristic functions are in one-to-one correspondence, for all $g_1, g_2 \geq 1$ and $\nu \geq 0$ the following composition rule holds:

$$\Phi_{g_1,\nu} \circ \Phi_{g_2,\nu} = \Phi_{g_1 g_2,\nu}\,. \tag{5.78}$$

In Theorem 6 we will provide a simple Kraus representation of the thermal amplifier.

Bounds on two-way capacities of the thermal amplifier

The best known upper bound on the two-way capacities of the thermal amplifier, shown by PLOB [17], is

$$K(\Phi_{g,\nu}) \leq \begin{cases} -h(\nu) + \log_2\left(\frac{g^{\nu+1}}{g-1}\right), & \text{if } g \in [1, 1+\frac{1}{\nu}), \\ 0, & \text{otherwise} \end{cases} \tag{5.79}$$

where $h(\nu)$ is the bosonic entropy defined in (5.63). The parameter region in which such an upper bound vanishes coincides with the parameter region in which the thermal amplifier $\Phi_{g,\nu}$ is entanglement breaking, i.e. $\nu \geq 0$ and $g \geq 1 + \frac{1}{\nu}$ [33, 34]. The best known lower bound (before our work) on $Q_2(\Phi_{\lambda,\nu})$ is given by [25]

$$Q_2(\Phi_{g,\nu}) \geq \max\left\{0, -h(\nu) + \log_2\left(\frac{g}{g-1}\right)\right\}\,, \tag{5.80}$$

which can be proved, analogously as it has been done in (5.65), by showing that the coherent information satisfies

$$\lim_{N_s\to\infty} I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \Phi_{g,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) = -h(\nu) + \log_2\left(\frac{g}{g-1}\right)\,. \tag{5.81}$$

The best known lower bound on the secret-key capacity $K(\Phi_{g,\nu})$ has been shown by Wong-Ottaviani-Guo-Pirandola (WOGP) [36]. In the energy-constrained scenario, the best known lower bound is the NPJ bound [26], which is given by

$$Q_2(\Phi_{g,\nu}, N_s) \geq \sup_{x\in[0,1]} x\,I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \Phi_{g,\nu}(|\Psi_{\frac{N_s}{x}}\rangle\langle\Psi_{\frac{N_s}{x}}|)\right)\,, \tag{5.82}$$

where [17, 22, 25, 26]

$$
\begin{aligned}
&I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \Phi_{g,\nu}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) \\
&= h\left(gN_s + (g-1)(\nu+1)\right) - h\left(\frac{D' + (g-1)(N_s+\nu+1) - 1}{2}\right) - h\left(\frac{D' - (g-1)(N_s+\nu+1) - 1}{2}\right),
\end{aligned}
\tag{5.83}
$$

with $D' := \sqrt{\left((g+1)N_s + (g-1)(\nu+1) + 1\right)^2 - 4g\,N_s(N_s+1)}$. Fixed g and ν, if the energy constraint N_s is sufficiently large, the NPJ lower bound is equal to the coherent information bound (i.e. the optimal value of the supremum problem in 5.82 is $x = 1$).

Additive Gaussian noise

Let $\mathcal{H}_S$ be a single-mode system and let $\{D_{\mathbf{r}}\}_{\mathbf{r}\in\mathbb{R}^2}$ be its dispacement operators. For all $\xi \geq 0$, the additive Gaussian noise $\Lambda_\xi : \mathfrak{S}(\mathcal{H}_S) \to \mathfrak{S}(\mathcal{H}_S)$ is a quantum channel defined by

$$
\Lambda_\xi(\rho) := \frac{1}{2\pi\xi} \int_{\mathbb{R}^2} \mathrm{d}^2\mathbf{r}\, e^{-\frac{1}{2\xi}\mathbf{r}^{\intercal}\mathbf{r}} D_{\mathbf{r}} \rho D_{\mathbf{r}}^\dagger \,. \tag{5.84}
$$

By using that $\hat{\mathbf{R}} = (\hat{x}, \hat{p})$, $a = \frac{\hat{x}+\hat{p}}{\sqrt{2}}$, and by defining $\mathbf{r} := (x,p)^{\mathrm{T}}$, $z := \frac{x+ip}{\sqrt{2}}$, and

$$
D(z) := \exp\left[z a^\dagger - z^* a\right] = D_{-\mathbf{r}} \,, \tag{5.85}
$$

the additive Gaussian noise can be expressed in the following equivalent form:

$$
\Lambda_\xi(\rho) = \frac{1}{\pi\xi} \int_{\mathbb{C}} \mathrm{d}^2 z\, e^{-\frac{|z|^2}{\xi}} D(z)\, \rho\, D(z)^\dagger \,, \tag{5.86}
$$

where we have used that $\mathrm{d}^2\mathbf{r} = \mathrm{d}x\,\mathrm{d}p = \frac{\mathrm{dRe}(z)\,\mathrm{dIm}(z)}{2} = \frac{\mathrm{d}^2 z}{2}$ and we have performed the integral variable substitution $z \to -z$. It can be shown that for all single-mode states ρ it holds that

$$
\begin{aligned}
&\mathbf{m}\left(\Lambda_\xi(\rho)\right) = \mathbf{m}(\rho) \,, \\
&V\left(\Lambda_\xi(\rho)\right) = V(\rho) + 2\xi\, \mathbb{1}_2 \,.
\end{aligned}
\tag{5.87}
$$

and, in terms of the characteristic function, for all $\mathbf{r} \in \mathbb{R}^2$ it holds that

$$
\chi_{\Lambda_\xi(\rho)}(\mathbf{r}) = \chi_\rho(\mathbf{r}) e^{-\frac{1}{2}\xi|\mathbf{r}|^2} \,. \tag{5.88}
$$

In Theorem 6 we will provide a simple Kraus representation of the additive Gaussian noise.

Additive Gaussian noise as the strong limit of thermal attenuator or thermal amplifier

For completeness, let us remark that the Additive Gaussian noise Λ_ξ is the strong limit of the thermal attenuator $\mathscr{E}_{1-\frac{\xi}{\nu},\nu}$ and thermal amplifier $\Phi_{1+\frac{\xi}{\nu},\nu}$ for $\nu \to \infty$, i.e. it holds that

$$
\begin{aligned}
&\lim_{\nu\to\infty} \|\mathscr{E}_{1-\frac{\xi}{\nu},\nu}(\rho) - \Lambda_\xi(\rho)\|_1 = 0 \,, \\
&\lim_{\nu\to\infty} \|\Phi_{1+\frac{\xi}{\nu},\nu}(\rho) - \Lambda_\xi(\rho)\|_1 = 0 \,,
\end{aligned}
\tag{5.89}
$$

for any single-mode state ρ. Indeed, (7.45), (5.77), and (5.88) imply that for any single-mode state ρ and any $\mathbf{r} \in \mathbb{R}^2$ it holds that

$$
\begin{aligned}
&\lim_{\nu\to\infty} \chi_{\mathscr{E}_{1-\frac{\xi}{\nu},\nu}(\rho)}(\mathbf{r}) = \chi_{\Lambda_\xi(\rho)}(\mathbf{r}) \,, \\
&\lim_{\nu\to\infty} \chi_{\Phi_{1+\frac{\xi}{\nu},\nu}(\rho)}(\mathbf{r}) = \chi_{\Lambda_\xi(\rho)}(\mathbf{r}) \,.
\end{aligned}
\tag{5.90}
$$

Consequently, by exploiting the fact that a sequence of states $\{\sigma_k\}_{k\in\mathbb{N}} \subseteq \mathfrak{S}(L^2(\mathbb{R}))$ converges in trace norm to a quantum state $\sigma \in \mathfrak{S}(L^2(\mathbb{R}))$ if and only if the sequence of characteristic functions $\{\chi_{\sigma_k}(\mathbf{r})\}_{k\in\mathbb{N}}$ converges pointwise to the characteristic function $\chi_\sigma(\mathbf{r})$ [55, Theorem 2], the thermal attenuator $\mathscr{E}_{1-\frac{\xi}{\nu},\nu}$ and thermal amplifier $\Phi_{1+\frac{\xi}{\nu},\nu}$ strongly converge to the additive Gaussian noise Λ_ξ for $\nu \to \infty$.

Bounds on two-way capacities of the additive Gaussian noise

The best known upper bound on the two-way capacities of the additive Gaussian noise, shown by PLOB [17], is

$$K(\Lambda_\xi) \leq \begin{cases} \frac{\xi-1}{\ln 2} - \log_2(\xi), & \text{if } \xi < 1, \\ 0, & \text{otherwise} \end{cases} \tag{5.91}$$

where $h(\nu)$ is the bosonic entropy defined in (5.63). The parameter region in which such an upper bound vanishes coincides with the parameter region in which the Additive Gaussian noise Λ_ξ is entanglement breaking, i.e. $\xi \geq 1$ [33, 34]. The best known lower bound (before our work) on $Q_2(\Lambda_\xi)$ is given by [25]

$$Q_2(\Lambda_\xi) \geq \max\{0, -\log_2(e\,\xi)\}\,, \tag{5.92}$$

which can be proved, analogously as it has been done in (5.65), by showing that the coherent information satisfies

$$\lim_{N_s\to\infty} I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \Lambda_\xi(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'})\right) = \log_2(e\,\xi)\,. \tag{5.93}$$

In the energy-constrained scenario, the best known lower bound is the NPJ bound [26], which is given by

$$Q_2(\Lambda_\xi, N_s) \geq \sup_{x\in[0,1]} x\, I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \Lambda_\xi(|\Psi_{\frac{N_s}{x}}\rangle\langle\Psi_{\frac{N_s}{x}}|)\right)\,, \tag{5.94}$$

where [17, 22, 25, 26]

$$I_{\mathrm{c}}\left(\mathrm{Id}_A \otimes \Lambda_\xi(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) = h\left(N_s + \xi\right) - h\left(\frac{D'' + \xi - 1}{2}\right) - h\left(\frac{D'' - \xi - 1}{2}\right)\,, \tag{5.95}$$

with $D'' := \sqrt{(2N_s + \xi + 1)^2 - 4N_s(N_s+1)}$. Fixed ξ, if the energy constraint N_s is sufficiently large, the NPJ lower bound is equal to the coherent information bound (i.e. the optimal value of the supremum problem in 5.94 is $x = 1$).

Action of phase-insensitive bosonic Gaussian channels on generic operators

In this section we establish properties of the channel composition between pure amplifier channel and pure loss channel.

Definition 3. For all $\lambda \in [0,1]$ and $g \geq 1$ let us define the channel $\mathscr{N}_{g,\lambda}$ as the composition between pure amplifier channel $\Phi_{g,0}$ and pure loss channel $\mathscr{E}_{\lambda,0}$, i.e.

$$\mathscr{N}_{g,\lambda} := \Phi_{g,0} \circ \mathscr{E}_{\lambda,0}\,. \tag{5.96}$$

Lemma 4. *The channel $\mathscr{N}_{g,\lambda}$ is entanglement breaking if and only if $(1-\lambda)g \geq 1$.*

Proof. First, let us determine the parameter region of g and λ where the channel $\mathscr{N}_{g,\lambda}$ is entanglement breaking. Since $\mathscr{N}_{g,\lambda}$ is a Gaussian channel, we can apply Lemma 3. By using (5.102), one can show that $\mathscr{N}_{g,\lambda}$ transforms the first moment and the covariance matrix as

$$\begin{aligned} &\mathbf{m}\left(\mathscr{N}_{g,\lambda}(\rho)\right) = \sqrt{g\lambda}\,\mathbf{m}(\rho)\,, \\ &V\left(\mathscr{N}_{g,\lambda}(\rho)\right) = g\lambda\, V(\rho) + (2g - 1 - g\lambda)\,\mathbb{1}_2\,, \end{aligned} \tag{5.97}$$

for all quantum states ρ. Hence, Lemma 3 establishes that Φ is entanglement breaking if and only if there exists $\alpha, \gamma \in \mathbb{R}^{2\times 2}$ with $\alpha \geq i\,\Omega_1$ and $\gamma \geq i\lambda g\,\Omega_1$ such that

$$(2g - 1 - g\lambda)\,\mathbb{1}_2 = \alpha + \gamma\,. \tag{5.98}$$

The condition in (5.98) is equivalent to

$$(2g-1-g\lambda)\,\mathbb{1}_2 \geq i\,(1+\lambda g)\,\Omega_1\,. \tag{5.99}$$

Indeed, if the condition in (5.98) is satisfied, then $(2g-1-g\lambda)\,\mathbb{1}_2 = \alpha+\gamma \geq i\,(1+\lambda g)\,\Omega_1$, i.e. also the condition in (5.99) is satisfied. Conversely, assume that the condition in (5.99) is satisfied. Then, the fact that

$$x\,\mathbb{1}_2 \geq i\,\Omega_1 \quad \text{if and only if} \quad x \geq 1\,, \tag{5.100}$$

implies that $(1-\lambda)g \geq 1$. Consequently, by choosing $\alpha := \mathbb{1}_2$ and $\gamma := (2g-2-g\lambda)\mathbb{1}_2$ and by using (5.100), it holds that the condition in (5.98) is satisfied with $\alpha \geq i\,\Omega_1$ and $\gamma \geq i\lambda g\,\Omega_1$. By exploiting (5.100), we deduce that $\mathcal{N}_{g,\lambda}$ is entanglement breaking if and only if $(1-\lambda)g \geq 1$. □

Lemma 5. *Let $\nu \geq 0$, $\lambda \in [0,1]$, $g \geq 1$, and $\xi \geq 0$. The thermal attenuator $\mathscr{E}_{\lambda,\nu}$, the thermal amplifier $\Phi_{g,\nu}$, and the additive Gaussian noise Λ_ξ can be expressed in terms of the composition between pure amplifier channel and pure loss channel as*

$$\begin{aligned}
\mathscr{E}_{\lambda,\nu} &= \mathcal{N}_{1+(1-\lambda)\nu,\,\frac{\lambda}{1+(1-\lambda)\nu}}\,,\\
\Phi_{g,\nu} &= \mathcal{N}_{g+(g-1)\nu,\,\frac{g}{g+(g-1)\nu}}\,,\\
\Lambda_\xi &= \mathcal{N}_{1+\xi,\,\frac{1}{1+\xi}}\,.
\end{aligned} \tag{5.101}$$

Proof. Let ρ be a single-mode state. The characteristic function of $\mathcal{N}_{g,\lambda}(\rho)$ is

$$\chi_{\mathcal{N}_{g,\lambda}(\rho)}(\mathbf{r}) = \chi_{\Phi_{g,0}\circ\mathscr{E}_{\lambda,0}(\rho)}(\mathbf{r}) = \chi_{\mathscr{E}_{\lambda,0}(\rho)}(\sqrt{g}\mathbf{r})e^{-\frac{1}{4}(g-1)|\mathbf{r}|^2} = \chi_\rho(\sqrt{g\lambda}\,\mathbf{r})e^{-\frac{1}{4}[2g-1-g\lambda]|\mathbf{r}|^2} \tag{5.102}$$

for all $\mathbf{r} \in \mathbb{R}^2$, where we have used (7.45) and (5.77). Consequently, by exploiting (7.45), (5.77), and (5.88), one can check that for all $\mathbf{r} \in \mathbb{R}^2$ it holds that

$$\begin{aligned}
\chi_{\mathscr{E}_{\lambda,\nu}(\rho)}(\mathbf{r}) &= \chi_{\mathcal{N}_{1+(1-\lambda)\nu,\,\frac{\lambda}{1+(1-\lambda)\nu}}(\rho)}(\mathbf{r})\,,\\
\chi_{\Phi_{g,\nu}(\rho)}(\mathbf{r}) &= \chi_{\mathcal{N}_{g+(g-1)\nu,\,\frac{g}{g+(g-1)\nu}}(\rho)}(\mathbf{r})\,,\\
\chi_{\Lambda_\xi(\rho)}(\mathbf{r}) &= \chi_{\mathcal{N}_{1+\xi,\,\frac{1}{1+\xi}}(\rho)}(\mathbf{r})\,.
\end{aligned} \tag{5.103}$$

Hence, by exploiting the fact that quantum states and characteristic functions are in one-to-one correspondence, (5.101) is proved. □

The forthcoming Theorem 6 provides a simple Kraus representation of $\mathcal{N}_{g,\lambda}$ and allows one to easily calculate the output of $\mathcal{N}_{g,\lambda}$ for generic input operators.

Theorem 6. *Let $\lambda \in [0,1]$ and $g \geq 1$. The quantum channel $\mathcal{N}_{g,\lambda}$, defined in Definition (3), admits the following Kraus representation:*

$$\mathcal{N}_{g,\lambda}(\rho) = \sum_{k,m=0}^{\infty} M_{k,m}^{(comp)}(g,\lambda)\,\rho\left(M_{k,m}^{(comp)}(g,\lambda)\right)^\dagger\,, \tag{5.104}$$

where

$$M_{k,m}^{(comp)}(g,\lambda) := M_k^{(pure\ amp)}(g)\,M_m^{(pure\ loss)}(\lambda) = \sqrt{\frac{(g-1)^k\,(1-\lambda)^m}{k!\,m!\,g^{k+1}}}(a^\dagger)^k\left(\sqrt{\frac{\lambda}{g}}\right)^{a^\dagger a}a^m \tag{5.105}$$

and where we have introduced the Kraus operators of pure loss channel and pure amplifier channel:

$$M_k^{(pure\ amp)}(g) := \frac{1}{\sqrt{g\,k!}}\left(\sqrt{\frac{g-1}{g}}\right)^k(a^\dagger)^k\left(\frac{1}{\sqrt{g}}\right)^{a^\dagger a}\,, \tag{5.106}$$

$$M_m^{(pure\ loss)}(\lambda) := \sqrt{\frac{(1-\lambda)^m}{m!}}(\sqrt{\lambda})^{a^\dagger a}\,a^m\,. \tag{5.107}$$

In particular, by letting $|n\rangle$ and $|i\rangle$ two Fock states, it holds that

$$\mathscr{N}_{g,\lambda}(|n\rangle\langle i|) = \sum_{l=\max(i-n,0)}^{\infty} f_{n,i,l}(g,\lambda)\,|l+n-i\rangle\langle l|\,. \tag{5.108}$$

where

$$f_{n,i,l}(g,\lambda) := \sum_{m=\max(i-l,0)}^{\min(n,i)} \frac{\sqrt{n!i!l!(l+n-i)!}}{(n-m)!(i-m)!m!(l+m-i)!} \frac{(g-1)^{l+m-i}(1-\lambda)^m \lambda^{\frac{n+i-2m}{2}}}{g^{l+1+\frac{n-i}{2}}}\,. \tag{5.109}$$

Proof. By using (7.42), the pure loss channel can be written as

$$\mathscr{E}_{\lambda,0}(\rho) = \sum_{m=0}^{\infty} M_m^{(\text{pure loss})}(\lambda)\,\rho\left(M_m^{(\text{pure loss})}(\lambda)\right)^\dagger\,, \tag{5.110}$$

where for all $m \in \mathbb{N}$ the Kraus operator $M_m^{(\text{pure loss})}(\lambda)$ is

$$M_m^{(\text{pure loss})}(\lambda) := (-1)^m \langle m|_E\, U_\lambda^{SE}\, |0\rangle_E\,. \tag{5.111}$$

Hence, by using the disentangling formula for beam splitter unitary [54, Appendix 5]

$$U_\lambda^{SE} = e^{-\sqrt{\frac{1-\lambda}{\lambda}}ab^\dagger} e^{\frac{1}{2}\ln\lambda\left(a^\dagger a - b^\dagger b\right)} e^{\sqrt{\frac{1-\lambda}{\lambda}}a^\dagger b} \tag{5.112}$$

and the fact that

$$e^{-\frac{1}{2}\ln\lambda\, a^\dagger a}\, a^m\, e^{\frac{1}{2}\ln\lambda\, a^\dagger a} = \lambda^{m/2} a\,, \tag{5.113}$$

it holds that

$$M_m^{(\text{pure loss})}(\lambda) = \frac{1}{\sqrt{m!}}\left(\sqrt{\frac{1-\lambda}{\lambda}}\right)^m a^m e^{\frac{1}{2}\ln\lambda\, a^\dagger a} = \sqrt{\frac{(1-\lambda)^m}{m!}}(\sqrt{\lambda})^{a^\dagger a}\, a^m\,. \tag{5.114}$$

By using (5.69), the pure amplifier channel can be written as

$$\Phi_{g,0}(\rho) = \sum_{k=0}^{\infty} M_k^{(\text{pure amp})}(g)\,\rho\left(M_k^{(\text{pure amp})}(g)\right)^\dagger\,, \tag{5.115}$$

where for all $k \in \mathbb{N}$ the Kraus operator $M_k^{(\text{pure amp})}(g)$ is

$$M_k^{(\text{pure amp})}(g) := \langle k|_E\, U_g^{SE}\, |0\rangle_E\,. \tag{5.116}$$

Hence, by using the disentangling formula for the two-mode squeezing unitary [54, Appendix 5]

$$U_g^{SE} = e^{\sqrt{\frac{g-1}{g}}a^\dagger b^\dagger} e^{\frac{1}{2}\ln\left(\frac{1}{g}\right)\left(a^\dagger a - b^\dagger b + 1\right)} e^{-\sqrt{\frac{g-1}{g}}ab}\,, \tag{5.117}$$

it holds that

$$M_k^{(\text{pure amp})}(g) = \frac{1}{\sqrt{g\,k!}}\left(\sqrt{\frac{g-1}{g}}\right)^k (a^\dagger)^k \left(\frac{1}{\sqrt{g}}\right)^{a^\dagger a}\,. \tag{5.118}$$

By using (5.110), (5.115), and the fact $\mathscr{N}_{g,\lambda} = \Phi_{g,0} \circ \mathscr{E}_{\lambda,0}$, (5.104) is proved. Now, let us calculate $\mathscr{N}_{g,\lambda}(|n\rangle\langle i|) = \sum_{m,n=0}^{\infty} M_{k,m}^{(\text{comp})}(g,\lambda)\,|n\rangle\langle i|\,(M_{k,m}^{(\text{comp})}(g,\lambda))^\dagger$ in order to prove (5.108). By exploiting the following formulae

$$\begin{aligned} a^m\,|n\rangle &= \begin{cases} \sqrt{\frac{n!}{(n-m)!}}\,|n-m\rangle\,, & \text{if } n \geq m, \\ 0, & \text{otherwise} \end{cases} \\ \left(a^\dagger\right)^k |n-m\rangle &= \sqrt{\frac{(n-m+k)!}{(n-m)!}}\,|n-m+k\rangle\,, \end{aligned} \tag{5.119}$$

for $m > n$ it holds that $M_{k,m}\,|n\rangle = 0$, otherwise for $m \leq n$ it holds that

$$M_{k,m}^{(comp)}(g,\lambda)\,|n\rangle = \frac{1}{(n-m)!}\sqrt{\frac{n!(n-m+k)!}{k!m!}}\sqrt{\frac{(g-1)^k(1-\lambda)^m\lambda^{n-m}}{g^{k+1+n-m}}}\,|n-m+k\rangle\ . \tag{5.120}$$

Consequently, we conclude that

$$\begin{aligned}\mathscr{N}_{g,\lambda}(|n\rangle\langle i|) &= \sum_{k=0}^{\infty}\sum_{m=0}^{\min(n,i)}\frac{\sqrt{n!(n-m+k)!i!(i-m+k)!}}{(n-m)!(i-m)!k!m!}\sqrt{\frac{(g-1)^{2k}(1-\lambda)^{2m}\lambda^{n+i-2m}}{g^{2k+2+n+i-2m}}}\,|n-m+k\rangle\langle i-m+k| \\ &= \sum_{l=\max(i-n,0)}^{\infty} f_{n,i,l}(g,\lambda)\,|l+n-i\rangle\langle l|\ .\end{aligned} \tag{5.121}$$

Hence, (5.108) is proved. □

Calculating the action of Gaussian channels on non-Gaussian states is cumbersome in general. The forthcoming Theorem 7 overcomes this difficulty and allows one to easily calculate the output of all piBGCs for generic input operators.

Theorem 7. *Let $\nu \geq 0$, $\lambda \in [0,1]$, $g \geq 1$, and $\xi \geq 0$. The thermal attenuator $\mathscr{C}_{\lambda,\nu}$, the thermal amplifier $\Phi_{g,\nu}$, and the additive Gaussian noise Λ_ξ admit the following Kraus representations:*

$$\mathscr{C}_{\lambda,\nu}(\rho) = \sum_{k,m=0}^{\infty} M_{k,m}^{(att)}(\lambda,\nu)\,\rho\left(M_{k,m}^{(att)}(\lambda,\nu)\right)^\dagger\ , \tag{5.122}$$

$$\Phi_{g,\nu}(\rho) = \sum_{k,m=0}^{\infty} M_{k,m}^{(amp)}(g,\nu)\,\rho\left(M_{k,m}^{(amp)}(g,\nu)\right)^\dagger\ , \tag{5.123}$$

$$\Lambda_{\xi}(\rho) = \sum_{k,m=0}^{\infty} M_{k,m}^{(add)}(\xi)\,\rho\left(M_{k,m}^{(add)}(\xi)\right)^\dagger\ , \tag{5.124}$$

where

$$M_{k,m}^{(att)}(\lambda,\nu) := M_{k,m}^{(comp)}\left(1+(1-\lambda)\nu, \frac{\lambda}{1+(1-\lambda)\nu}\right)\ , \tag{5.125}$$

$$M_{k,m}^{(amp)}(g,\nu) := M_{k,m}^{(comp)}\left(g+(g-1)\nu, \frac{g}{g+(g-1)\nu}\right)\ , \tag{5.126}$$

$$M_{k,m}^{(add)}(\xi) := M_{k,m}^{(comp)}\left(1+\xi, \frac{1}{1+\xi}\right)\ , \tag{5.127}$$

with $M_{k,m}^{(comp)}$ being defined in (5.105). In particular, by letting $|n\rangle$ and $|i\rangle$ two Fock states, it holds that

$$\mathscr{C}_{\lambda,\nu}(|n\rangle\langle i|) = \sum_{l=\max(i-n,0)}^{\infty} f_{n,i,l}\left(1+(1-\lambda)\nu, \frac{\lambda}{1+(1-\lambda)\nu}\right)\,|l+n-i\rangle\langle l|\ , \tag{5.128}$$

$$\Phi_{g,\nu}(|n\rangle\langle i|) = \sum_{l=\max(i-n,0)}^{\infty} f_{n,i,l}\left(g+(g-1)\nu, \frac{g}{g+(g-1)\nu}\right)\,|l+n-i\rangle\langle l|\ , \tag{5.129}$$

$$\Lambda_{\xi}(|n\rangle\langle i|) = \sum_{l=\max(i-n,0)}^{\infty} f_{n,i,l}\left(1+\xi, \frac{1}{1+\xi}\right)\,|l+n-i\rangle\langle l|\ , \tag{5.130}$$

with $f_{n,i,l}$ being defined in (5.109).

Proof. Theorem 7 is a direct consequence of Lemma 5 and Theorem 6. □

We observe here that the Kraus representation in (5.125) of the thermal attenuator is precisely the one obtained in [56] via the "master equation trick".

5.3.2 Results

In this section we expound our results. In subsection 5.3.2 first we prove preliminary results on the two-way capacities of generic quantum channels and second we apply them to the composition between pure amplifier channel and pure loss channel. In subsection 5.3.2 we specialise these results to the case of piBGCs (thermal attenuator, thermal amplifier, and additive Gaussian noise) and we find the following two main results:

- The parameter regions where the (EC) two-way capacities of piBGCs are strictly positive are precisely those where these channels are not entanglement breaking;
- We find a new lower bound on the secret-key and two-way quantum capacity of piBGCs, which constitutes a significant improvement with respect the state-of-the-art lower bounds [12, 25–27, 36] in many parameter regions.

Preliminary results

Let us begin by introducing the concept of a *generalized Choi state* of a quantum channel.

Definition 8 Generalised Choi state of a quantum channel [57, 58]**.** *Let $\mathcal{H}_B$ be a possibly infinite-dimensional Hilbert space. Let $\mathcal{H}_A, \mathcal{H}_{A'}$ be isomorphic (possibly infinite dimensional) Hilbert spaces. Let $|\psi\rangle_{A'A}$ be a pure state of the form*

$$|\psi\rangle_{A'A} = \sum_i \sqrt{\lambda_i}\, |e_i\rangle_A \otimes |e_i\rangle_{A'} \,, \tag{5.131}$$

where $(\lambda_i)_i$ are strictly positive numbers such that $\sum_i \lambda_i = 1$, and $(|e_i\rangle_A)_i$ and $(|e_i\rangle_{A'})_i$ form an orthonormal basis of $\mathcal{H}_A$ and $\mathcal{H}_{A'}$, respectively. Let $\Phi_{A'\to B}$ be a quantum channel from $\mathcal{H}_{A'}$ to $\mathcal{H}_B$. Then, the state

$$C_{AB} \coloneqq \mathrm{Id}_A \otimes \Phi_{A'\to B}(|\psi\rangle\langle\psi|_{AA'}) \tag{5.132}$$

is said to be a generalised Choi state of Φ.

In finite dimensions, if the state $|\psi\rangle$ in (5.131) is a *maximally entangled state*, then the state $C_{AB} \coloneqq \mathrm{Id}_A \otimes \Phi_{A'\to B}(|\psi\rangle\langle\psi|_{AA'})$ is simply referred to as the Choi state of the channel Φ. Additionally, it is well known that a quantum channel is completely characterised by its Choi state [14]. The following lemma extends this result, showing that a quantum channel can also be completely characterised in terms of its *generalised* Choi state.

Lemma 9. *Let $\Phi_{A'\to B}$ be a quantum channel and let be C_{AB} a generalised Choi state. Then, it holds that*

$$\Phi_{A'\to B}(X_{A'}) = \mathrm{Tr}_A\Big[(X_A \otimes \mathbb{1}_B)\,(D_A \otimes \mathbb{1}_B)\, C_{AB}^{t_A}\,(D_A \otimes \mathbb{1}_B)\Big] \quad \text{for all linear operators } X_{A'}\,, \tag{5.133}$$

where, by using the notation introduced in Definition 69, $D_A \coloneqq \sum_i \lambda_i^{-1/2} |e_i\rangle\langle e_i|_A$ and t_A is the partial transpose on A, i.e. $(|e_i\rangle\langle e_j|_A)^{t_A} = |e_j\rangle\langle e_i|_A$ for all i, j.

Proof. By exploiting that $(|e_i\rangle_A)_i$ are orthonormal, we have that

$$\Phi_{A'\to B}\Big(|e_i\rangle\langle e_j|_{A'}\Big) = \frac{1}{\sqrt{\lambda_i\lambda_j}}\,\mathrm{Tr}_A[(|e_j\rangle\langle e_i|_A \otimes \mathbb{1}_B)\; C_{AB}]\,, \quad \forall\, i, j\,. \tag{5.134}$$

By writing

$$X_{A'} = \sum_{ij} \langle e_i|\, X\, |e_j\rangle\, |e_i\rangle\langle e_j|_{A'} \tag{5.135}$$

and by exploiting the linearity of $\Phi_{A'\to B}$, it thus follows that

$$\begin{aligned}\Phi_{A'\to B}(X_{A'}) &= \mathrm{Tr}_A\Big[\Big((D_A X_A D_A)^{t_A}\otimes \mathbb{1}_B\Big)\, C_{AB}\Big]\\ &= \mathrm{Tr}_A\Big[(D_A X_A D_A\otimes \mathbb{1}_B)\, C_{AB}^{t_A}\Big]\\ &= \mathrm{Tr}_A\Big[(X_A\otimes \mathbb{1}_B)\,(D_A\otimes \mathbb{1}_B)\, C_{AB}^{t_A}\,(D_A\otimes \mathbb{1}_B)\Big]\,.\end{aligned} \tag{5.136}$$

□

In finite dimensions, it is well known that a quantum channel is entanglement breaking if and only if its Choi state is separable. The following lemma generalises this result, demonstrating that a quantum channel is entanglement breaking if and only if its *generalised* Choi state is separable.

Lemma 10. *Let Φ be a quantum channel and let C_{AB} be a generalised Choi state of Φ. Then, Φ is entanglement breaking if and only if C_{AB} is separable.*

Proof. By definition, if Φ is entanglement breaking, then $\mathrm{Id}_R\otimes\Phi_{A'\to B}(\rho_{RA'})$ is separable for all bipartite states $\rho_{RA'}$. In particular, any generalised Choi state of an entanglement breaking channel is separable.

Conversely, let us assume that the generalised Choi state C_{AB} is separable, that is there exists a probability distribution p_x and states $(\rho_A^{(x)})_x, (\sigma_B^{(x)})_x$ such that

$$C_{AB} = \sum_x p_x\, \rho_A^{(x)}\otimes\sigma_B^{(x)}\,. \tag{5.137}$$

Let us show that Φ is entanglement breaking. To this end let us consider an arbitrary bipartite state $\rho_{RA'}$ and let us show that $\mathrm{Id}_R\otimes\Phi_{A'\to B}(\rho_{RA'})$ is separable. By exploiting Lemma 9, it holds that

$$\begin{aligned}\mathrm{Id}_R\otimes\Phi_{A'\to B}(\rho_{RA'}) &= \mathrm{Tr}_A\Big[(\rho_{RA}\otimes\mathbb{1}_B)\,\Big(\mathbb{1}_R\otimes(D_A\otimes\mathbb{1}_B)\,C_{AB}^{t_A}\,(D_A\otimes\mathbb{1}_B)\Big)\Big]\\ &= \sum_x p_x\ \mathrm{Tr}_A\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big]\otimes\sigma_B^{(x)}\,.\end{aligned} \tag{5.138}$$

Since the operator $D_A\,(\rho_A^{(x)})^{t_A}\,D_A$ is positive semidefinite, we can write its spectral decomposition as

$$D_A\,(\rho_A^{(x)})^{t_A}\,D_A = \sum_i \eta_i^{(x)}\,|\phi_i^{(x)}\rangle\langle\phi_i^{(x)}| \tag{5.139}$$

with the eigenvalues $(\eta_i^{(x)})_i$ being positive. This implies that the operator $\mathrm{Tr}_A\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big]$ is positive semidefinite, as it can be written as

$$\mathrm{Tr}_A\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big] = \sum_i \eta_i^{(x)}\,\langle\phi_i^{(x)}|_A\,\rho_{RA}\,|\phi_i^{(x)}\rangle_A\,. \tag{5.140}$$

and $\langle\phi_i^{(x)}|_A\,\rho_{RA}\,|\phi_i^{(x)}\rangle_A$ is positive semidefinite. In particular, the trace of the operator

$$\mathrm{Tr}_A\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big]$$

vanishes if and only if it is the zero operator. Consequently, (5.138) implies that

$$\mathrm{Id}_R\otimes\Phi_{A'\to B}(\rho_{RA'}) = \sum_{x:\,q_x\neq 0} q_x\,\omega_R^{(x)}\otimes\sigma_B^{(x)}\,, \tag{5.141}$$

where we defined

$$\begin{aligned}q_x &:= p_x\,\mathrm{Tr}_{RA}\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big]\,,\\ \omega_R^{(x)} &:= \frac{\mathrm{Tr}_A\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big]}{\mathrm{Tr}_{RA}\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big]}\,.\end{aligned} \tag{5.142}$$

The fact that the operator $\mathrm{Tr}_A\Big[\rho_{RA}\,\Big(\mathbb{1}_R\otimes D_A\,(\rho_A^{(x)})^{t_A}\,D_A\Big)\Big]$ is positive semidefinite implies that $(q_x)_x$ is a probability distribution and that $\omega_R^{(x)}$ is a quantum state. Hence, we conclude that $\mathrm{Id}_R\otimes\Phi_{A'\to B}(\rho_{RA'})$ is separable. □

The following theorem establishes that the energy-constrained two-way capacities of a single-mode Gaussian channel are strictly positive if and only if the channel is not entanglement breaking.

Theorem 11. *Let Φ be a single-mode Gaussian channel and let $N_s > 0$. The energy-constrained two-way quantum capacity $Q_2(\Phi, N_s)$ and secret-key capacity $K(\Phi, N_s)$ are strictly positive if and only if Φ is not entanglement breaking.*

Proof. Since any entanglement-breaking channel has vanishing two-way capacities [14], it suffices to consider the case where Φ is not entanglement breaking. Assume that Alice prepares many copies of the two-mode squeezed vacuum state $|\Psi_{N_s}\rangle_{AA'}$ with mean local photon number N_s and sends the systems A' through the Gaussian channel $\Phi_{A'\to B}$. Now Alice and Bob share many copies of the two-mode Gaussian state $C_{N_s} \coloneqq \Big(\mathrm{Id}_A\otimes\Phi_{A'\to B}\Big)(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'})$, which is a generalised Choi state of Φ [2, 57]. As such, C_{N_s} is entangled, as established by Lemma 10. By exploiting the fact that a two-mode Gaussian state is entangled if and only if it is not PPT [2, 30, 31], it thus follows that C_{N_s} is not PPT. Since any two-mode Gaussian state that it is not PPT is also *distillable* [32] — i.e. it can be converted into ebits with a strictly positive rate — we conclude that $K(\Phi, N_s) \geq Q_2(\Phi, N_s) > 0$. □

In the following, we provide an alternative, more explicit proof of the above result. We start by proving the following lemma.

Lemma 12. *Let $\Phi : \mathfrak{S}(L^2(\mathbb{R})) \to \mathfrak{S}(L^2(\mathbb{R}))$ be a single-mode Gaussian quantum channel and let $N_s > 0$. Suppose that $f\left(\mathrm{Id}\otimes\Phi(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) < 0$, where $|\Psi_{N_s}\rangle$ is the two-mode squeezed vacuum state defined in (5.39) and f is the function defined in Lemma 2. The energy-constrained two-way capacities $Q_2(\Phi, N_s)$ and $K(\Phi, N_s)$ are strictly positive. In particular, the (unconstrained) two-way capacities $Q_2(\Phi)$ and $K(\Phi)$ are strictly positive.*

Proof. Since the state $\mathrm{Id}\otimes\Phi(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)$ is a two-mode Gaussian state, we can apply Lemma 2 to conclude that it is entangled. Consequently, since any two-mode Gaussian entangled state is distillable [32], then $\mathrm{Id}\otimes\Phi(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)$ is distillable. Hence, by exploiting (5.50), we deduce that $Q_2(\Phi, N_s) > 0$. In addition, (5.46) implies that $K(\Phi, N_s) > 0$. Finally, since the energy-constrained capacities are lower bounds on the corresponding unconstrained capacities, we conclude that the unconstrained two-way capacities of Φ are strictly positive. □

The forthcoming Theorem 13 determines the parameter region of $g \geq 1$ and $\lambda \in [0, 1]$ where the composition $\mathcal{N}_{g,\lambda} \coloneqq \Phi_{g,0} \circ \mathscr{E}_{\lambda,0}$ between pure amplifier channel $\Phi_{g,0}$ and pure loss channel $\mathscr{E}_{\lambda,0}$ has strictly positive (EC) two-way capacities. In particular, we show that the (EC) two-way capacities of $\mathcal{N}_{g,\lambda}$ are strictly positive if and only if $\mathcal{N}_{g,\lambda}$ is not entanglement breaking.

Theorem 13. *Let $\lambda \in [0, 1]$, $g \geq 1$, and $N_s > 0$. The energy-constrained two-way capacities $Q_2(\mathcal{N}_{g,\lambda}, N_s)$ and $K(\mathcal{N}_{g,\lambda}, N_s)$ are strictly positive if and only if $(1-\lambda)g < 1$, i.e. if and only if $\mathcal{N}_{g,\lambda}$ is not entanglement breaking. In particular, the same holds for the unconstrained two-way capacities.*

Proof. Suppose that $(1-\lambda)g \geq 1$. Then Lemma 4 implies that $\mathcal{N}_{g,\lambda}$ is entanglement breaking and hence [20] its two way-capacities vanish.

Now, suppose that $(1-\lambda)g < 1$. Let us check that the hypothesis of Lemma 12 is fulfilled, i.e. we need to check that $f\left(\mathrm{Id}\otimes\mathcal{N}_{g,\lambda}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|)\right) < 0$, where $|\Psi_{N_s}\rangle$ is the two-mode squeezed

vacuum state defined in (5.39) and f is the function defined in Lemma 2. Let us calculate the covariance matrix of the state

$$\begin{aligned}&\mathrm{Id}_A \otimes \mathscr{N}_{g,\lambda}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}) \\ &= \mathrm{Tr}_{E_1E_2}\left[\left(\mathbb{1}_A \otimes U_g^{A'E_1} \otimes U_\lambda^{A'E_2}\right) |\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'} \otimes |0\rangle\langle 0|_{E_1} \otimes |0\rangle\langle 0|_{E_2} \left(\mathbb{1}_A \otimes U_g^{A'E_1} \otimes U_\lambda^{A'E_2}\right)^\dagger\right]\end{aligned} \tag{5.143}$$

with respect the ordering (A, A', E_1, E_2). By using (5.57) and (5.74), one can show that the covariance matrix of

$$\left(\mathbb{1}_A \otimes U_g^{A'E_1} \otimes U_\lambda^{A'E_2}\right) |\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'} \otimes |0\rangle\langle 0|_{E_1} \otimes |0\rangle\langle 0|_{E_2} \left(\mathbb{1}_A \otimes U_g^{A'E_1} \otimes U_\lambda^{A'E_2}\right)^\dagger \tag{5.144}$$

with respect the ordering (A, A', E_1, E_2) is

$$(\mathbb{1}_2 \oplus S_g \oplus \mathbb{1}_2)\left(\mathbb{1}_2 \oplus \bar{S}_\lambda\right)\left(V(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}) \oplus V(|0\rangle\langle 0|) \oplus V(|0\rangle\langle 0|)\right)\left(\mathbb{1}_2 \oplus \bar{S}_\lambda^\intercal\right)\left(\mathbb{1}_2 \oplus S_g^\intercal \oplus \mathbb{1}_2\right), \tag{5.145}$$

where

$$\bar{S}_\lambda := \begin{pmatrix} \sqrt{\lambda}\,\mathbb{1}_2 & 0_{2\times 2} & \sqrt{1-\lambda}\,\mathbb{1}_2 \\ 0_{2\times 2} & \mathbb{1}_2 & 0_{2\times 2} \\ -\sqrt{1-\lambda}\,\mathbb{1}_2 & 0_{2\times 2} & \sqrt{\lambda}\,\mathbb{1}_2 \end{pmatrix} \tag{5.146}$$

and $0_{2\times 2} := \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Hence, since $V\left(\mathrm{Id}_A \otimes \mathscr{N}_{g,\lambda}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'})\right)$ is the 4×4 upper-left block of the covariance matrix in (5.145), one can show that

$$V\left(\mathrm{Id}_A \otimes \mathscr{N}_{g,\lambda}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'})\right) = \begin{pmatrix} (2N_s+1)\mathbb{1}_2 & 2\sqrt{g\lambda N_s(N_s+1)}\sigma_z \\ 2\sqrt{g\lambda N_s(N_s+1)}\sigma_z & [2g\,(1+\lambda N_s)-1]\mathbb{1}_2 \end{pmatrix}, \tag{5.147}$$

where we used (5.38) and (5.41). Consequently, since

$$f\left(\mathrm{Id}_A \otimes \mathscr{N}_{g,\lambda}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}\right) = -16N_s(1+N_s)g\,(1-(1-\lambda)g), \tag{5.148}$$

and since $(1-\lambda)g < 1$, we have that $f\left(\mathrm{Id}_A \otimes \mathscr{N}_{g,\lambda}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}\right) < 0$, i.e. the hypothesis of Lemma 12 is fulfilled. Hence, Lemma 12 implies that the energy-constrained two-way capacities of $\mathscr{N}_{g,\lambda}$ are strictly positive. This concludes the proof of Theorem 13. In Remark 16 we will provide an alternative proof. □

In the forthcoming Theorem 14 we obtain a lower bound on the two-way capacities of a quantum channel $\Phi : \mathfrak{S}(L^2(\mathbb{R})) \to \mathfrak{S}(L^2(\mathbb{R}))$ by introducing a protocol to distribute ebits though Φ. The idea of such a protocol is the following. First, Alice prepares states of the form

$$|\Psi_{M,c}\rangle_{AA'} := c\,|0\rangle_A\,|0\rangle_{A'} + \sqrt{1-c^2}\,|M\rangle_A\,|M\rangle_{A'}, \tag{5.149}$$

where $M \in \mathbb{N}^+$ and $c \in (0,1)$. Then, she sends the halves A' to Bob trough Φ, who makes a measurement on each half in order to project his half onto the span of $\{|0\rangle, |M\rangle\}$. Then, Alice and Bob run k times the P1-or-P2 recurrence protocol [41] on the resulting states. After this, Alice and Bob run the improved hashing protocol introduced in [42] in order to generate ebits. Let $R(\Phi, M, c, k)$ be the rate of distributed ebits of this protocol. A lower bound on $Q_2(\Phi)$ (and hence on $K(\Phi)$) can be obtained by maximising $R(\Phi, M, c, k)$ over $M \in \mathbb{N}^+$, $c \in (0,1)$, and $k \in \mathbb{N}$.

Theorem 14. *Let $\Phi : \mathfrak{S}(L^2(\mathbb{R})) \to \mathfrak{S}(L^2(\mathbb{R}))$ be a quantum channel which maps a single-mode system A' into another single-mode system B. The EC two-way capacities $Q_2(\Phi, N_s)$ and $K(\Phi, N_s)$ satisfy the following lower bound*

$$K(\Phi, N_s) \geq Q_2(\Phi, N_s) \geq \sup_{\substack{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N} \\ (1-c^2)M \leq N_s}} R(\Phi, M, c, k), \tag{5.150}$$

and, in particular, the unconstrained two-way capacities satisfy

$$K(\Phi) \geq Q_2(\Phi) \geq \sup_{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N}} R(\Phi, M, c, k)\,, \tag{5.151}$$

where

$$R(\Phi, M, c, k) := \mathscr{C}(\Phi, c, M)\frac{\prod_{t=0}^{k-1} P_t}{2^k}\mathscr{I}(\alpha_{00}^{(k)}, \alpha_{01}^{(k)}, \alpha_{10}^{(k)}, \alpha_{11}^{(k)})\,. \tag{5.152}$$

Fixed $c \in (0,1)$, $M \in \mathbb{N}^+$, and $k \in \mathbb{N}$, the quantities present in (5.152) are defined as follows. $\mathscr{C}(\Phi, c, M)$ is defined as

$$\mathscr{C}(\Phi, c, M) := \operatorname{Tr}\left[\mathbb{1}_A \otimes \Pi_M\ \mathrm{Id}_A \otimes \Phi(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|_{AA'})\right]\,, \tag{5.153}$$

where $\Pi_M := |0\rangle\langle 0|_B + |M\rangle\langle M|_B$ and the state $|\Psi_{M,c}\rangle_{AA'}$ is defined in (5.149). Let us define for all $m, n \in \{0,1\}$ the coefficients $\alpha_{mn}^{(0)}$ as

$$\alpha_{mn}^{(0)} := \frac{\langle\psi_{mn}^{(M)}|_{AB}\, \mathbb{1}_A \otimes \Pi_M\ \mathrm{Id}_A \otimes \Phi(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|)\, \mathbb{1}_A \otimes \Pi_M\, |\psi_{mn}^{(M)}\rangle_{AB}}{\mathscr{C}(\Phi, c, M)}\,, \tag{5.154}$$

where $|\psi_{mn}^{(M)}\rangle_{AB}$ is defined as

$$|\psi_{mn}^{(M)}\rangle_{AB} := \frac{1}{\sqrt{2}}\sum_{j=0}^{1}(-1)^{mj}\,|jM\rangle_A \otimes |(j\oplus n)M\rangle_B\,. \tag{5.155}$$

For all $t \in \{0, 1, \ldots, k-1\}$ and all $m, n \in \{0,1\}$ the coefficients $\alpha_{mn}^{(t+1)}$ and P_t are defined in the following way:

- *If $\alpha_{10}^{(t)} < \alpha_{01}^{(t)}$, then*

$$\alpha_{mn}^{(t+1)} := \frac{1}{P_t}\sum_{\substack{m_1,m_2=0\\ m_1\oplus m_2=m}}^{1} \alpha_{m_1 n}^{(t)}\alpha_{m_2 n}^{(t)}\,, \tag{5.156}$$

  where

$$P_t := \sum_{n=0}^{1}\left(\sum_{m=0}^{1}\alpha_{mn}^{(t)}\right)^2\,. \tag{5.157}$$

- *If $\alpha_{10}^{(t)} \geq \alpha_{01}^{(t)}$, then*

$$\alpha_{mn}^{(t+1)} := \frac{1}{P_t}\sum_{\substack{n_1,n_2=0\\ n_1\oplus n_2=n}}^{1} \alpha_{mn_1}^{(t)}\alpha_{mn_2}^{(t)}\,, \tag{5.158}$$

  where

$$P_t := \sum_{m=0}^{1}\left(\sum_{n=0}^{1}\alpha_{mn}^{(t)}\right)^2\,. \tag{5.159}$$

For all $\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11} \geq 0$ with $\alpha_{00}+\alpha_{01}+\alpha_{10}+\alpha_{11} = 1$, the quantity $\mathscr{I}(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$ is defined as

$$\mathscr{I}(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}) := \max\left(Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}), Y(\alpha_{00}, \alpha_{10}, \alpha_{01}, \alpha_{11}),\, Y(\alpha_{01}, \alpha_{00}, \alpha_{10}, \alpha_{11})\right)\,, \tag{5.160}$$

where the function Y is defined in (5.51).

Proof. We introduce a protocol to distribute ebits through the channel Φ, which depends on three parameters: $M \in \mathbb{N}^+$, $c \in (0,1)$, $k \in \mathbb{N}$. Our lower bound on $Q_2(\Phi, N_s)$ in (5.150) can be obtained by optimising over these parameters the rate of ebits of such a protocol. The lower bound on the other EC two-way capacities follows from (5.46). The steps of the protocol are the following.

-Step 1: Alice prepares n_0 copies of the state $|\Psi_{M,c}\rangle_{AA'}$ in (5.149) and she sends the halves A' to Bob through the channel Φ. Hence, Alice and Bob share n_0 copies of the state $\mathrm{Id}_A \otimes \Phi(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|)$.

-Step 2: Bob performs the local POVM $\{\Pi_M, \mathbb{1} - \Pi_M\}$ on each pair $\mathrm{Id}_A \otimes \Phi(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|)$, where $\Pi_M := |0\rangle\langle 0| + |M\rangle\langle M|$. If Bob finds the outcome which corresponds to Π_M, then Alice and Bob keep the pair, otherwise they discard it. They keep the pair with probability

$$\mathscr{C}(\Phi, c, M) := \mathrm{Tr}\left[\mathbb{1}_A \otimes \Pi_M \ \mathrm{Id}_A \otimes \Phi(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|)\right] . \tag{5.161}$$

At this point, Alice and Bob shares $\approx n_0 \mathscr{C}(\Phi, c, M)$ pairs. Each of these pairs are in the state ρ' given by

$$\rho' = \frac{\mathbb{1}_A \otimes \Pi_M \ \mathrm{Id}_A \otimes \Phi(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|) \ \mathbb{1}_A \otimes \Pi_M}{\mathscr{C}(\Phi, c, M)} . \tag{5.162}$$

Note that the support of ρ' is equal to $\mathrm{Span}\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |M\rangle, |M\rangle \otimes |0\rangle, |M\rangle \otimes |M\rangle\}$. For simplicity, in the following we will use the notation $|1\rangle \equiv |M\rangle$. This formally corresponds to consider the state $\rho'' := U_M \otimes U_M \, \rho' \, U_M^\dagger \otimes U_M^\dagger$, which is obtained once both Alice and Bob have applied the unitary

$$U_M := \sum_{i \neq \{0,M\}}^{\infty} |i\rangle\langle i| + |1\rangle\langle M| + |M\rangle\langle 1| \tag{5.163}$$

on the remaining state ρ'. Hence, since the support of ρ'' is equal to $\mathrm{Span}\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}$, in the following we consider transformations which act on qubit systems.

-Step 3: For each of the $\approx n_0 \mathscr{C}(\Phi, c, M)$ pairs, Alice and Bob choose randomly two bits $\mu, \nu \in \{0, 1\}$ and they both apply the unitary $\sigma_{\mu\nu}$ defined by

$$\sigma_{\mu\nu} := \sum_{i=0}^{1} (-1)^{\mu i} |i \oplus \nu\rangle\langle i| \tag{5.164}$$

(in terms of the Pauli matrices it holds that $\sigma_{00} = \mathbb{1}_2$, $\sigma_{01} = \sigma_x$, $\sigma_{10} = \sigma_z$, and $\sigma_{11} = i\sigma_y$). Hence, each pair is transformed into the state ρ_0 defined by

$$\rho_0 := \frac{1}{4} \sum_{\mu,\nu=0}^{1} (\sigma_{\mu\nu} \otimes \sigma_{\mu\nu}) \ \rho'' \ (\sigma_{\mu\nu} \otimes \sigma_{\mu\nu})^\dagger . \tag{5.165}$$

By exploiting the fact that the Bell states defined in 5.31 form an orthonormal basis, one can show that ρ is diagonal in the Bell basis:

$$\rho_0 = \sum_{m,n=0}^{1} \alpha_{mn}^{(0)} |\psi_{mn}\rangle\langle\psi_{mn}| , \tag{5.166}$$

where the coefficients $\alpha_{mn}^{(0)}$ are given by

$$\alpha_{mn}^{(0)} = \langle\psi_{mn}| \rho'' |\psi_{mn}\rangle = \langle\psi_{mn}^{(M)}| \rho' |\psi_{mn}^{(M)}\rangle , \tag{5.167}$$

with $|\psi_{mn}^{(M)}\rangle$ being defined in 5.155.

-Step 4: Alice and Bob run the following sub-routine, which is a recurrence protocol dubbed *P1-or-P2* [41].

- **Step 4.0**: Let $t = 0$.
- **Step 4.1**: At this point, all the pairs are in the state ρ_t. Alice and Bob collect all the pairs in groups of two pairs. Let $\rho_t^{(A_1B_1)}$ denote the first pair of each group and let $\rho_t^{(A_2B_2)}$ denote the second one. If $\alpha_{10}^{(t)} < \alpha_{01}^{(t)}$, then Alice and Bob apply the bi-local unitary U_1 defined as

$$U_1 := U_{\mathrm{CNOT}}^{(A_1A_2)} \otimes U_{\mathrm{CNOT}}^{(B_1B_2)} , \tag{5.168}$$

where for all $S = A, B$ the operator $U_{\text{CNOT}}^{(S_1S_2)}$ is the CNOT gate on S_1 and S_2 with control qubit S_1, i.e.

$$U_{\text{CNOT}}^{(S_1S_2)} |i\rangle_{S_1} \otimes |j\rangle_{S_2} = |i\rangle_{S_1} \otimes |i \oplus j\rangle_{S_2} \ . \tag{5.169}$$

Otherwise if $\alpha_{10}^{(t)} \geq \alpha_{01}^{(t)}$, they apply the bi-local unitary U_2 defined as

$$U_2 := (H^{(A_1)} \otimes H^{(B_1)})(U_{\text{CNOT}}^{(A_1A_2)} \otimes U_{\text{CNOT}}^{(B_1B_2)})(H^{(A_1)} \otimes H^{(A_2)} \otimes H^{(B_1)} \otimes H^{(B_2)}) \, , \tag{5.170}$$

where for all $S = A_1, A_2, B_1, B_2$ the operator $H^{(S)}$ on S is the Hadamard gate, i.e.

$$H^{(S)} = \frac{1}{\sqrt{2}} \sum_{m,n=0}^{1} (-1)^{mn} |n\rangle\langle m|_S \ . \tag{5.171}$$

At this point, the state of $A_1A_2B_1B_2$ is

$$\rho_t^{(A_1A_2B_1B_2)} := U_p \, \rho_t^{(A_1B_2)} \otimes \rho_k^{(A_2B_2)} \, U_p^\dagger \, . \tag{5.172}$$

with $p = 1$ if $\alpha_{10}^{(t)} < \alpha_{01}^{(t)}$, and $p = 2$ otherwise.

- **Step 4.2**: Alice and Bob measure the pair A_2B_2 of each group with respect to the local POVM $\{M_{i,j}\}_{i,j\in\{0,1\}}$ with $M_{i,j} := |i\rangle\langle i|_{A_2} \otimes |j\rangle\langle j|_{B_2}$ for all $i, j \in \{0, 1\}$. Then they discard the pair A_2B_2. They discard also the pair A_1B_1 if the outcome of the previous measurement corresponds to $M_{i,j}$ with $i \neq j$. The probability that a pair A_1B_1 is not discarded is given by

$$P_t := \sum_{i=0}^{1} \langle i|_{A_2} \langle i|_{B_2} \, \text{Tr}_{A_1B_1} \left[\rho_t^{(A_1A_2B_1B_2)}\right] |i\rangle_{A_2} |i\rangle_{B_2} \ . \tag{5.173}$$

By using that for all $k_1, k_2, j_1, j_2 \in \{0, 1\}$ it holds that

$$U_{\text{CNOT}}^{(A_1A_2)} \otimes U_{\text{CNOT}}^{(B_1B_2)} |\psi_{k_1j_1}\rangle_{A_1B_1} \otimes |\psi_{k_2j_2}\rangle_{A_2B_2} = |\psi_{k_1\oplus k_2, \, j_1}\rangle_{A_1B_1} \otimes |\psi_{k_2, \, j_1\oplus j_2}\rangle_{A_2B_2} \tag{5.174}$$

and that

$$H^{(A)} \otimes H^{(B)} |\psi_{k_1j_1}\rangle_{AB} = (-1)^{k_1j_1} |\psi_{j_1k_1}\rangle_{AB} \ , \tag{5.175}$$

one can show that P_t can be expressed as in (5.157) if $\alpha_{10}^{(t)} < \alpha_{01}^{(t)}$, and as in (5.159) otherwise. At this point, the number of remaining pairs is

$$\approx n_0 \, \mathscr{C}(\Phi, c, M) \frac{1}{2^{t+1}} \prod_{m=0}^{t} P_m \tag{5.176}$$

and each of these is in the state ρ_{t+1} given by

$$\rho_{t+1} = \frac{1}{2} \sum_{i=0}^{1} \frac{\langle i|_{A_2} \langle i|_{B_2} \, \rho_t^{(A_1A_2B_1B_2)} \, |i\rangle_{A_2} |i\rangle_{B_2}}{\text{Tr}_{A_1B_1} \left[\langle i|_{A_2} \langle i|_{B_2} \, \rho_t^{(A_1A_2B_1B_2)} \, |i\rangle_{A_2} |i\rangle_{B_2}\right]} \, . \tag{5.177}$$

By using (5.174) and (5.175), one can show that

$$\rho_{t+1} = \sum_{m,n=0}^{1} \alpha_{mn}^{(t+1)} \, |\psi_{mn}\rangle\langle\psi_{mn}| \ , \tag{5.178}$$

where the coefficients $\alpha_{mn}^{(t+1)}$ are given by (5.156) if $\alpha_{10}^{(t)} < \alpha_{01}^{(t)}$, and by (5.158) otherwise.

- **Step 4.3**: Let $t = t + 1$.
- **Step 4.4**: If the condition $t < k$ is satisfied, then go back to Step 4.1.

Before introducing Step 5, let us recall that if the improved hashing protocol of [42] is applied on states of the form $\rho = \sum_{ij=0}^{1} \alpha_{ij} |\psi_{ij}\rangle\langle\psi_{ij}|$ then it can generate ebits with a yield $Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$ given by (5.51). Note that such a yield is not invariant under permutations of the variables $\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}$. Hence, one may achieve a yield which is larger than $Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$ by applying suitable bi-local unitaries, which suitably permutes the Bell states, just before running the improved hashing protocol. Since the yield function $Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$ satisfies

$$\begin{aligned} Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}) &= Y(\alpha_{10}, \alpha_{01}, \alpha_{00}, \alpha_{11}) , \\ Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}) &= Y(\alpha_{00}, \alpha_{11}, \alpha_{10}, \alpha_{01}) , \\ Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}) &= Y(\alpha_{01}, \alpha_{00}, \alpha_{11}, \alpha_{10}) , \end{aligned} \tag{5.179}$$

then by permuting the four variables α_{ij} it is possible to obtain at most three different values of the rate function, which are: $Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$, $Y(\alpha_{00}, \alpha_{10}, \alpha_{01}, \alpha_{11})$, and $Y(\alpha_{01}, \alpha_{00}, \alpha_{10}, \alpha_{11})$. Let us define the function $\mathscr{I}$ as

$$\mathscr{I}(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}) := \max\left(Y(\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}), Y(\alpha_{00}, \alpha_{10}, \alpha_{01}, \alpha_{11}), Y(\alpha_{01}, \alpha_{00}, \alpha_{10}, \alpha_{11}) \right) . \tag{5.180}$$

Note that at the beginning of Step 5, the number of remaining pairs is

$$\approx n_0 \, \mathscr{C}(\Phi, c, M) \frac{1}{2^k} \prod_{t=0}^{k-1} P_t \tag{5.181}$$

and each of these is in $\rho_k = \sum_{m,n=0}^{1} \alpha_{mn}^{(k)} |\psi_{mn}\rangle\langle\psi_{mn}|$.

-Step 5: If $\mathscr{I}(\alpha_{00}^{(k)}, \alpha_{01}^{(k)}, \alpha_{10}^{(k)}, \alpha_{11}^{(k)}) = Y(\alpha_{00}^{(k)}, \alpha_{10}^{(k)}, \alpha_{01}^{(k)}, \alpha_{11}^{(k)})$, then both Alice and Bob apply the Hadamard gate defined by (5.171). Therefore, in this case, the state of each of pairs becomes

$$(H \otimes H) \, \rho_k \, (H \otimes H)^\dagger = \alpha_{00}^{(k)} |\psi_{00}\rangle\langle\psi_{00}| + \alpha_{10}^{(k)} |\psi_{01}\rangle\langle\psi_{01}| + \alpha_{01}^{(k)} |\psi_{10}\rangle\langle\psi_{10}| + \alpha_{11}^{(k)} |\psi_{11}\rangle\langle\psi_{11}| , \tag{5.182}$$

where we have exploited (5.175). If $\mathscr{I}(\alpha_{00}^{(k)}, \alpha_{01}^{(k)}, \alpha_{10}^{(k)}, \alpha_{11}^{(k)}) = Y(\alpha_{01}^{(k)}, \alpha_{00}^{(k)}, \alpha_{10}^{(k)}, \alpha_{11}^{(k)})$, then both Alice and Bob apply $B_x := \frac{\mathbb{1}_2 - i\sigma_{01}}{\sqrt{2}}$, where σ_{01} is defined by (5.164), and hence the state becomes

$$(B_x \otimes B_x) \, \rho_k \, (B_x \otimes B_x)^\dagger = \alpha_{01}^{(k)} |\psi_{00}\rangle\langle\psi_{00}| + \alpha_{00}^{(k)} |\psi_{01}\rangle\langle\psi_{01}| + \alpha_{10}^{(k)} |\psi_{10}\rangle\langle\psi_{10}| + \alpha_{11}^{(k)} |\psi_{11}\rangle\langle\psi_{11}| . \tag{5.183}$$

-Step 6: Alice and Bob run the improved hashing protocol of [42], which can achieve the yield $\mathscr{I}(\alpha_{00}^{(k)}, \alpha_{01}^{(k)}, \alpha_{10}^{(k)}, \alpha_{11}^{(k)})$. Hence, in the end, Alice and Bob can generate a number of ebits equal to

$$\approx n_0 \, \mathscr{C}(\Phi, c, M) \frac{\prod_{t=0}^{k-1} P_t}{2^k} \mathscr{I}(\alpha_{00}^{(k)}, \alpha_{01}^{(k)}, \alpha_{10}^{(k)}, \alpha_{11}^{(k)}) . \tag{5.184}$$

Since the channel Φ is used n_0 times (during Step 1) to send the n_0 halves of the state $|\Psi_{M,c}\rangle_{AA'}$, the rate of distributed ebits of the presented protocol is

$$\mathscr{C}(\Phi, c, M) \frac{\prod_{t=0}^{k-1} P_t}{2^k} \mathscr{I}(\alpha_{00}^{(k)}, \alpha_{01}^{(k)}, \alpha_{10}^{(k)}, \alpha_{11}^{(k)}) . \tag{5.185}$$

Since the local mean photon number of $|\Psi_{M,c}\rangle_{AA'}$ is $(1-c^2)M$, the rate in (5.185) is a lower bound on the energy-constrained two-way quantum capacity $Q_2(\Phi, N_s)$ for all $M \in \mathbb{N}^+$, $c \in (0,1)$, $k \in \mathbb{N}$ such that $(1-c^2)M \leq N_s$. The optimisation over these parameters of the rate in (5.185) leads to the lower bound on $Q_2(\Phi, N_s)$ in (5.150). In addition, since $K(\Phi, N_s) \geq Q_2(\Phi, N_s)$ thanks to (5.46), we have proved (5.150). By taking the limit $N_s \to \infty$ of (5.150), the lower bound on the unconstrained two-way capacities in (5.151) is also proved. □

In the forthcoming Theorem 15 we apply Theorem 14 to the composition $\mathscr{N}_{g,\lambda} := \Phi_{g,0} \circ \mathscr{E}_{\lambda,0}$ between pure amplifier channel $\Phi_{g,0}$ and pure loss channel $\mathscr{E}_{\lambda,0}$.

Theorem 15. *Let $g \geq 1$, $\lambda \in [0,1]$, and $N_s \geq 0$. The EC two-way capacities $Q_2(\mathcal{N}_{g,\lambda}, N_s)$ and $K(\mathcal{N}_{g,\lambda}, N_s)$ of the composition $\mathcal{N}_{g,\lambda} := \Phi_{g,0} \circ \mathscr{E}_{\lambda,0}$ between pure amplifier channel and pure loss channel satisfy the following lower bound*

$$K(\mathcal{N}_{g,\lambda}, N_s) \geq Q_2(\mathcal{N}_{g,\lambda}, N_s) \geq \sup_{\substack{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N} \\ (1-c^2)M\leq N_s}} \mathscr{R}(g,\lambda,M,c,k)\,, \tag{5.186}$$

and, in particular, the unconstrained two-way capacities satisfy

$$K(\mathcal{N}_{g,\lambda}) \geq Q_2(\mathcal{N}_{g,\lambda}) \geq \sup_{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N}} \mathscr{R}(g,\lambda,M,c,k)\,, \tag{5.187}$$

where

$$\mathscr{R}(g,\lambda,M,c,k) := R(\mathcal{N}_{g,\lambda}, M, c, k)\,, \tag{5.188}$$

with the quantity $R(\mathcal{N}_{g,\lambda}, M, c, k)$ being defined in Theorem 14. The quantities $\mathscr{C}(\mathcal{N}_{g,\lambda}, c, M)$ and $\alpha^{(0)}_{mn}$, which appear in the definition of $R(\mathcal{N}_{g,\lambda}, M, c, k)$ in Theorem 14, can be expressed as

$$\begin{aligned}
\mathscr{C}(\mathcal{N}_{g,\lambda}, c, M) &:= \sum_{n,l=0}^{1} c_n^2\, f_{Mn,Mn,Ml}(g,\lambda)\,, \\
\alpha^{(0)}_{mn} &:= \frac{1}{2\mathscr{C}(\mathcal{N}_{g,\lambda}, c, M)} \sum_{x,y=0}^{1} \sum_{l=\max(y-x,0)}^{1+\min(y-x,0)} \delta_{x\oplus n, l+x-y}\, \delta_{y\oplus n,l}\, (-1)^{m(x+y)} c_x c_y\, f_{Mx,My,Ml}(g,\lambda)\,,
\end{aligned} \tag{5.189}$$

where $c_0 := c$, $c_1 := \sqrt{1-c^2}$, $f_{n,i,l}(g,\lambda)$ is defined in (5.109), and $\delta_{x,y}$ denotes the Kronecker delta.

Proof. (5.186) and (5.187) follows by applying Theorem 14 to $\mathcal{N}_{g,\lambda}$. We only need to show the expressions of $\mathscr{C}(\mathcal{N}_{g,\lambda}, c, M)$ and $\alpha^{(0)}_{mn}$ in (5.189). In this proof we use the notation introduced in the statement of Theorem 14. By using (5.108), we deduce that

$$\mathrm{Id}_A \otimes \mathcal{N}_{g,\lambda}(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|) = \sum_{n,i=0}^{1} \sum_{l=M\max(i-n,0)}^{\infty} c_n c_i f_{Mn,Mi,l}(g,\lambda)\, |Mn\rangle\langle Mi|_A \otimes |l+M(n-i)\rangle\langle l|_B\,. \tag{5.190}$$

Consequently, it holds that

$$\begin{aligned}
&\mathbb{1}_A \otimes \Pi_M\ \mathrm{Id}_A \otimes \mathcal{N}_{g,\lambda}(|\Psi_{M,c}\rangle\langle\Psi_{M,c}|)\, \mathbb{1}_A \otimes \Pi_M \\
&= \sum_{n,i=0}^{1} \sum_{l=\max(i-n,0)}^{1+\min(i-n,0)} c_n c_i f_{Mn,Mi,Ml}(g,\lambda)\, |Mn\rangle\langle Mi|_A \otimes |M(l+n-i)\rangle\langle Ml|_B\,.
\end{aligned} \tag{5.191}$$

By inserting this into the definition of $\mathscr{C}(\mathcal{N}_{g,\lambda}, c, M)$ in (5.153) and of $\alpha^{(0)}_{mn}$ in (5.154), one obtains the expressions in (5.189). □

Remarks

Let us consider the entanglement distribution protocol shown in the proof of Theorem 14 applied to the the composition $\mathcal{N}_{g,\lambda} := \Phi_{g,0} \circ \mathscr{E}_{\lambda,0}$ between pure amplifier channel $\Phi_{g,0}$ and pure loss channel $\mathscr{E}_{\lambda,0}$. After completing Step 2 of this protocol, the entanglement distribution process is reduced to an entanglement distillation protocol on the two-qubit state reported in (5.162). We will denote this two-qubit state as $\rho_{AB}^{(g,\lambda,M,c)}$, where $M \in \mathbb{N}^+$ and $c \in (0,1)$ correspond to the constants appearing in the state in (5.149) that Alice produces during Step 1. The natural question that arises is: "Under what conditions is $\rho_{AB}^{(g,\lambda,M,c)}$ distillable?" In Remark 16 we answer this question.

Remark 16. *$\rho_{AB}^{(g,\lambda,M,c)}$ is distillable if and only if λ and g satisfy the inequality $(1-\lambda)g < 1$, meaning that $\mathcal{N}_{g,\lambda}$ is not entanglement breaking. This provides an alternative proof of Theorem 13.*

Proof. By exploiting (5.191), for all $g > 1, \lambda \in (0,1), M \in \mathbb{N}^+, c \in (0,1)$ the state in (5.162) can be expressed as

$$\rho_{AB}^{(g,\lambda,M,c)} := \frac{\sum_{n,i=0}^{1} \sum_{l=\max(i-n,0)}^{1+\min(i-n,0)} c_n c_i f_{Mn,Mi,Ml}(g,\lambda) \, |Mn\rangle\langle Mi|_A \otimes |M(l+n-i)\rangle\langle Ml|_B}{\sum_{n,l=0}^{1} c_n^2 \, f_{Mn,Mn,Ml}(g,\lambda)}, \quad (5.192)$$

where $c_0 := c$, $c_1 := \sqrt{1-c^2}$, and $f_{n,i,l}(g,\lambda)$ is defined in (5.109). Consequently, it holds that

$$\begin{aligned}\rho_{AB}^{(g,\lambda,M,c)} = \frac{1}{\sum_{n,l=0}^{1} c_n^2 \, f_{Mn,Mn,Ml}(g,\lambda)} \Big[& c^2 f_{0,0,0}(g,\lambda) \, |0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B \\ & + c^2 f_{0,0,M}(g,\lambda) \, |0\rangle\langle 0|_A \otimes |M\rangle\langle M|_B \\ & + c\sqrt{1-c^2} f_{0,M,M}(g,\lambda) \, |0\rangle\langle M|_A \otimes |0\rangle\langle M|_B \\ & + c\sqrt{1-c^2} f_{M,0,0}(g,\lambda) \, |M\rangle\langle 0|_A \otimes |M\rangle\langle 0|_B \\ & + (1-c^2) f_{M,M,0}(g,\lambda) \, |M\rangle\langle M|_A \otimes |0\rangle\langle 0|_B \\ & + (1-c^2) f_{M,M,M}(g,\lambda) \, |M\rangle\langle M|_A \otimes |M\rangle\langle M|_B \Big], \end{aligned} \quad (5.193)$$

Hence, the matrix associated with the partial transpose on B of $\rho_{AB}^{(g,\lambda,M,c)}$, written with respect the basis $\{|0\rangle_A \otimes |0\rangle_B, |0\rangle_A \otimes |M\rangle_B, |M\rangle_A \otimes |0\rangle_B, |M\rangle_A \otimes |M\rangle_B\}$, is

$$\frac{1}{\sum_{n,l=0}^{1} c_n^2 \, f_{Mn,Mn,Ml}(g,\lambda)} \begin{pmatrix} c^2 f_{0,0,0}(g,\lambda) & 0 & 0 & 0 \\ 0 & c^2 f_{0,0,M}(g,\lambda) & c\sqrt{1-c^2} f_{0,M,M}(g,\lambda) & 0 \\ 0 & c\sqrt{1-c^2} f_{M,0,0}(g,\lambda) & (1-c^2) f_{M,M,0}(g,\lambda) & 0 \\ 0 & 0 & 0 & (1-c^2) f_{M,M,M}(g,\lambda) \end{pmatrix}.$$

It follows that $\rho_{AB}^{(g,\lambda,M,c)}$ is not PPT if and only if

$$f_{M,0,0}(g,\lambda) \, f_{M,M,0}(g,\lambda) < f_{0,M,M}(g,\lambda) \, f_{0,0,M}(g,\lambda). \quad (5.194)$$

The definition of $f_{\cdot,\cdot,\cdot}(g,\lambda)$ in (5.109) yields

$$\begin{aligned} f_{0,0,M}(g,\lambda) &= \frac{(g-1)^M}{g^{1+M}}, \\ f_{M,M,0}(g,\lambda) &= \frac{(1-\lambda)^M}{g}, \\ f_{0,M,M}(g,\lambda) &= f_{M,0,0}(g,\lambda) = \frac{\lambda^{\frac{M}{2}}}{g^{1+\frac{M}{2}}}. \end{aligned} \quad (5.195)$$

Consequently, (5.194) establishes that $\rho_{AB}^{(g,\lambda,M,c)}$ is not PPT if and only if $(1-\lambda)g < 1$, independentely of c and M. The fact that any two-qubit state is distillable if and only if it is not PPT [59] implies that $\rho_{AB}^{(g,\lambda,M,c)}$ is distillable if and only if $(1-\lambda)g < 1$ for all $c \in (0,1)$ and all $M \in \mathbb{N}^+$.

Let us now show that this fact constitutes an alternative proof of Theorem 13, i.e. let us show that the energy-constrained two-way capacities $Q_2(\mathcal{N}_{g,\lambda}, N_s)$ and $K(\mathcal{N}_{g,\lambda}, N_s)$ are strictly positive if and only if $(1-\lambda)g < 1$, i.e. if and only if $\mathcal{N}_{g,\lambda}$ is not entanglement breaking. The entanglement distribution protocol's Steps S1 and S2 imply that for all $N_s \geq 0$ it holds that $Q_2(\mathcal{N}_{g,\lambda}, N_s) \geq E_d\left(\rho_{AB}^{(g,\lambda,M,c)}\right)$, for any $c \in (0,1)$ and $M \in \mathbb{N}^+$ satisfying $(1-c^2)M \leq N_s$. Here, $E_d(\cdot)$ denotes the distillable entanglement. As we have proved above, if $(1-\lambda)g < 1$ then the state $\rho_{AB}^{(g,\lambda,M,c)}$ is distillable, i.e. $E_d(\rho_{AB}^{(g,\lambda,M,c)}) > 0$. This implies that if $(1-\lambda)g < 1$, then the energy-constrained two-way capacities of $\mathcal{N}_{g,\lambda}$ are strictly positive, i.e. $K(\mathcal{N}_{g,\lambda}, N_s) \geq Q_2(\mathcal{N}_{g,\lambda}, N_s) > 0$. Conversely, by exploiting Theorem 4 and the fact that any entanglement-breaking channel has vanishing two-way capacities, it follows that if $(1-\lambda)g \geq 1$ then $K(\mathcal{N}_{g,\lambda}) = Q_2(\mathcal{N}_{g,\lambda}) = 0$ and hence $K(\mathcal{N}_{g,\lambda}, N_s) = Q_2(\mathcal{N}_{g,\lambda}, N_s) = 0$. □

Remark 16 ensures that if the channel $\mathcal{N}_{g,\lambda}$ is not entanglement breaking, then the state $\rho_{AB}^{(g,\lambda,M,c)}$ obtained at the end of Step 2 is distillable for any $M \in \mathbb{N}^+$ and $c \in (0,1)$. We now turn our attention to the state, denoted as $\sigma_{AB}^{(g,\lambda,M,c)}$, which is obtained at the end of Step 3 through Pauli-based twirling of $\rho_{AB}^{(g,\lambda,M,c)}$. It is possible for this operation to map distillable states to undistillable states, so we ask the question: "Under what conditions is $\sigma_{AB}^{(g,\lambda,M,c)}$ distillable?" In Remark 17 we will demonstrate that for any $\lambda \in (0,1)$ and $g > 1$, if $\mathcal{N}_{g,\lambda}$ is not entanglement breaking, then for all $M \in \mathbb{N}^+$ the state $\sigma_{AB}^{(g,\lambda,M,\bar{c})}$ is distillable, where $\bar{c} := \frac{1}{\sqrt{1+(g-1)^M}}$. This means that Alice and Bob can choose the value of c appropriately such that the Pauli-based twirling does not affect the distillability of the shared state.

Remark 17. *Let $M \in \mathbb{N}^+$, $\lambda \in (0,1)$, and $g > 1$ with $(1-\lambda)g < 1$ (meaning that $\mathcal{N}_{g,\lambda}$ is not entanglement breaking). Then, the state $\sigma_{AB}^{(g,\lambda,M,\bar{c})}$ is distillable, where $\bar{c} := \frac{1}{\sqrt{1+(g-1)^M}}$.*

Proof. After applying the Pauli-based twirling on the state $\rho_{AB}^{(g,\lambda,M,\bar{c})}$, the resulting state $\sigma_{AB}^{(g,\lambda,M,\bar{c})}$ is transformed into a Bell-diagonal form, that is

$$\sigma_{AB}^{(g,\lambda,M,\bar{c})} = \sum_{i,j=0}^{1} p_{ij} \, |\psi_{ij}^{(M)}\rangle\langle\psi_{ij}^{(M)}|_{AB} \, , \tag{5.196}$$

where $p_{ij} := \langle\psi_{ij}^{(M)}| \, \rho_{AB}^{(g,\lambda,M,\bar{c})} \, |\psi_{ij}^{(M)}\rangle$ and $\{|\psi_{ij}^{(M)}\rangle_{AB}\}_{i,j\in\{0,1\}}$ are the Bell states defined in (5.155). In particular, it holds that

$$\begin{aligned}
p_{00} + p_{10} &= \langle 0|_A \langle 0|_B \, \rho_{AB}^{(g,\lambda,M,c)} \, |0\rangle_A |0\rangle_B + \langle M|_A \langle M|_B \, \rho_{AB}^{(g,\lambda,M,c)} \, |M\rangle_A |M\rangle_B \, , \\
p_{01} - p_{11} &= 2 \langle 0|_A \langle 0|_B \, \rho_{AB}^{(g,\lambda,M,c)} \, |M\rangle_A |M\rangle_B \, , \\
p_{01} + p_{11} &= \langle 0|_A \langle M|_B \, \rho_{AB}^{(g,\lambda,M,c)} \, |0\rangle_A |M\rangle_B + \langle M|_A \langle 0|_B \, \rho_{AB}^{(g,\lambda,M,c)} \, |M\rangle_A |0\rangle_B \, , \\
p_{00} - p_{10} &= 2 \langle 0|_A \langle 0|_B \, \rho_{AB}^{(g,\lambda,M,c)} \, |M\rangle_A |M\rangle_B \, .
\end{aligned} \tag{5.197}$$

Lemma 18 guarantees that if $p_{01} + p_{11} - |p_{00} - p_{10}| < 0$ then the state $\sigma_{AB}^{(g,\lambda,M,\bar{c})}$ is distillable. By using (5.193) and (5.197), the condition $p_{01} + p_{11} - |p_{00} - p_{10}| < 0$ is satisfied if and only if

$$\bar{c}^2 f_{0,0,M}(g,\lambda) + (1-\bar{c}^2) f_{M,M,0}(g,\lambda) - 2\bar{c}\sqrt{1-\bar{c}^2} f_{M,0,0}(g,\lambda) < 0 \, , \tag{5.198}$$

that is

$$\bar{c}^2 (g-1)^M + (1-\bar{c}^2)(1-\lambda)^M g^M - 2\bar{c}\sqrt{1-\bar{c}^2}(\lambda g)^{M/2} < 0 \, , \tag{5.199}$$

where we have exploited (5.195). By hypothesis, the channel $\mathcal{N}_{g,\lambda}$ is entanglement breaking and hence $(1-\lambda)g < 1$, as established by Lemma 5. Consequently, for all $g > 1$ and $\lambda \in (0,1)$ it holds that

$$\begin{aligned}
\bar{c}^2 (g-1)^M + (1-\bar{c}^2)(1-\lambda)^M g^M - 2\bar{c}\sqrt{1-\bar{c}^2}(\lambda g)^{M/2} &< \bar{c}^2 (g-1)^M + (1-\bar{c}^2) - 2\bar{c}\sqrt{1-\bar{c}^2}(g-1)^{M/2} \\
&= \left(\bar{c}(g-1)^{M/2} - \sqrt{1-\bar{c}^2}\right)^2 = 0 \, ,
\end{aligned} \tag{5.200}$$

where we have used that $\bar{c} := \frac{1}{\sqrt{1+(g-1)^M}}$. Hence, for all $M \in \mathbb{N}^+$, $\lambda \in (0,1)$, and $g > 1$ with $(1-\lambda)g < 1$, it holds that $\sigma_{AB}^{(g,\lambda,M,\bar{c})}$ is distillable. □

Lemma 18. *Let $\{|\psi_{ij}\rangle\}_{i,j\in\{0,1\}}$ be the Bell states defined in (5.31). A convex combination of Bell states $\rho_{AB} = \sum_{i,j=0}^{1} p_{ij} \, |\psi_{ij}\rangle\langle\psi_{ij}|$ is distillable if and only if $p_{00} + p_{10} < |p_{01} - p_{10}|$ or $p_{01} + p_{11} < |p_{00} - p_{10}|$.*

Proof. The matrix associated with $\rho_{AB} = \sum_{i,j=0}^{1} p_{ij} \left|\psi_{ij}\right\rangle\!\left\langle\psi_{ij}\right|$, written with respect the basis $\{\left|0\right\rangle_A \otimes \left|0\right\rangle_B, \left|0\right\rangle_A \otimes \left|1\right\rangle_B, \left|1\right\rangle_A \otimes \left|0\right\rangle_B, \left|1\right\rangle_A \otimes \left|1\right\rangle_B\}$, is

$$\frac{1}{2}\begin{pmatrix} p_{00}+p_{10} & 0 & 0 & p_{00}-p_{10} \\ 0 & p_{10}+p_{11} & p_{10}-p_{11} & 0 \\ 0 & p_{10}-p_{11} & p_{10}+p_{11} & 0 \\ p_{00}-p_{10} & 0 & 0 & p_{00}+p_{10} \end{pmatrix}.$$

Its partial transpose on B is

$$\frac{1}{2}\begin{pmatrix} p_{00}+p_{10} & 0 & 0 & p_{10}-p_{11} \\ 0 & p_{10}+p_{11} & p_{00}-p_{10} & 0 \\ 0 & p_{00}-p_{10} & p_{10}+p_{11} & 0 \\ p_{10}-p_{11} & 0 & 0 & p_{00}+p_{10} \end{pmatrix}.$$

Hence, the state ρ_{AB} is PPT if and only if $p_{00}+p_{10} \geq |p_{01}-p_{10}|$ and $p_{01}+p_{11} \geq |p_{00}-p_{10}|$. Consequently, the fact that any two-qubit state is distillable if and only if it is not PPT [59] implies the validity of the thesis. □

Experimental challenges regarding our protocol

As demonstrated in the main text, applying our main result regarding the maximum tolerable excess noise to the current Internet infrastructure shows that continuous-variable quantum key distribution is feasible if and only if the fibre length is approximately less than 1000 kilometres. Hence, any practical QKD protocol, which is based on the existing Internet infrastructure, must adhere to this fundamental limit. Furthermore, this limit of 1000 kilometres can now serve as a benchmark for evaluating the quality of any new CV-QKD protocol, underscoring the significant impact of our results on practical implementations.

The potential benefit of our protocol (presented both in the main text and in the proof of Theorem 14 above) lies in its *faithfulness* — it can distil entanglement (and hence generate secret keys) whenever the channel is not entanglement breaking. With the current Internet infrastructure based on optical fibres, our protocol could *theoretically* achieve the ultimate limit set by quantum physics of transmitting entanglement and secret keys over distances up to 1000 kilometres. This is a unique feature of our protocol, which stands in stark contrast with *all* existing entanglement distribution and key distribution protocols.

While there exist CV-QKD protocols that are relatively easy to implement with current technology (capable of distributing secret keys across optical fibres of at most 200 kilometres [6–10]), this is not the case for entanglement distribution. Indeed, *all* known entanglement-distribution protocols are experimentally challenging with current technology. For example, the best known entanglement-distribution protocol prior to our work [25] — i.e. the hashing protocol applied to the Choi state of the channel — is not experimentally feasible.

Our protocol is an entanglement-distribution protocol and, as such, is experimentally challenging at present. We emphasise that this limitation is not unique to our protocol but is a common challenge faced by *all* entanglement distribution protocols due to current technological constraints. A major factor is the lack of a noiseless quantum memory, which makes it challenging to perform even a few iterations of a recurrence entanglement distillation protocol. Nevertheless, given the significant recent experimental advancements regarding quantum memories [60, 61] and entanglement distillation [45–47], we are optimistic about the future experimental viability of our protocol. Hence, we stress that, although our protocol is experimentally challenging with current technology, there is no way that a protocol as simple as our ours will *not* be realisable in a few decades at worst.

Let us provide further details about a possible practical realisation of our protocol. To perform Step 1, it suffices that Alice produces the state

$$|\Psi\rangle_{AA'} := \frac{|0\rangle_A \otimes |0\rangle_{A'} + |1\rangle_A \otimes |1\rangle_{A'}}{\sqrt{2}} \tag{5.201}$$

in order to make the rate of the protocol faithful. However, without changing the rate (as explained below), Alice can instead produce the *NOON state* [43]

$$|\Psi'\rangle_{AA'} := \frac{|0\rangle_A \otimes |1\rangle_{A'} + |1\rangle_A \otimes |0\rangle_{A'}}{\sqrt{2}} , \tag{5.202}$$

which can be experimentally prepared [62, 63]. After Alice has sent the sub-system A' through the channel to Bob, Step 2 involves performing a non-demolition measurement with the POVM operator $|0\rangle\langle 0|_B + |1\rangle\langle 1|_B$. Although this measurement appears challenging to implement experimentally, fortunately the problem has been studied already, and several promising approaches do exist. Specifically, one may exploit either: the pre-certification scheme employed in [64]; the coupling scheme between optical signals and trapped cold atomic gas designed in [65]; single photon filters based on Rydberg blockade [66, 67] to implement single photon subtraction [68, 69]; single atoms inside an optical cavity to perform single photon subtraction [70, 71]. After this non-demolition measurement, Alice and Bob share a two-mode state in the subspace spanned by $\{|0\rangle \otimes |0\rangle , |0\rangle \otimes |1\rangle , |1\rangle \otimes |0\rangle , |1\rangle \otimes |1\rangle\}$. At this point, they can transfer their state from the optical modes to a qubit solid-state platform (e.g., superconducting or trapped ion platform). This transfer can be experimentally performed in several ways, for example by exploiting: the quantum-memory based approaches introduced in [60, 61]; the aforementioned single photon subtraction methods [66–71], which map the photonic state onto the atomic state; quantum transduction from optical to microwave photons that are compatible with the superconducting qubits [72–74]. This means that only the first two steps of the protocol involve optical platforms, which is advantageous because all the two-qubit unitaries used in the subsequent steps of our protocol are much easier to experimentally implement in a qubit platform. To address the fact that Alice has sent the NOON state in (5.202) instead of the state in (5.201), she simply needs to apply the Pauli σ_x before initiating Step 3.

Multi-rail strategies

In this section we introduce an additional protocol for distributing ebits across the piBGC $\mathcal{N}_{g,\lambda}$ by combining and optimising the multi-rail protocol introduced in [44] and the qudit P1-or-P2 protocol introduced in [41]. To begin, we will establish some notation and we will prove a useful lemma. For any $K \in \mathbb{N}$ with $K \geq 2$ and any $\mathbf{n} := (n_1, \ldots, n_K) \in \mathbb{N}^K$, we denote as $|\mathbf{n}\rangle_{A_1 \ldots A_K}$ the following K-mode Fock state with total photon number equal to $\|\mathbf{n}\|_1$:

$$|\mathbf{n}\rangle_{A_1 \ldots A_K} := |n_1\rangle_{A_1} \otimes |n_2\rangle_{A_2} \otimes \ldots \otimes |n_K\rangle_{A_K} , \tag{5.203}$$

where we have used the notation $\|\mathbf{n}\|_1 := \sum_{j=1}^K n_j$. For any $N, K \in \mathbb{N}^+$ with $K \geq 2$, let us order the set

$$\{|\mathbf{n}\rangle_{A_1 \ldots A_K} : \mathbf{n} \in \mathbb{N}^K , \|\mathbf{n}\|_1 = N\} \tag{5.204}$$

according to the restricted lexicographic ordering. More formally, the relation $\preceq$ is defined as

$$|\mathbf{n}\rangle_{A_1 \ldots A_K} \preceq |\mathbf{m}\rangle_{A_1 \ldots A_K} \iff \sum_{j=1}^K n_j \, (N+1)^j < \sum_{j=1}^K m_j \, (N+1)^j . \tag{5.205}$$

The set has $\binom{N+K-1}{N}$ elements, and for all $n = 0, 1, \ldots, \binom{N+K-1}{N}-1$, we define the state $|\phi_n^{(N)}\rangle_{A_1 \ldots A_K}$ as the nth element of the ordered set. For example, if $N = 2$ and $K = 3$, we have that

$$\begin{aligned}
|\phi_0^{(2)}\rangle_{A_1A_2A_3} &:= |0\rangle_{A_1} \otimes |0\rangle_{A_2} \otimes |2\rangle_{A_3} \ , \\
|\phi_1^{(2)}\rangle_{A_1A_2A_3} &:= |0\rangle_{A_1} \otimes |1\rangle_{A_2} \otimes |1\rangle_{A_3} \ , \\
|\phi_2^{(2)}\rangle_{A_1A_2A_3} &:= |0\rangle_{A_1} \otimes |2\rangle_{A_2} \otimes |0\rangle_{A_3} \ , \\
|\phi_3^{(2)}\rangle_{A_1A_2A_3} &:= |1\rangle_{A_1} \otimes |0\rangle_{A_2} \otimes |1\rangle_{A_3} \ , \\
|\phi_4^{(2)}\rangle_{A_1A_2A_3} &:= |1\rangle_{A_1} \otimes |1\rangle_{A_2} \otimes |0\rangle_{A_3} \ , \\
|\phi_5^{(2)}\rangle_{A_1A_2A_3} &:= |2\rangle_{A_1} \otimes |0\rangle_{A_2} \otimes |0\rangle_{A_3} \ .
\end{aligned} \tag{5.206}$$

In addition, for all $N, K \in \mathbb{N}^+$ with $K \geq 2$ let us define the following state of $K + K$ modes $A_1, \ldots, A_k, A'_1, \ldots, A'_K$:

$$\begin{aligned}
|\Psi_{N,K}\rangle_{A_1 \ldots A_K, A'_1, \ldots, A'_K} &:= \frac{1}{\sqrt{\binom{N+K-1}{N}}} \sum_{n=0}^{\binom{N+K-1}{N}-1} |\phi_n^{(N)}\rangle_{A_1 \ldots A_k} \otimes |\phi_n^{(N)}\rangle_{A'_1 \ldots A'_k} \\
&= \frac{1}{\sqrt{\binom{N+K-1}{N}}} \sum_{\substack{\mathbf{n} \in \mathbb{N}^K \\ \|\mathbf{n}\|_1 = N}} |\mathbf{n}\rangle_{A_1 \ldots A_K} \otimes |\mathbf{n}\rangle_{A'_1 \ldots A'_K} \ ,
\end{aligned} \tag{5.207}$$

which is a $\binom{N+K-1}{N}$-dimensional maximally entangled state that corresponds to the subspace of the Hilbert space of K modes with total photon number equal to N. Moreover, let us define for all $K, F \in \mathbb{N}$ the projector $\Pi_F^{(K)}$ onto the subspace of K modes $A_1, \ldots, A_K$ whose total photon number equals F, i.e.

$$\Pi_F^{(K)} := \sum_{\substack{\mathbf{m} \in \mathbb{N}^K \\ \|\mathbf{m}\|_1 = F}} |\mathbf{m}\rangle\langle\mathbf{m}|_{A_1 \ldots A_K} \ . \tag{5.208}$$

The following lemma will be useful in order to calculate the rate of our entanglement distribution protocol.

Lemma 19. *Let $\lambda \in [0, 1], g \geq 1, N \in \mathbb{N}, K \in \mathbb{N}^+$, and $\boldsymbol{n} \in \mathbb{N}^K$ such that $\|\boldsymbol{n}\|_1 = N$. Assume that Alice transmits the K-mode Fock state $|\boldsymbol{n}\rangle$ to Bob via K parallel uses of the piBGC $\mathcal{N}_{g,\lambda} := \Phi_{g,0} \circ \mathscr{E}_{\lambda,0}$ and suppose further that Bob measures the total photon number of the K received modes. The probability $\mathscr{P}_F$ that Bob gets the outcome $F \in \mathbb{N}$ is*

$$\mathscr{P}_F := \operatorname{Tr}\left[\mathcal{N}_{g,\lambda}^{\otimes K}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\,\Pi_F^{(K)}\right] = \sum_{P=0}^{\min(F,N)} \binom{N}{P}\binom{K+F-1}{F-P}\lambda^P(1-\lambda)^{N-P}\frac{(g-1)^{F-P}}{g^{K+F}} \ . \tag{5.209}$$

In particular, note that $\mathscr{P}_F$ depends on $\boldsymbol{n}$ only through the total photon number $\|\boldsymbol{n}\|_1 = N$. Specifically, if the communication channel is the pure loss channel $\mathscr{E}_{\lambda,0} = \mathcal{N}_{1,\lambda}$, the probability of getting the outcome $F \in \mathbb{N}$ is

$$\operatorname{Tr}\left[\mathscr{E}_{\lambda,0}^{\otimes K}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\,\Pi_F^{(K)}\right] = \binom{N}{F}\lambda^F(1-\lambda)^{N-F}\Theta(N-F) \, , \tag{5.210}$$

where we have introduced the Heaviside function $\Theta(x)$ defined as $\Theta(x) = 1$ if $x \geq 0$, and $\Theta(x) = 0$ if $x < 0$. In addition, if the communication channel is the pure amplifier channel $\Phi_{g,0} = \mathcal{N}_{g,1}$, the probability of getting the outcome $F \in \mathbb{N}$ is

$$\operatorname{Tr}\left[\Phi_{g,0}^{\otimes K}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\,\Pi_F^{(K)}\right] = \binom{K+F-1}{F-N}\frac{(g-1)^{F-N}}{g^{K+F}}\Theta(F-N) \, . \tag{5.211}$$

Therefore, the probability $\mathscr{P}_F$ in (5.209) of getting F photons at the output of K parallel uses of the composition between pure loss channel and pure amplifier channel can be expressed as the sum over $P \in \mathbb{N}$ of the conditional probability of getting F photons at the output of the K pure amplifier channels conditioned on the event of getting P photons at the output of the K pure loss channels, multiplied by the probability of the latter event.

Proof. As a consequence of (5.128), for all $n \in \mathbb{N}$ it holds that

$$\mathscr{E}_{\lambda,0}(|n\rangle\langle n|) = \sum_{l=0}^{n} \binom{n}{l} \lambda^l (1-\lambda)^{n-l} \, |l\rangle\langle l| \tag{5.212}$$

and hence

$$\mathscr{E}_{\lambda,0}^{\otimes K}(|\mathbf{n}\rangle\langle\mathbf{n}|) = \sum_{\substack{\mathbf{l}\in\mathbb{N}^K \\ \mathbf{l}\leq\mathbf{n}}} \left(\prod_{j=1}^{K} \binom{n_j}{l_j} \right) \lambda^{\|\mathbf{l}\|_1} (1-\lambda)^{N-\|\mathbf{l}\|_1} \, |\mathbf{l}\rangle\langle\mathbf{l}| \, , \tag{5.213}$$

where the inequality between vectors $\mathbf{a} \geq \mathbf{b}$ means that $a_j \geq b_j$ for all $j = 1, \ldots, K$. Consequently, by using that $\sum_{P=0}^{\infty} \Pi_P^{(K)} = \mathbb{1}$, it holds that

$$\begin{aligned} \mathscr{P}_F &:= \operatorname{Tr}\left[\Phi_{g,0}^{\otimes K} \left(\mathscr{E}_{\lambda,0}^{\otimes K}(|\mathbf{n}\rangle\langle\mathbf{n}|) \right) \Pi_F^{(K)} \right] = \sum_{P,P'=0}^{\infty} \operatorname{Tr}\left[\Phi_{g,0}^{\otimes K} \left(\Pi_P^{(K)} \mathscr{E}_{\lambda,0}^{\otimes K}(|\mathbf{n}\rangle\langle\mathbf{n}|) \Pi_{P'}^{(K)} \right) \Pi_F^{(K)} \right] \\ &= \sum_{P=0}^{N} \lambda^P (1-\lambda)^{N-P} \sum_{\substack{\mathbf{l}\in\mathbb{N}^K \\ \mathbf{l}\leq\mathbf{n} \\ \|\mathbf{l}\|_1 = P}} \left(\prod_{j=1}^{K} \binom{n_j}{l_j} \right) \operatorname{Tr}\left[\Phi_{g,0}^{\otimes K} \left(|\mathbf{l}\rangle\langle\mathbf{l}| \right) \Pi_F^{(K)} \right] . \end{aligned} \tag{5.214}$$

Moreover, (5.129) implies that for all $l \in \mathbb{N}$ it holds that

$$\Phi_{g,0}(|l\rangle\langle l|) = \frac{1}{g^{l+1}} \sum_{m=0}^{\infty} \binom{l+m}{l} \left(\frac{g-1}{g} \right)^m |m+l\rangle\langle m+l| \tag{5.215}$$

and hence

$$\Phi_{g,0}^{\otimes K}(|\mathbf{l}\rangle\langle\mathbf{l}|) = \frac{1}{g^{P+K}} \sum_{\mathbf{m}\in\mathbb{N}^K} \left(\prod_{j=1}^{K} \binom{l_j+m_j}{l_j} \right) \left(\frac{g-1}{g} \right)^{\|\mathbf{m}\|_1} |\mathbf{m}+\mathbf{l}\rangle\langle\mathbf{m}+\mathbf{l}| \, . \tag{5.216}$$

Consequently, it holds that

$$\operatorname{Tr}\left[\Phi_{g,0}^{\otimes K} \left(|\mathbf{l}\rangle\langle\mathbf{l}| \right) \Pi_F^{(K)} \right] = \frac{(g-1)^{F-P}}{g^{K+F}} \sum_{\substack{\mathbf{m}\in\mathbb{N}^K \\ \|\mathbf{m}\|_1 = F-P}} \prod_{j=1}^{K} \binom{l_j+m_j}{l_j} . \tag{5.217}$$

The sum

$$\sum_{\substack{\mathbf{m}\in\mathbb{N}^K \\ \|\mathbf{m}\|_1 = F-P}} \prod_{j=1}^{K} \binom{l_j+m_j}{l_j} , \tag{5.218}$$

which appears in (5.217), is the coefficient of the term x^{F-P} of the power series $Q(x)$ in the variable $x \in (0,1)$ defined as

$$Q(x) := \sum_{\mathbf{m}\in\mathbb{N}^K} \left(\prod_{j=1}^{K} \binom{l_j+m_j}{l_j} \right) x^{\|\mathbf{m}\|_1} . \tag{5.219}$$

By exploiting that for all $l \in \mathbb{N}$ it holds that

$$\sum_{m=0}^{\infty} \binom{m+l}{m} x^m = \frac{1}{(1-x)^{l+1}} , \tag{5.220}$$

one obtains that

$$Q(x) = \sum_{\mathbf{m}\in\mathbb{N}^K} \left(\prod_{j=1}^K \binom{l_j+m_j}{l_j}\right) x^{\|\mathbf{m}\|_1} = \prod_{j=1}^K \left(\sum_{m=0}^\infty \binom{l_j+m}{m} x^m\right) = \prod_{j=1}^K \frac{1}{(1-x)^{l_j+1}} = \frac{1}{(1-x)^{P+K}}$$
$$= \sum_{m=0}^\infty \binom{P+K+m-1}{m} x^m \,. \tag{5.221}$$

It follows that

$$\sum_{\substack{\mathbf{m}\in\mathbb{N}^K \\ \|\mathbf{m}\|_1=F-P}} \prod_{j=1}^K \binom{l_j+m_j}{l_j} = \binom{K+F-1}{F-P}\Theta(F-P) \tag{5.222}$$

and hence

$$\mathrm{Tr}\left[\Phi_{g,0}^{\otimes K}\left(|\mathbf{l}\rangle\langle\mathbf{l}|\right)\Pi_F^{(K)}\right] = \frac{(g-1)^{F-P}}{g^{K+F}}\binom{K+F-1}{F-P}\Theta(F-P)\,, \tag{5.223}$$

where we have introduced the Heaviside function $\Theta(x)$ defined as $\Theta(x)=1$ if $x\geq 0$, and $\Theta(x)=0$ if $x<0$. Consequently, (5.214) implies that

$$\mathscr{P}_F = \sum_{P=0}^{\min(F,N)} \lambda^P(1-\lambda)^{N-P}\frac{(g-1)^{F-P}}{g^{K+F}}\binom{K+F-1}{F-P} \sum_{\substack{\mathbf{l}\in\mathbb{N}^K \\ \mathbf{l}\leq\mathbf{n} \\ \|\mathbf{l}\|_1=P}} \left(\prod_{j=1}^K \binom{n_j}{l_j}\right)$$
$$= \sum_{P=0}^{\min(F,N)} \binom{N}{P}\lambda^P(1-\lambda)^{N-P}\frac{(g-1)^{F-P}}{g^{K+F}}\binom{K+F-1}{F-P}\,, \tag{5.224}$$

where in the last equality we have exploited that

$$\sum_{\substack{\mathbf{l}\in\mathbb{N}^K \\ \mathbf{l}\leq\mathbf{n} \\ \|\mathbf{l}\|_1=P}} \left(\prod_{j=1}^K \binom{n_j}{l_j}\right) = \binom{N}{P}\,. \tag{5.225}$$

This follows from the fact that the sum in (5.225) is equal to the coefficient of the term x^P of the following polynomial in the variable $x\in\mathbb{R}$:

$$\sum_{\substack{\mathbf{l}\in\mathbb{N}^K \\ \mathbf{l}\leq\mathbf{n}}} \left(\prod_{j=1}^K \binom{n_j}{l_j}\right) x^{\|\mathbf{l}\|_1} = \prod_{j=1}^K (1+x)^{n_j} = (1+x)^N = \sum_{l=0}^N \binom{N}{l} x^l\,. \tag{5.226}$$

□

Remark 20. *Here we present an alternative method to calculate the probability $\mathscr{P}_F$ reported in (5.209). For all $x\in(0,1)$ let us consider the tensor product of K thermal states with mean photon number $\frac{x}{1-x}$, i.e.*

$$\tau_{\frac{x}{1-x}}^{\otimes K} = (1-x)^K \sum_{\boldsymbol{l}\in\mathbb{N}^K} x^{\|\boldsymbol{l}\|_1}\,|\boldsymbol{l}\rangle\langle\boldsymbol{l}|\,. \tag{5.227}$$

Consequently, the quantity

$$\mathscr{P}_F = \mathrm{Tr}\left[\mathscr{M}_{g,\lambda}^{\otimes K}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\,\Pi_F^{(K)}\right] \tag{5.228}$$

is the coefficient of the term x^F of the power series $P(x)$ in the variable $x\in(0,1)$ defined as

$$P(x) := \frac{1}{(1-x)^K}\,\mathrm{Tr}\left[\mathscr{M}_{g,\lambda}^{\otimes K}(|\boldsymbol{n}\rangle\langle\boldsymbol{n}|)\,\tau_{\frac{x}{1-x}}^{\otimes K}\right]\,. \tag{5.229}$$

By using the characteristic function properties reported in (7.33), (5.33), (5.102), and the fact that the characteristic function of a thermal state τ_ν is $\chi_{\tau_\nu}(\mathbf{r}) = e^{-\frac{1}{4}(2\nu+1)|\mathbf{r}|^2}$, one obtains that for any single-mode state ρ it holds that

$$\begin{aligned}\mathrm{Tr}\left[\mathcal{N}_{g,\lambda}(\rho)\,\tau_{\frac{x}{1-x}}\right] &= \int_{\mathbb{R}^2}\frac{\mathrm{d}^2\mathbf{r}}{2\pi}\chi_{\mathcal{N}_{g,\lambda}(\rho)}(\mathbf{r})\,\chi_{\tau_{\frac{x}{1-x}}}(\mathbf{r}) = \int_{\mathbb{R}^2}\frac{\mathrm{d}^2\mathbf{r}}{2\pi}\chi_\rho(\sqrt{g\lambda}\,\mathbf{r})e^{-\frac{1}{4}\left(2g-g\lambda+2\frac{x}{1-x}\right)|\mathbf{r}|^2}\\ &= \frac{1}{g\lambda}\int_{\mathbb{R}^2}\frac{\mathrm{d}^2\mathbf{r}}{2\pi}\chi_\rho(\mathbf{r})e^{-\frac{1}{4g\lambda}\left(2g-g\lambda+2\frac{x}{1-x}\right)|\mathbf{r}|^2} = \frac{1}{g\lambda}\,\mathrm{Tr}\left[\rho\;\tau_{\frac{g-g\lambda+(1+g\lambda-g)x}{g\lambda(1-x)}}\right].\end{aligned} \tag{5.230}$$

Hence, by exploiting (5.220) and the fact that $\|\boldsymbol{n}\|_1 = N$, the power series $P(x)$ can be expressed as

$$\begin{aligned}P(x) &= \frac{1}{(1-x)^K(g\lambda)^K}\,\mathrm{Tr}\left[|\boldsymbol{n}\rangle\langle\boldsymbol{n}|\,\tau^{\otimes K}_{\frac{g-g\lambda+(1+g\lambda-g)x}{g\lambda(1-x)}}\right] = \frac{[g(1-\lambda)+(1+g\lambda-g)x]^N}{[g-(g-1)x]^{N+K}}\\ &= \sum_{P=0}^{N}\binom{N}{P}(1+g\lambda-g)^P(1-\lambda)^{N-P}g^{-P-K}x^P\sum_{l=0}^{\infty}\binom{N+K-1+l}{l}\left(\frac{g-1}{g}\right)^l x^l.\end{aligned} \tag{5.231}$$

It follows that

$$\mathscr{P}_F = \sum_{P=0}^{\min(F,N)}\binom{N}{P}\binom{N+K+F-P-1}{F-P}(1+g\lambda-g)^P(1-\lambda)^{N-P}\frac{(g-1)^{F-P}}{g^{F+K}}. \tag{5.232}$$

Incidentally, by comparing the two expressions of $\mathscr{P}_F$ in (5.209) and (5.232), one deduces the following identity:

$$\sum_{P=0}^{\min(F,N)}\binom{N}{P}\binom{K+F-1+N-P}{F-P}\left(\frac{g\lambda-(g-1)}{(1-\lambda)(g-1)}\right)^P = \sum_{P=0}^{\min(F,N)}\binom{N}{P}\binom{K+F-1}{F-P}\left(\frac{\lambda}{(1-\lambda)(g-1)}\right)^P. \tag{5.233}$$

Let us now introduce an additional entanglement distribution protocol to distribute ebits across any piBGC $\mathcal{N}_{g,\lambda}$. The protocol depends on two parameters, $K, N \in \mathbb{N}^+$ with $K \geq 2$, and it is composed of five steps named S1-S5, which we now outline.

S1: Alice prepares the state $|\Psi_{N,K}\rangle_{A_1\ldots A_K,A'_1,\ldots,A'_K}$ of $K+K$ modes $A_1,\ldots,A_k,A'_1,\ldots,A'_K$, sending the systems $A'_1,\ldots,A'_K$ to Bob through K uses of the channel $\mathcal{N}_{g,\lambda}$. Now Alice and Bob share the state $\mathrm{Id}_{A_1\ldots A_k}\otimes\mathcal{N}^{\otimes K}_{g,\lambda}(|\Psi_{N,K}\rangle\langle\Psi_{N,K}|)$. By using (5.108), such a state can be expressed as

$$\begin{aligned}&\mathrm{Id}_{A_1\ldots A_k}\otimes\mathcal{N}^{\otimes K}_{g,\lambda}(|\Psi_{N,K}\rangle\langle\Psi_{N,K}|)\\ &= \frac{1}{\binom{N+K-1}{N}}\sum_{\substack{\mathbf{n}\in\mathbb{N}^K\\ \|\mathbf{n}\|_1=N}}\sum_{\substack{\mathbf{i}\in\mathbb{N}^K\\ \|\mathbf{i}\|_1=N}}\sum_{\substack{\mathbf{l}\in\mathbb{N}^K\\ \mathbf{l}\geq\max(\mathbf{i}-\mathbf{n},\mathbf{0})}}\left(\prod_{j=1}^{K}f_{n_j,i_j,l_j}(g,\lambda)\right)|\mathbf{n}\rangle\langle\mathbf{i}|_{A_1\ldots A_K}\otimes|\mathbf{l}+\mathbf{n}-\mathbf{i}\rangle\langle\mathbf{l}|_{B_1\ldots B_K},\end{aligned} \tag{5.234}$$

where $\mathbf{0}\in\mathbb{N}^K$ is the zero vector and the inequality between vectors $\mathbf{a}\geq\mathbf{b}$ means that $a_j \geq b_j$ for all $j = 1,\ldots,K$.

S2: Bob performs the local POVM $\{\Pi^{(K)}_F\}_{F\in\mathbb{N}}$, where $\Pi^{(K)}_F$ is the projector onto the subspace whose total photon number equals F (see (5.208)), on the K modes he has received. The probability of getting the outcome F is denoted by $\mathscr{P}_F$ and it can be calculated as

$$\begin{aligned}\mathscr{P}_F &:= \mathrm{Tr}\left[\left(\mathbb{1}_{A_1\ldots A_k}\otimes\Pi^{(K)}_F\right)\left(\mathrm{Id}_{A_1\ldots A_k}\otimes\mathcal{N}^{\otimes K}_{g,\lambda}(|\Psi_{N,K}\rangle\langle\Psi_{N,K}|)\right)\right] = \frac{1}{\binom{N+K-1}{N}}\sum_{\substack{\mathbf{n}\in\mathbb{N}^K\\ \|\mathbf{n}\|_1=N}}\mathrm{Tr}\left[\Pi^{(K)}_F\,\mathcal{N}^{\otimes K}_{g,\lambda}(|\mathbf{n}\rangle\langle\mathbf{n}|)\right]\\ &= \sum_{P=0}^{\min(F,N)}\tbinom{N}{P}\tbinom{K+F-1}{F-P}\lambda^P(1-\lambda)^{N-P}\frac{(g-1)^{F-P}}{g^{K+F}},\end{aligned} \tag{5.235}$$

where we have exploited Lemma 19. The post-measurement state $\rho^{(F)}_{A_1\ldots A_k B_1\ldots B_k}$ conditioned on the outcome $F \in \mathbb{N}$ is given by

$$\begin{aligned}
&\rho^{(F)}_{A_1\ldots A_k B_1\ldots B_k} \\
&= \frac{1}{\mathscr{P}_F} \left(\mathbb{1}_{A_1\ldots A_k} \otimes \Pi_F^{(K)}\right) \left(\mathrm{Id}_{A_1\ldots A_k} \otimes \mathscr{N}_{g,\lambda}^{\otimes K}(|\Psi_{N,K}\rangle\langle\Psi_{N,K}|)\right) \left(\mathbb{1}_{A_1\ldots A_k} \otimes \Pi_F^{(K)}\right) \\
&= \frac{1}{\mathscr{P}_F \binom{N+K-1}{N}} \sum_{\substack{\mathbf{n}\in\mathbb{N}^K \\ \|\mathbf{n}\|_1 = N}} \sum_{\substack{\mathbf{i}\in\mathbb{N}^K \\ \|\mathbf{i}\|_1 = N}} \sum_{\substack{\mathbf{l}\in\mathbb{N}^K \\ \|\mathbf{l}\|_1 = F \\ \mathbf{l}\geq \max(\mathbf{i}-\mathbf{n},\mathbf{0})}} \left(\prod_{j=1}^K f_{n_j,i_j,l_j}(g,\lambda)\right) |\mathbf{n}\rangle\langle\mathbf{i}|_{A_1\ldots A_K} \otimes |\mathbf{l}+\mathbf{n}-\mathbf{i}\rangle\langle\mathbf{l}|_{B_1\ldots B_K} \\
&= \sum_{n,i=0}^{\binom{N+K-1}{N}-1} \sum_{h,l=0}^{\binom{F+K-1}{F}-1} c_{n,i,h,l}\, |\phi_n^{(N)}\rangle\langle\phi_i^{(N)}|_{A_1\ldots A_k} \otimes |\phi_h^{(F)}\rangle\langle\phi_l^{(F)}|_{B_1\ldots B_k} \,,
\end{aligned} \tag{5.236}$$

where for all $n, i = 0, 1, \ldots, \binom{N+K-1}{N} - 1$ and all $h, l = 0, 1, \ldots, \binom{F+K-1}{F} - 1$ the coefficient $c_{n,i,h,l}$ is defined as follows. Let $\mathbf{n}, \mathbf{i}, \mathbf{h}, \mathbf{l} \in \mathbb{N}^K$ such that $|\phi_n^{(N)}\rangle = |\mathbf{n}\rangle$, $|\phi_i^{(N)}\rangle = |\mathbf{i}\rangle$, $|\phi_h^{(F)}\rangle = |\mathbf{h}\rangle$, and $|\phi_l^{(F)}\rangle = |\mathbf{l}\rangle$. If $\mathbf{l} \geq \max(\mathbf{i}-\mathbf{n}, \mathbf{0})$ and $\mathbf{h} = \mathbf{l}+\mathbf{n}-\mathbf{i}$, then

$$c_{n,i,h,l} := \frac{\left(\prod_{j=1}^K f_{n_j,i_j,l_j}(g,\lambda)\right)}{\mathscr{P}_F \binom{N+K-1}{N}} \,, \tag{5.237}$$

otherwise $c_{n,i,h,l} = 0$. By setting

$$d := \max\left(\binom{N+K-1}{N}, \binom{F+K-1}{F}\right) \,, \tag{5.238}$$

the resulting state in (5.236) can be seen as a bipartite two-qu*d*it state $\rho_{AB}^{(F)} \in \mathfrak{S}(\mathcal{H}_d \otimes \mathcal{H}_d)$ of the form

$$\rho_{AB}^{(F)} = \sum_{n,i,h,l=0}^{d-1} \eta_{n,i,h,l}\, |n\rangle\langle i|_A \otimes |h\rangle\langle l|_B \,, \tag{5.239}$$

where $\mathcal{H}_d$ is the qu*d*it Hilbert space with $\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$ as an orthonormal basis, and where the coefficients $\eta_{n,i,h,l}$ are defined as follows:

- if $n, i \leq \binom{N+K-1}{N} - 1$ and $h, l \leq \binom{F+K-1}{F} - 1$, then $\eta_{n,i,h,l} := c_{n,i,h,l}$;
- otherwise, $\eta_{n,i,h,l} := 0$.

Consequently, Alice and Bob have reduced the problem in distilling ebits from the two-qu*d*it state $\rho_{AB}^{(F)}$.

S3: Now Alice and Bob decide whether or not to run the reverse hashing protocol, which can distil ebits from $\rho_{AB}^{(F)}$ with a rate equal to its reverse coherent information, i.e.

$$I_{\mathrm{rc}}(\rho_{AB}^{(F)}) = S(\mathrm{Tr}_B\, \rho_{AB}^{(F)}) - S(\rho_{AB}^{(F)}) \,, \tag{5.240}$$

where $S(\cdot)$ denotes the von Neumann entropy. By exploiting that

$$\mathrm{Tr}_B\, \rho_{AB}^{(F)} = \frac{1}{\binom{N+K-1}{N}} \sum_{n=0}^{\binom{N+K-1}{N}-1} |n\rangle\langle n| \,, \tag{5.241}$$

as guaranteed by (5.236) and Lemma 19, it follows that the reverse coherent information can be calculated as

$$I_{\mathrm{rc}}(\rho_{AB}^{(F)}) = \log_2 \binom{N+K-1}{N} - S\left(\sum_{n,i,h,l=0}^{d-1} \eta_{n,i,h,l}\, |n\rangle\langle i| \otimes |h\rangle\langle l|\right) \,. \tag{5.242}$$

If Alice and Bob choose to run the reverse hashing protocol, the protocol terminates. Otherwise, they apply the qu*d*it Pauli-based twirling reported in [41, Eq. (18)] in order to transform their state in a Bell-diagonal state of the form

$$\rho_{AB}^{\prime(F)} = \sum_{m,n=0}^{d} \alpha_{mn}^{(F,0)} \, |\psi_{mn}^{(d)}\rangle\langle\psi_{mn}^{(d)}|_{AB} \, , \tag{5.243}$$

where

$$|\psi_{mn}^{(d)}\rangle_{AB} := \frac{1}{\sqrt{d}} \sum_{r=0}^{d-1} e^{i\frac{2\pi mr}{d}} \, |r\rangle_A \otimes |(r-n) \bmod d\rangle_B \tag{5.244}$$

and

$$\alpha_{mn}^{(F,0)} := \langle\psi_{mn}^{(d)}| \, \rho_{AB}^{(F)} \, |\psi_{mn}^{(d)}\rangle = \frac{1}{d} \sum_{r_1,r_2=0}^{d-1} \cos\left(\frac{2\pi m(r_2-r_1)}{d}\right) \eta_{r_1,r_2,(r_1-n) \bmod d,(r_2-n) \bmod d} \, . \tag{5.245}$$

S4: Alice and Bob run $\bar{k}$ times the P1-or-P2 sub-routine for qu*d*its [41], where $\bar{k}$ is chosen in order to maximise the ebit rate. The goal of this step is to bring the shared state closer to the d-dimensional maximally-entangled state $|\psi_{00}^{(d)}\rangle$. This step is successful, i.e. the protocol is not aborted, with a probability of success equal to $\prod_{t=0}^{\bar{k}-1} P_t^{(F)}$ and it allows Alice and Bob to transform $2^{\bar{k}}$ copies of $\rho_{AB}^{\prime(F)} = \sum_{m,n=0}^{d} \alpha_{mn}^{(F,0)} \, |\psi_{mn}^{(d)}\rangle\langle\psi_{mn}^{(d)}|_{AB}$ in a state of the form

$$\rho_{AB}^{\prime(F,\bar{k})} := \sum_{m,n=0}^{d} \alpha_{mn}^{(F,\bar{k})} \, |\psi_{mn}^{(d)}\rangle\langle\psi_{mn}^{(d)}|_{AB} \, . \tag{5.246}$$

For all $t \in \{0,1,\ldots,\bar{k}-1\}$ and all $m,n \in \{0,1,\ldots,d-1\}$ the coefficients $\alpha_{mn}^{(F,t+1)}$ and the probabilities $P_t^{(F)}$ are recursively defined in the following way [41]:

- If $\sum_{m_1=0}^{d-1} \alpha_{m_1 0}^{(F,t)} < \sum_{n_1=0}^{d-1} \alpha_{0 n_1}^{(F,t)}$, then

$$\alpha_{mn}^{(F,t+1)} := \frac{1}{P_t^{(F)}} \sum_{\substack{m_1,m_2=0 \\ (m_1+m_2) \bmod d = m}}^{d-1} \alpha_{m_1 n}^{(F,t)} \alpha_{m_2 n}^{(F,t)} \, , \tag{5.247}$$

  where

$$P_t^{(F)} := \sum_{m_1,m_2,n=0}^{d-1} \alpha_{m_1 n}^{(F,t)} \alpha_{m_2 n}^{(F,t)} \, . \tag{5.248}$$

- Otherwise,

$$\alpha_{mn}^{(F,t+1)} := \frac{1}{P_t^{(F)}} \sum_{\substack{n_1,n_2=0 \\ (n_1+n_2) \bmod d = n}}^{d-1} \alpha_{m n_1}^{(F,t)} \alpha_{m n_2}^{(F,t)} \, , \tag{5.249}$$

  where

$$P_t^{(F)} := \sum_{m,n_1,n_2=0}^{d-1} \alpha_{m n_1}^{(F,t)} \alpha_{m n_2}^{(F,t)} \, . \tag{5.250}$$

S5: Alice and Bob distil ebits from the state $\rho_{AB}^{\prime(F,\bar{k})} = \sum_{m,n=0}^{d} \alpha_{mn}^{(F,\bar{k})} \, |\psi_{mn}^{(d)}\rangle\langle\psi_{mn}^{(d)}|_{AB}$ with a yield denoted as $\mathscr{J}_d(\alpha^{(F,\bar{k})})$ by running the following protocol:

- If $d = 2$, then Alice and Bob run the Step 5 and Step 6 of the entanglement distribution protocol introduced in the proof of Theorem 14 in order to distil ebits from $\rho_{AB}^{\prime(F,\bar{k})}$ with a yield equal to

$$\mathscr{J}_2(\alpha^{(F,\bar{k})}) := \mathscr{J}(\alpha_{00}^{(F,\bar{k})}, \alpha_{01}^{(F,\bar{k})}, \alpha_{10}^{(F,\bar{k})}, \alpha_{11}^{(F,\bar{k})}) \, , \tag{5.251}$$

  where $\mathscr{J}$ is defined in (5.160).

- If $d > 2$, then Alice and Bob run the hashing protocol on $\rho_{AB}^{\prime(F,\bar{k})}$ and thus they distil ebits with a yield equal to the coherent information of $\rho_{AB}^{\prime(F,\bar{k})}$, i.e.

$$\mathscr{I}_d(\alpha^{(F,\bar{k})}) := I_{\mathrm{c}}(\rho_{AB}^{\prime(F,\bar{k})}) = \log_2 d + \sum_{m,n=0}^{d-1} \alpha_{mn}^{(F,\bar{k})} \log_2 \alpha_{mn}^{(F,\bar{k})} \,. \tag{5.252}$$

The ebit rate of the protocol is given by

$$R(g,\lambda,N,K) := \frac{1}{K} \sum_{F=1}^{\infty} \mathscr{P}_F \max\left(I_{\mathrm{rc}}(\rho_{AB}^{(F)}) \,,\, \sup_{\bar{k}\in\mathbb{N}} \frac{\prod_{t=0}^{\bar{k}-1} P_t^{(F)}}{2^{\bar{k}}} \mathscr{I}_d(\alpha^{(F,\bar{k})}) \right) . \tag{5.253}$$

The term $\frac{1}{K}$ in the expression (5.253) arises from the fact that Alice uses the channel K times during step S1, and the variable F corresponds to the outcome of the total photon number measurement in step S2, with associated probability $\mathscr{P}_F$. The sum over F equals the expected value of the yield of ebits that can be distilled from the post-measurement state $\rho_{AB}^{(F)}$ by running steps S3, S4, and S5. The maximum comes from the fact that during step S3 Alice and Bob choose whether or not to run the reverse hashing protocol, which can distil ebits with a rate equal to $I_{\mathrm{rc}}(\rho_{AB}^{(F)})$. The supremum over $\bar{k}$ comes from the fact that Alice and Bob choose the number of iterations $\bar{k}$ of the P1-or-P2 subroutine in order to maximise the rate. The rate in (5.253) is a lower bound on the two-way quantum capacity of the piBGC $\mathscr{N}_{g,\lambda}$ for all $N, K \in \mathbb{N}^+$ with $K \geq 2$. Therefore, we have

$$K(\mathscr{N}_{g,\lambda}) \geq Q_2(\mathscr{N}_{g,\lambda}) \geq \sup_{\substack{N,K\in\mathbb{N}^+ \\ K\geq 2}} R(g,\lambda,N,K) \,. \tag{5.254}$$

Let us summarise this result in the following theorem.

Theorem 21. *For all $\lambda \in [0,1]$ and $g \geq 1$ the secret-key capacity $K(\mathscr{N}_{g,\lambda})$ and the two-way quantum capacity $Q_2(\mathscr{N}_{g,\lambda})$ of the piBGC $\mathscr{N}_{g,\lambda}$ satisfy*

$$K(\mathscr{N}_{g,\lambda}) \geq Q_2(\mathscr{N}_{g,\lambda}) \geq \sup_{\substack{N,K\in\mathbb{N}^+ \\ K\geq 2}} R(g,\lambda,N,K) \,, \tag{5.255}$$

where

$$R(g,\lambda,N,K) := \frac{1}{K} \sum_{F=1}^{\infty} \mathscr{P}_F \max\left(I_{rc}^{(F)} \,,\, \sup_{\bar{k}\in\mathbb{N}} \frac{\prod_{t=0}^{\bar{k}-1} P_t^{(F)}}{2^{\bar{k}}} \mathscr{I}_d(\alpha^{(F,\bar{k})}) \right) . \tag{5.256}$$

The quantities present in (5.256) are defined as follows. For all $F \in \mathbb{N}$ the dimension d is defined as

$$d := \max\left(\binom{N+K-1}{N}, \binom{F+K-1}{F} \right) \tag{5.257}$$

and the probability $\mathscr{P}_F$ is defined as

$$\mathscr{P}_F := \sum_{P=0}^{\min(F,N)} \binom{N}{P} \binom{K+F-1}{F-P} \lambda^P (1-\lambda)^{N-P} \frac{(g-1)^{F-P}}{g^{K+F}} \,. \tag{5.258}$$

Moreover, the probabilities $P_{\bar{k}}^{(F)}$ and the coefficients $\{\alpha_{mn}^{(F,\bar{k})}\}_{m,n\in\{0,1,\ldots,d-1\}}$ are recursively defined as follows. For all $t \in \{0,1,\ldots,\bar{k}-1\}$ and all $m,n \in \{0,1,\ldots,d-1\}$ it holds that:

- *If $\sum_{m_1=0}^{d-1} \alpha_{m_1 0}^{(F,t)} < \sum_{n_1=0}^{d-1} \alpha_{0 n_1}^{(F,t)}$, then*

$$\begin{aligned} \alpha_{mn}^{(F,t+1)} &:= \frac{1}{P_t^{(F)}} \sum_{\substack{m_1,m_2=0 \\ (m_1+m_2) \bmod d = m}}^{d-1} \alpha_{m_1 n}^{(F,t)} \alpha_{m_2 n}^{(F,t)} \,, \\ P_t^{(F)} &:= \sum_{m_1,m_2,n=0}^{d-1} \alpha_{m_1 n}^{(F,t)} \alpha_{m_2 n}^{(F,t)} \,. \end{aligned} \tag{5.259}$$

- *Otherwise,*

$$\begin{aligned} \alpha_{mn}^{(F,t+1)} &:= \frac{1}{P_t^{(F)}} \sum_{\substack{n_1,n_2=0 \\ (n_1+n_2)\ mod\ d=n}}^{d-1} \alpha_{mn_1}^{(F,t)} \alpha_{mn_2}^{(F,t)} \,, \\ P_t^{(F)} &:= \sum_{m,n_1,n_2=0}^{d-1} \alpha_{mn_1}^{(F,t)} \alpha_{mn_2}^{(F,t)} \,. \end{aligned} \tag{5.260}$$

Moreover, for all $m, n \in \{0, 1, \ldots, d-1\}$ *the coefficient* $\alpha_{mn}^{(F,0)}$ *is defined as*

$$\alpha_{mn}^{(F,0)} := \frac{1}{d} \sum_{r_1,r_2=0}^{d-1} \cos\left(\frac{2\pi m(r_2 - r_1)}{d}\right) \eta_{r_1,r_2,(r_1-n)\ mod\ d,(r_2-n)\ mod\ d} \,. \tag{5.261}$$

In addition, for all $n, i \in 0, 1, \ldots, \binom{N+K-1}{N} - 1$*, we define* $(n_1, \ldots, n_K)$ *and* $(i_1, \ldots, i_K)$ *as the* n*th and* i*th element of the ordered set* $S_{K,N}$*, where* $S_{K,N}$ *is defined as*

$$S_{K,N} := \{(f_1, \ldots, f_K) \in \mathbb{N}^K : \sum_{j=1}^{K} f_j = N\} \tag{5.262}$$

and it is ordered according to the relation $\preceq_{K,N}$*, given by*

$$(f_1, \ldots, f_K) \preceq_{K,N} (g_1, \ldots, g_K) \iff \sum_{j=1}^{K} f_j \, (N+1)^j < \sum_{j=1}^{K} g_j \, (N+1)^j \,. \tag{5.263}$$

Additionally, for all $h, l \in \{0, 1, \ldots, \binom{F+K-1}{F} - 1\}$*, we define* $(h_1, \ldots, h_K)$ *and* $(l_1, \ldots, l_K)$ *as the* h*th and* l*th element of the set* $S_{K,F}$ *ordered according to the relation* $\preceq_{F,N}$*. Furthermore, for all* $n, i, h, l \in \{0, 1, \ldots, d-1\}$ *the coefficients* $\eta_{n,i,h,l}$ *are defined as follows:*

- *If*

$$\begin{aligned} n, i &\leq \binom{N+K-1}{N} - 1 \,, \\ h, l &\leq \binom{F+K-1}{F} - 1 \,, \\ l_j &\geq \max(i_j - n_j, 0) \quad \textit{for all } j = 1, 2, \ldots, K \,, \\ h_j &= l_j + n_j - i_j \quad \textit{for all } j = 1, 2, \ldots, K \,, \end{aligned} \tag{5.264}$$

  then

$$\eta_{n,i,h,l} := \frac{\left(\prod_{j=1}^{K} f_{n_j,i_j,l_j}(g,\lambda)\right)}{\mathscr{P}_F\binom{N+K-1}{N}} \,, \tag{5.265}$$

  where $f_{n,i,l}(g,\lambda)$ *is defined in* (5.109)*.*

- *Otherwise,* $\eta_{n,i,h,l} = 0$*.*

Moreover, the quantity $I_{rc}^{(F)}$ *is defined as*

$$I_{rc}^{(F)} := \log_2 \binom{N+K-1}{N} - S\left(\sum_{n,i,h,l=0}^{d-1} \eta_{n,i,h,l} \, |n\rangle\langle i| \otimes |h\rangle\langle l|\right) \,, \tag{5.266}$$

where $S(\cdot)$ *denotes the von Neumann entropy. Finally, the term* $\mathscr{I}_d(\alpha^{(F,\bar{k})})$ *is defined differently depending on the value of* d*:*

- *If $d = 2$, then*
$$\mathscr{J}_2(\alpha^{(F,\bar{k})}) := \mathscr{J}(\alpha_{00}^{(F,\bar{k})}, \alpha_{01}^{(F,\bar{k})}, \alpha_{10}^{(F,\bar{k})}, \alpha_{11}^{(F,\bar{k})}), \tag{5.267}$$
where $\mathscr{J}$ is defined in (5.160).

- *If $d > 2$, then*
$$\mathscr{J}_d(\alpha^{(F,\bar{k})}) := \log_2 d + \sum_{m,n=0}^{d-1} \alpha_{mn}^{(F,\bar{k})} \log_2 \alpha_{mn}^{(F,\bar{k})}. \tag{5.268}$$

Results on the two-way capacities of piBGCs

In this subsection, for each of the piBGCs, first we determine the parameter region where the two-way capacities vanish, second we find a new lower bound on the two-way capacities, and finally we compare our results with the existing literature.

Results on the two-way capacities of the thermal attenuator

Let us consider the thermal attenuator $\mathscr{E}_{\lambda,\nu}$ of transmissivity $\lambda \in [0,1]$ and thermal noise $\nu \geq 0$. Since the PLOB bound in (5.62) vanishes for $\lambda \leq \frac{\nu}{\nu+1}$, it is already known that the two-way capacities of $\mathscr{E}_{\lambda,\nu}$ vanish for $\lambda < \frac{\nu}{\nu+1}$. The following theorem establishes that also the vice-versa is true.

Theorem 22. *Let $\lambda \in [0,1]$, $\nu \geq 0$, and $N_s > 0$. The energy-constrained two-way capacities of the thermal attenuator $Q_2(\mathscr{E}_{\lambda,\nu}, N_s)$ and $K(\mathscr{E}_{\lambda,\nu}, N_s)$ vanish if and only if $\lambda \leq \frac{\nu}{\nu+1}$, i.e. if and only if $\mathscr{E}_{\lambda,\nu}$ is entanglement breaking. In particular, the same holds for the unconstrained two-way capacities.*

Proof. Theorem 22 is a direct consequence of Lemma 5 and Theorem 13. □

The validity of Theorem 22 was not known before the present work. Indeed, in [4, 12] the authors says that it is an open problem to determine the exact value of the maximum tolerable excess noise, which is defined by
$$\epsilon(\lambda) := \frac{1-\lambda}{\lambda} \max\{\nu \geq 0 \,:\, K(\mathscr{E}_{\lambda,\nu}) > 0\}. \tag{5.269}$$
Theorem 13 implies that $\varepsilon(\lambda) = 1$ for all $\lambda \in (0,1)$. Hence, we have answered to the question, which was deemed "crucial" in [12, Section 7], "What is the maximum excess noise that is tolerable in QKD? I.e., optimizing over all QKD protocols?" In [4, 12] the authors showed, by applying the PLOB bound, the upper bound $\varepsilon(\lambda) \leq 1$ and provided also a lower bound on $\varepsilon(\lambda)$ which was far from 1.

Except for the special case $\nu = 0$, it is an open question whether the reverse coherent information lower bound in Eq. 5.64 equals the true two-way quantum capacity of the thermal attenuator $Q_2(\mathscr{E}_{\lambda,\nu})$: Theorem 22 provides a negative answer to this question. Indeed, although $Q_2(\mathscr{E}_{\lambda,\nu}) = 0$ if and only if $\lambda \leq \frac{\nu}{\nu+1}$ (thanks to Theorem 22), the reverse coherent information lower bound vanishes for all $\lambda \leq 1 - 2^{-h(\nu)}$. Hence, since $1 - 2^{-h(\nu)} > \frac{\nu}{\nu+1}$ for all $\nu > 0$, the reverse coherent information lower bound is not equal to $Q_2(\mathscr{E}_{\lambda,\nu})$ at least in the region $\nu > 0$ and $\lambda \in (\frac{\nu}{\nu+1}, 1 - 2^{-h(\nu)}]$. In the following theorem we obtain an improved lower bound on the two-way capacities of the thermal attenuator.

Theorem 23. *Let $\lambda \in [0,1]$, $\nu \geq 0$, and $N_s \geq 0$. The EC two-way capacities $Q_2(\mathscr{E}_{\lambda,\nu}, N_s)$ and $K(\mathscr{E}_{\lambda,\nu}, N_s)$ of the thermal attenuator $\mathscr{E}_{\lambda,\nu}$ satisfy the following lower bound*
$$K(\mathscr{E}_{\lambda,\nu}, N_s) \geq Q_2(\mathscr{E}_{\lambda,\nu}, N_s) \geq \sup_{\substack{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N} \\ (1-c^2)M \leq N_s}} \mathscr{R}\left(1 + (1-\lambda)\nu, \frac{\lambda}{1+(1-\lambda)\nu}, M, c, k\right), \tag{5.270}$$

and, in particular, the unconstrained two-way capacities satisfy

$$K(\mathcal{E}_{\lambda,\nu}) \geq Q_2(\mathcal{E}_{\lambda,\nu}) \geq \sup_{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N}} \mathscr{R}\left(1+(1-\lambda)\nu, \frac{\lambda}{1+(1-\lambda)\nu}, M, c, k\right), \tag{5.271}$$

where the quantity $\mathscr{R}$ is defined in (5.188).

Proof. Theorem 23 is a direct consequence of Lemma 5 and Theorem 15. □

Theorem 23 shows a new lower bound, reported in (5.271), on the two-way capacities of the thermal attenuator $\mathcal{E}_{\lambda,\nu}$. Our new lower bound outperforms all the previous known lower bounds in a large region of the parameters λ and ν. In Fig. 5.2a and in Fig. 5.2b we plot our new bound and its ratio with the PLOB bound, respectively, with respect to ν where the transmissivity is chosen to be equal to $\lambda(\nu) := 1 - 2^{-h(\nu)}$, which is the upper endpoint for the λ-range for which the best known lower bound on $Q_2(\mathcal{E}_{\lambda,\nu})$ (i.e. the reverse coherent information lower bound reported in (5.64)) vanishes. From Fig. 5.2a and Fig. 5.2b we see that for these choices of ν and $\lambda(\nu)$, our new lower bound is now the best lower bound on $Q_2(\mathcal{E}_{\lambda,\nu})$ and it achieves the $\simeq 14\%$ of the PLOB bound for $\nu \gg 1$. For example, if $\nu = 1$ and if the transmissivity is equal to $\lambda = 1 - 2^{-h(1)} = 0.75$, our new lower bound is $\simeq 0.033$, its ratio with the PLOB bound is $\simeq 0.08$, and the optimal parameters of the supremum present in the expression of our new bound in (5.271) are $c \simeq 0.703$, $M = 2$, and $k = 2$. In Fig. 5.3 we plot our new bound on $Q_2(\mathcal{E}_{\lambda,\nu})$ with respect to λ for $\nu = 1$ and $\nu = 10$.

Our new bound can outperform also the best known lower bound (before our work) on the secret-key capacity $K(\mathcal{E}_{\lambda,\nu})$ found by Ottaviani et al. [27]. To demonstrate that our new bound can be strictly tighter than the Ottaviani et al. lower bound, in Fig. 5.4 we plot the latter bound and our new bound with respect to λ for $\nu = 1$ and $\nu = 10$. From Fig. 5.4, we note that the Ottaviani et al. lower bound vanishes for larger transmissivities than our bound. In particular, fixed $\nu > 0$, we numerically observe that our new bound is strictly positive for all $\lambda > \frac{\nu}{\nu+1}$, which is the region where the two-way capacities of $\mathcal{E}_{\lambda,\nu}$ are strictly positive, as established by Theorem 13. As an example, for $\nu = 1$, in Fig. 5.5 we plot the ratio between our bound and the PLOB bound in logarithmic scale and we see that our bound is strictly positive for $\lambda \gtrsim \frac{\nu}{\nu+1} = 0.5$. In addition, fixed $\nu > 0$, we numerically observe that the optimal value of k of the supremum present in the expression of our new bound in (5.271) increases as λ decreases and tends to infinity as λ tends to $\frac{\nu}{\nu+1}$, where we recall that k represents the number of iterations of the P1-or-P2 sub-routine [41] in the entanglement distribution protocol we have introduced in the proof of Theorem 14.

We numerically observe that for all λ and ν the optimal choice of M of the supremum present in the expression of our bound in (5.271) is always less or equal to 3. Hence, since the mean photon number of each signal sent by Alice is $\mathrm{Tr}[a^\dagger a\, |\Psi_{M,c}\rangle\langle\Psi_{M,c}|] = (1-c^2)M$ (see (5.149)), the entanglement distribution protocol we have presented in the proof of Theorem 23 exploits a mean photon number per channel use which is strictly lower than 3. On the contrary, the entanglement distribution protocol which leads to the reverse coherent information lower bound in (5.64) requires infinite mean photon number per channel use, as we reviewed in 5.65.

Theorem 23 shows also the bound in (5.270), which constitutes a new lower bound on the EC two-way capacities of the thermal attenuator $\mathcal{E}_{\lambda,\nu}$. This new lower bound can outperform the NPJ lower bound [26] reported in (5.68), which is the best known lower bound on the EC two-way capacities of the thermal attenuator, as we show in Fig. 5.6 where we plot our new bound in (5.270) with respect to λ for different choices of ν and of the energy constraint N_s.

By using the results of Section 5.3.2, in the forthcoming Theorem 24 we show an additional lower bound on the two-way quantum capacity of the thermal attenuator $\mathcal{E}_{\lambda,\nu}$.

Theorem 24 Multi-rail lower bound**.** *For all $\lambda \in [0,1]$ and $\nu \geq 0$ the two-way capacities of the thermal attenuator $\mathscr{E}_{\lambda,\nu}$ satisfy*

$$K(\mathscr{E}_{\lambda,\nu}) \geq Q_2(\mathscr{E}_{\lambda,\nu}) \geq \sup_{\substack{N,K\in\mathbb{N}^+ \\ K\geq 2}} R\left(1+(1-\lambda)\nu, \frac{\lambda}{1+(1-\lambda)\nu}, N, K\right), \tag{5.272}$$

where the quantity R is defined in (5.256).

Proof. Theorem 24 is a direct consequence of Theorem 21 and Lemma 5. □

Theorem 24 shows an additional lower bound on $Q_2(\mathscr{E}_{\lambda,\nu})$, that we dub 'multi-rail lower bound'. This bound is the ebit rate of the entanglement distribution protocol presented in Section 5.3.2, which combines the multi-rail protocol introduced in [44] and the qudit P1-or-P2 protocol introduced in [41]. In Fig. 5.7 we plot both the multi-rail lower bound (reported in (5.272)) and our previously discussed lower bound (reported in (5.271)) as a function of λ for $\nu = 0.1$, $\nu = 0.5$, $\nu = 1$, and $\nu = 10$. Our numerical investigation shows that for $\nu \lesssim 1$, the multi-rail lower bound is tighter than the previously discussed lower bound, as confirmed by Fig. 5.7.

Results on the two-way capacities of the thermal amplifier

Let us consider the thermal amplifier $\Phi_{g,\nu}$ of gain $g \geq 1$ and thermal noise $\nu \geq 0$. Since the PLOB bound in (5.79) vanishes for $g \geq 1 + \frac{1}{\nu}$, it is already known that the two-way capacities of $\Phi_{g,\nu}$ vanish for $g \geq 1 + \frac{1}{\nu}$. The following theorem establishes that also the vice-versa is true.

Theorem 25. *Let $g \geq 1$, $\nu \geq 0$, and $N_s > 0$. The energy-constrained two-way capacities of the thermal amplifier $Q_2(\Phi_{g,\nu}, N_s)$ and $K(\Phi_{g,\nu}, N_s)$ vanish if and only if $g \geq 1 + \frac{1}{\nu}$, i.e. if and only if $\Phi_{g,\nu}$ is entanglement breaking. In particular, the same holds for the unconstrained two-way capacities.*

Proof. Theorem 25 is a direct consequence of Lemma 5 and Theorem 13. □

Except for the special case $\nu = 0$, it is an open question whether the coherent information lower bound in Eq. 5.80 equals the true two-way quantum capacity of the thermal amplifier $Q_2(\Phi_{g,\nu})$: Theorem 25 provides a negative answer to this question. Indeed, although $Q_2(\Phi_{g,\nu}) = 0$ if and only if $g > 1 + \frac{1}{\nu}$ (thanks to Theorem 25), the coherent information lower bound vanishes for all $g \geq \frac{1}{1-2^{-h(\nu)}}$. Hence, since $1 + \frac{1}{\nu} > \frac{1}{1-2^{-h(\nu)}}$ for all $\nu > 0$, the coherent information lower bound is not equal to $Q_2(\Phi_{g,\nu})$ at least in the region $\nu > 0$ and $g \in [\frac{1}{1-2^{-h(\nu)}}, 1 + \frac{1}{\nu})$. In the following theorem we obtain an improved lower bound on the two-way capacities of the thermal amplifier.

Theorem 26. *Let $g \geq 1$, $\nu \geq 0$, and $N_s \geq 0$. The EC two-way capacities $Q_2(\Phi_{g,\nu}, N_s)$ and $K(\Phi_{g,\nu}, N_s)$ of the thermal amplifier $\Phi_{g,\nu}$ satisfy the following lower bound*

$$K(\Phi_{g,\nu}, N_s) \geq Q_2(\Phi_{g,\nu}, N_s) \geq \sup_{\substack{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N} \\ (1-c^2)M\leq N_s}} \mathscr{R}\left(g+(g-1)\nu, \frac{g}{g+(g-1)\nu}, M, c, k\right), \tag{5.273}$$

and, in particular, the unconstrained two-way capacities satisfy

$$K(\Phi_{g,\nu}) \geq Q_2(\Phi_{g,\nu}) \geq \sup_{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N}} \mathscr{R}\left(g+(g-1)\nu, \frac{g}{g+(g-1)\nu}, M, c, k\right), \tag{5.274}$$

where the quantity $\mathscr{R}$ is defined in (5.188).

Proof. Theorem 26 is a direct consequence of Lemma 5 and Theorem 15. □

Theorem 26 shows a new lower bound, reported in (5.274), on the two-way capacities of the thermal amplifier $\Phi_{g,\nu}$. Our new lower bound outperforms all the previous known lower bounds in a large region of the parameters g and ν. In Fig. 5.8a and in Fig. 5.8b we plot our new bound and its ratio with the PLOB bound, respectively, with respect to ν where the transmissivity is chosen to be equal to $g(\nu) := \frac{1}{1-2^{-h(\nu)}}$, which is the lower endpoint for the g-range for which the best known lower bound on $Q_2(\Phi_{g,\nu})$ (i.e. the coherent information lower bound reported in (5.80)) vanishes. From Fig. 5.8a and Fig. 5.8b we see that for these choices of ν and $g(\nu)$, our new lower bound is now the best lower bound on $Q_2(\Phi_{g,\nu})$ and it achieves the $\simeq 14\%$ of the PLOB bound for $\nu \gg 1$. In Fig. 5.9 we plot our new bound with respect to g for $\nu = 1$ and $\nu = 10$.

Our new bound can outperform also the WOGP-bound [36], which is the best known lower bound (before our work) on the secret-key capacity $K(\Phi_{g,\nu})$. To demonstrate that our new bound can be strictly tighter than the WOGP lower bound, in Fig. 5.10 we plot the latter bound and our new bound with respect to g for $\nu = 1$ and $\nu = 10$. From Fig. 5.10 we note that the WOGP lower bound vanishes for smaller values of g than our bound. In particular, fixed $\nu > 0$, we numerically observe that our new bound is strictly positive for all $g < 1 + \frac{1}{\nu}$, which is the region where the two-way capacities of $\Phi_{g,\nu}$ are strictly positive, as established by Theorem 25. As an example, for $\nu = 1$, in Fig. 5.11 we plot the ratio between our bound and the PLOB bound in logarithmic scale and we see that our bound is strictly positive for $g \lesssim 1 + \frac{1}{\nu} = 2$.

Results on the two-way capacities of the additive Gaussian noise

Let us consider the additive Gaussian noise Λ_ξ of parameter $\xi \geq 0$. Since the PLOB bound in (5.91) vanishes for $\xi \geq 1$, it is already known that the two-way capacities of Λ_ξ vanish for $\xi \geq 1$. The following theorem establishes that also the vice-versa is true.

Theorem 27. *Let $\xi \geq 0$, and $N_s > 0$. The energy-constrained two-way capacities of the additive Gaussian noise $Q_2(\Lambda_\xi, N_s)$ and $K(\Lambda_\xi, N_s)$ vanish if and only if $\xi \geq 1$. In particular, the two-way capacities $Q_2(\Lambda_\xi)$ and $K(\Lambda_\xi)$ vanish if and only if $\xi \geq 1$.*

Proof. Theorem 27 is a direct consequence of Lemma 5 and Theorem 13. □

It is an open question whether the coherent information lower bound in Eq. 5.92 equals the true two-way quantum capacity of the additive Gaussian noise $Q_2(\Lambda_\xi)$: Theorem 27 provides a negative answer to this question. Indeed, although $Q_2(\Lambda_\xi) = 0$ if and only if $\xi \geq 1$ (thanks to Theorem 27), the coherent information lower bound vanishes for all $\xi \geq \frac{1}{e}$. Hence, the coherent information lower bound is not equal to $Q_2(\Lambda_\xi)$ at least in the region $\xi \in [\frac{1}{e}, 1)$. In the following theorem we obtain an improved lower bound on the two-way capacities of the additive Gaussian noise.

Theorem 28. *Let $\xi \in [0, 1)$ and $N_s \geq 0$. The EC two-way capacities $Q_2(\Lambda_\xi, N_s)$ and $K(\Lambda_\xi, N_s)$ of the additive Gaussian noise Λ_ξ satisfy the following lower bound*

$$K(\Lambda_\xi, N_s) \geq Q_2(\Lambda_\xi, N_s) \geq \sup_{\substack{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N} \\ (1-c^2)M\leq N_s}} \mathscr{R}\left(1+\xi, \frac{1}{1+\xi}, M, c, k\right), \tag{5.275}$$

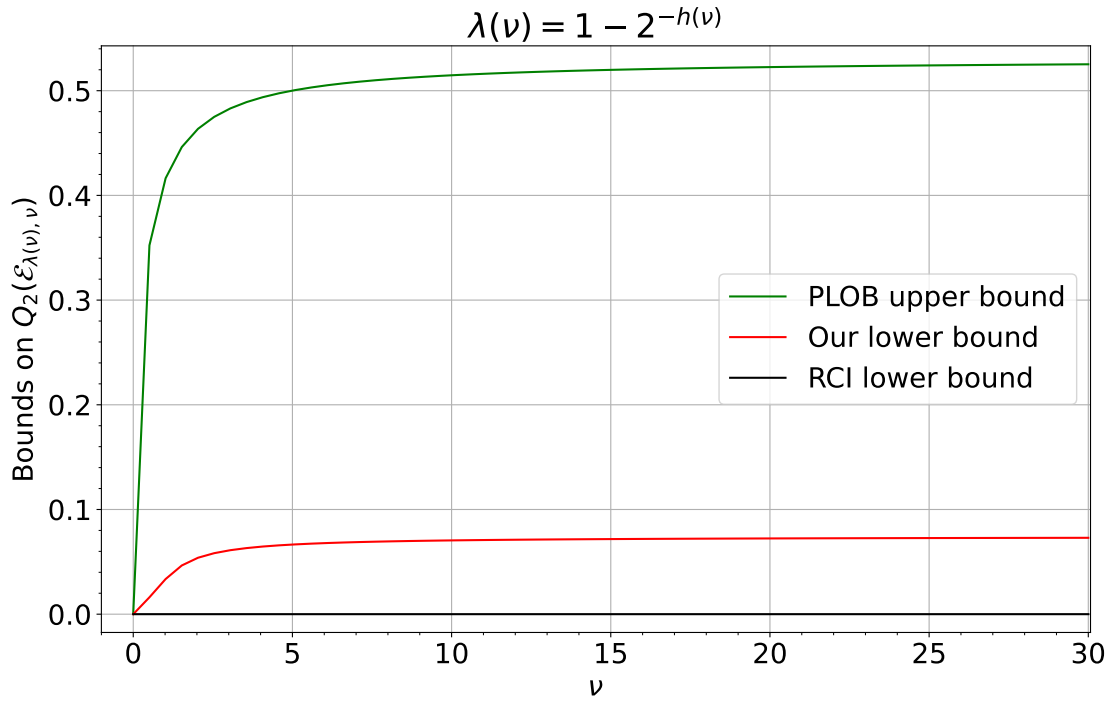
and, in particular, the unconstrained two-way capacities satisfy

$$K(\Lambda_\xi) \geq Q_2(\Lambda_\xi) \geq \sup_{c\in(0,1),\, M\in\mathbb{N}^+,\, k\in\mathbb{N}} \mathscr{R}\left(1+\xi, \frac{1}{1+\xi}, M, c, k\right), \tag{5.276}$$

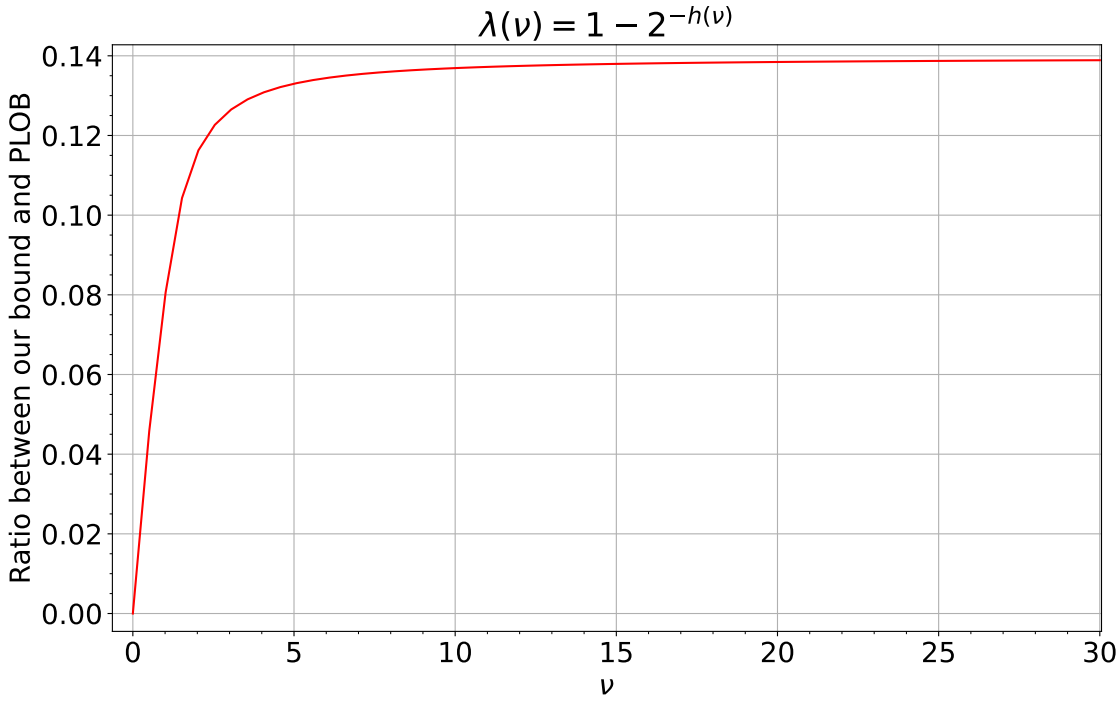
where the quantity $\mathscr{R}$ is defined in (5.188).

Proof. Theorem 28 is a direct consequence of Lemma 5 and Theorem 15. □

Theorem 28 shows a new lower bound, reported in (5.276), on the two-way capacities of the additive Gaussian noise Λ_ξ. Our new lower bound outperforms all the previous known lower bounds in a large region of the parameter ξ, as it can been seen from Fig. 5.12.



(a)



(b)

Figure 5.2: **(a).** Bounds on the two-way quantum capacity of the thermal attenuator $Q_2(\mathcal{E}_{\lambda(\nu),\nu})$ plotted with respect to ν, where the transmissivity is equal to the critical value $\lambda(\nu) := 1 - 2^{-h(\nu)}$. The red curve is our lower bound calculated by exploiting (5.271). The black curve is the best known lower bound on $Q_2(\mathcal{E}_{\lambda(\nu),\nu})$, which is the reverse coherent information lower bound reported in (5.64) (which is zero since $\lambda(\nu) = 1 - 2^{-h(\nu)}$). The green curve is the PLOB upper bound reported in (5.62). These bounds are also bounds on the secret-key capacity $K(\mathcal{E}_{\lambda(\nu),\nu})$. **(b).** Ratio between our new lower bound on the two-way quantum and secret-key capacities in (5.271) and the PLOB bound in (5.62) as a function of ν where the transmissivity is $\lambda(\nu) := 1 - 2^{-h(\nu)}$.

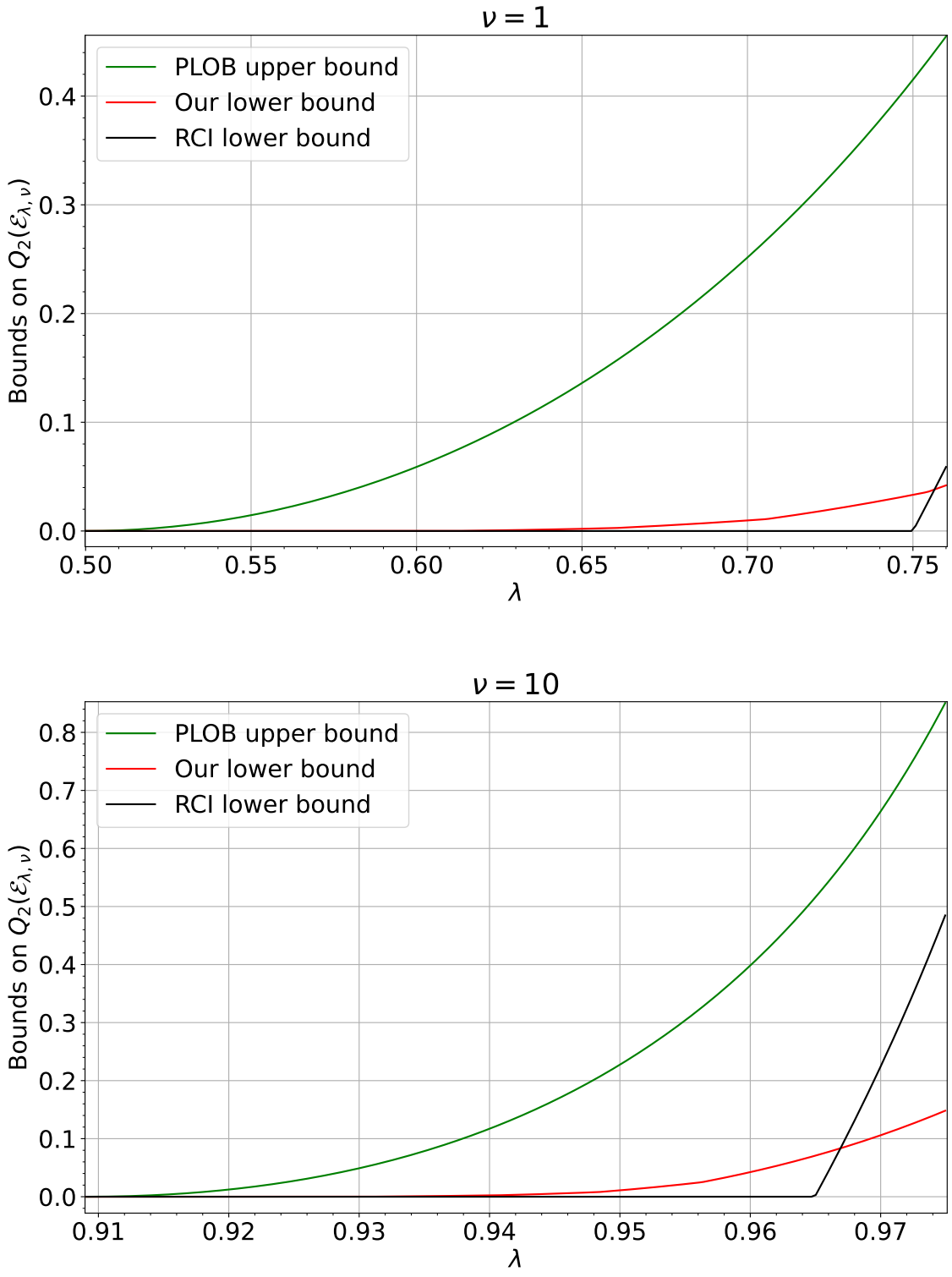


Figure 5.3: Bounds on the two-way quantum capacity of the thermal attenuator $Q_2(\mathscr{E}_{\lambda,\nu})$ plotted with respect to λ. The red line is our new lower bound obtained by exploiting (5.271). The black line is the best known lower bound on $Q_2(\mathscr{E}_{\lambda,\nu})$, which is the reverse coherent information lower bound reported in (5.64). The green line is the PLOB upper bound reported in (5.62). These bounds are also bounds on the secret-key capacity $K(\mathscr{E}_{\lambda,\nu})$.

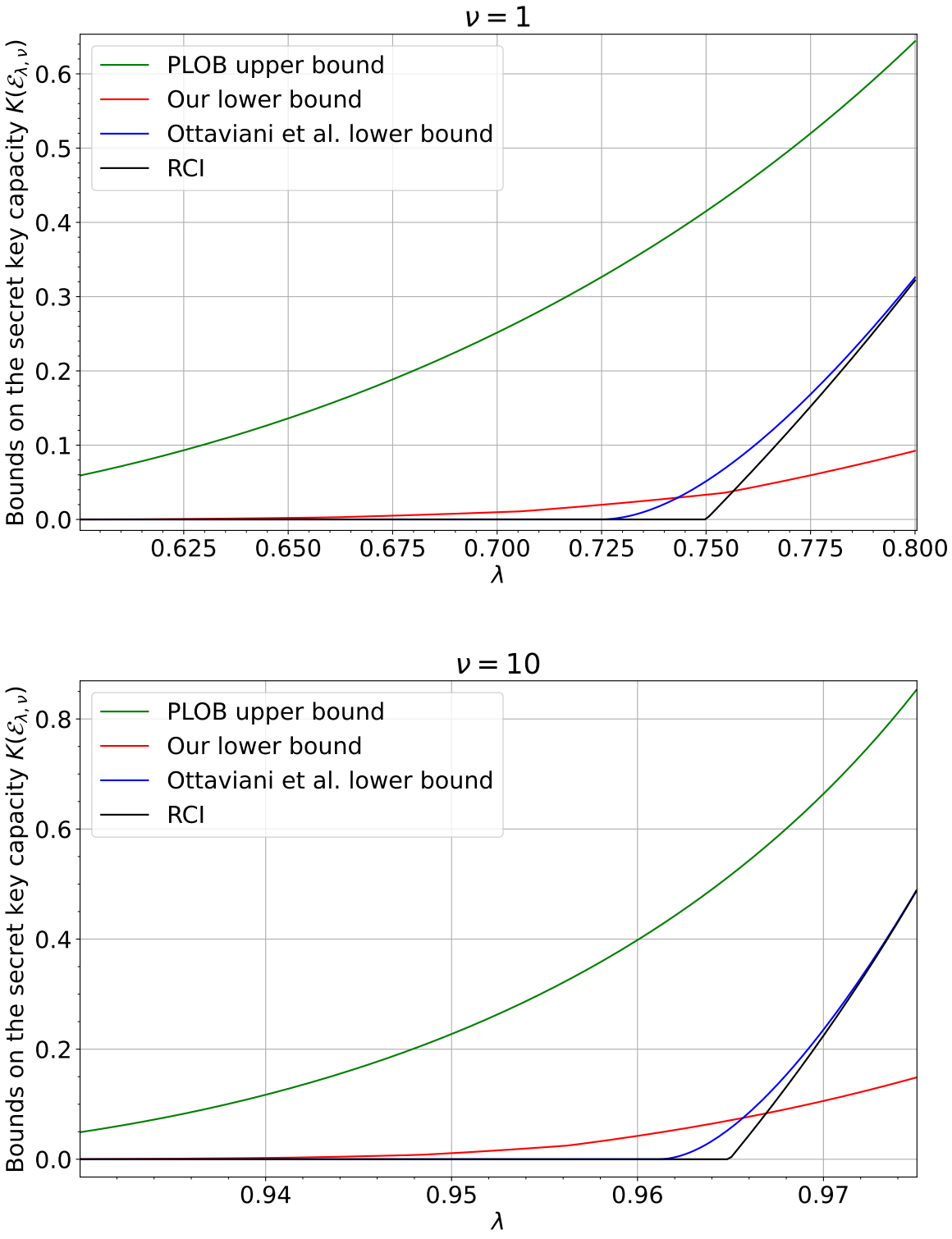


Figure 5.4: Bounds on the secret-key capacity of the thermal attenuator $K(\mathscr{E}_{\lambda,\nu})$ plotted with respect to λ. The red line is our new lower bound obtained by exploiting (5.271), the black line is the bound in (5.64) calculated by evaluating the reverse coherent information in (5.65), the blue line is the best known lower bound discovered by [27], and the green line is the PLOB upper bound reported in (5.62).

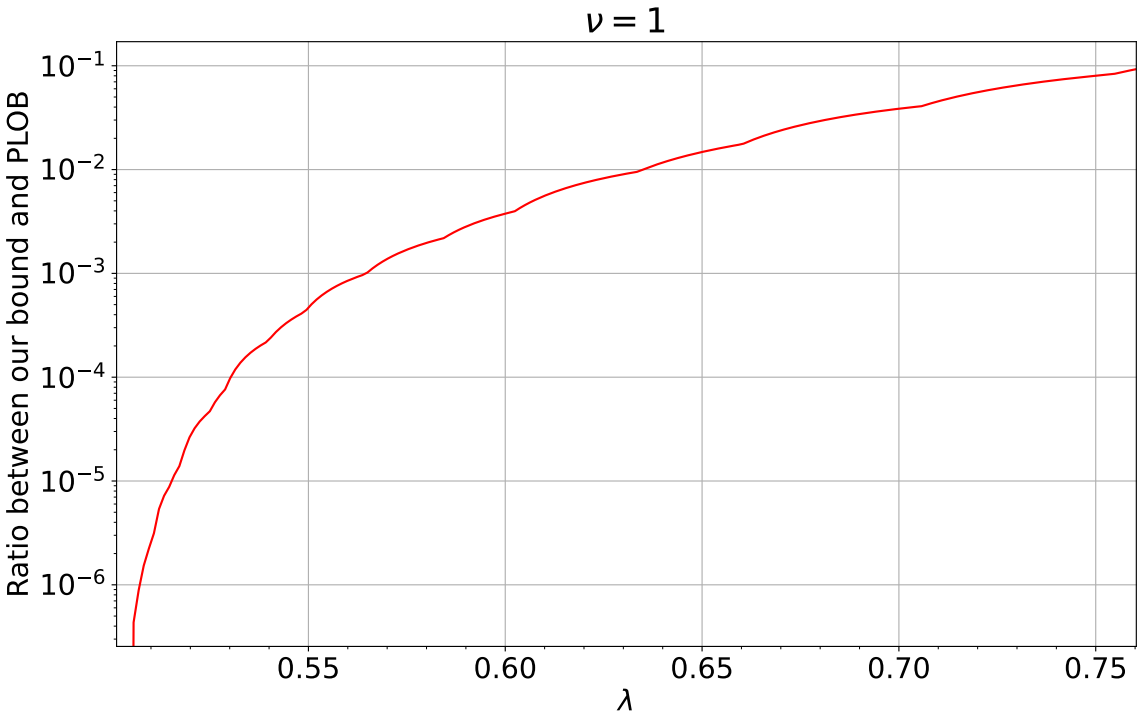


Figure 5.5: Ratio between our lower bound on the two-way quantum and secret-key capacities of the thermal attenuator in (5.271) and the PLOB bound in (5.62) as a function of λ for $\nu = 1$.

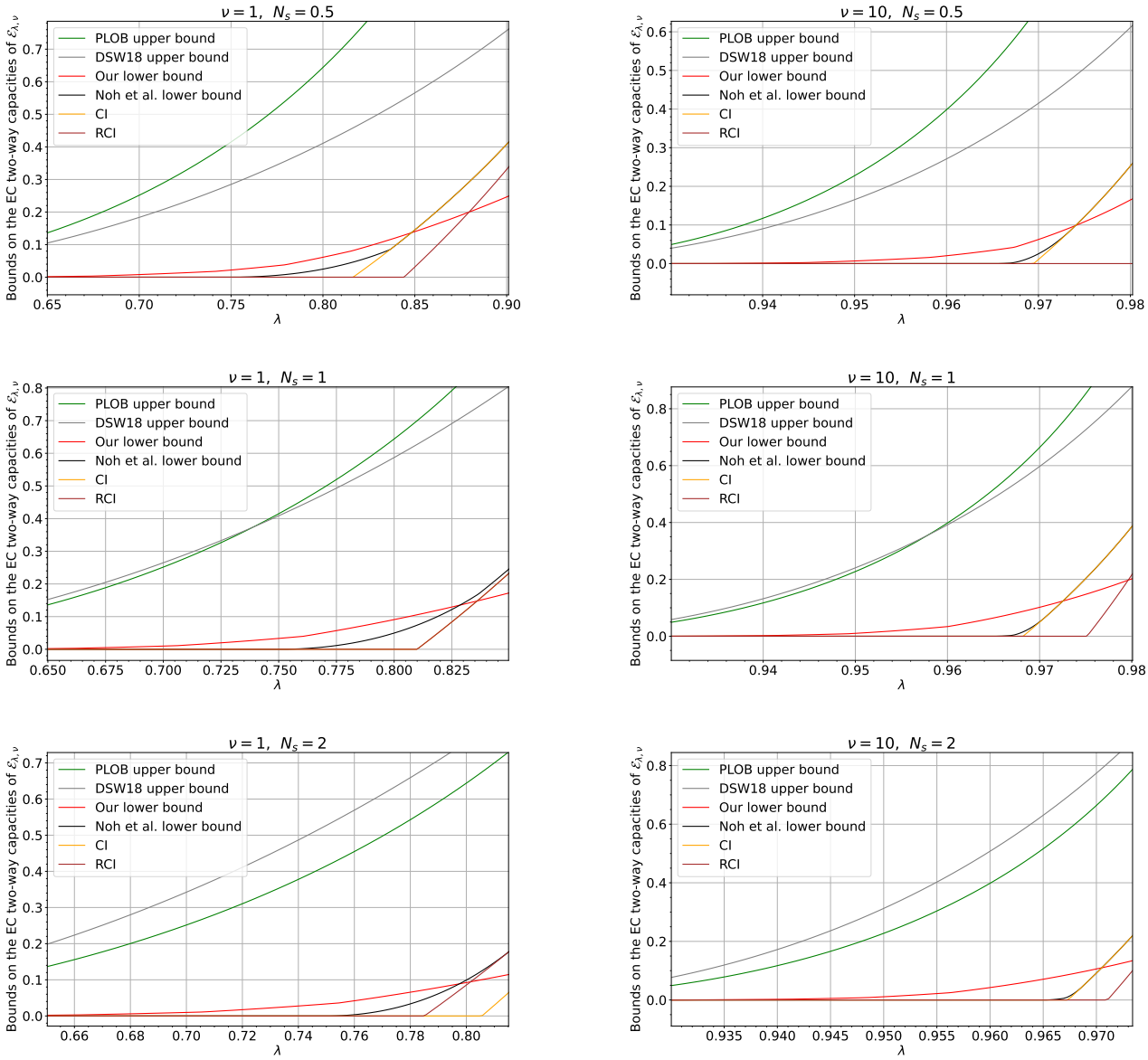


Figure 5.6: Bounds on the energy-constrained two-way quantum capacity $Q_2(\mathscr{E}_{\lambda,\nu}, N_s)$ and secret-key capacity $K(\mathscr{E}_{\lambda,\nu}, N_s)$ of the thermal attenuator plotted with respect to λ for different choices of ν and of the energy constraint N_s. The red line is our new lower bound obtained by exploiting (5.270), the black line is the NPJ lower bound [26] reported in (5.68), the yellow line is the coherent information lower bound reported in (5.67), the brown line is the reverse coherent information lower bound reported in (5.67), the grey line is the DSW18 upper bound [15], and the green line is the PLOB upper bound reported in (5.62).

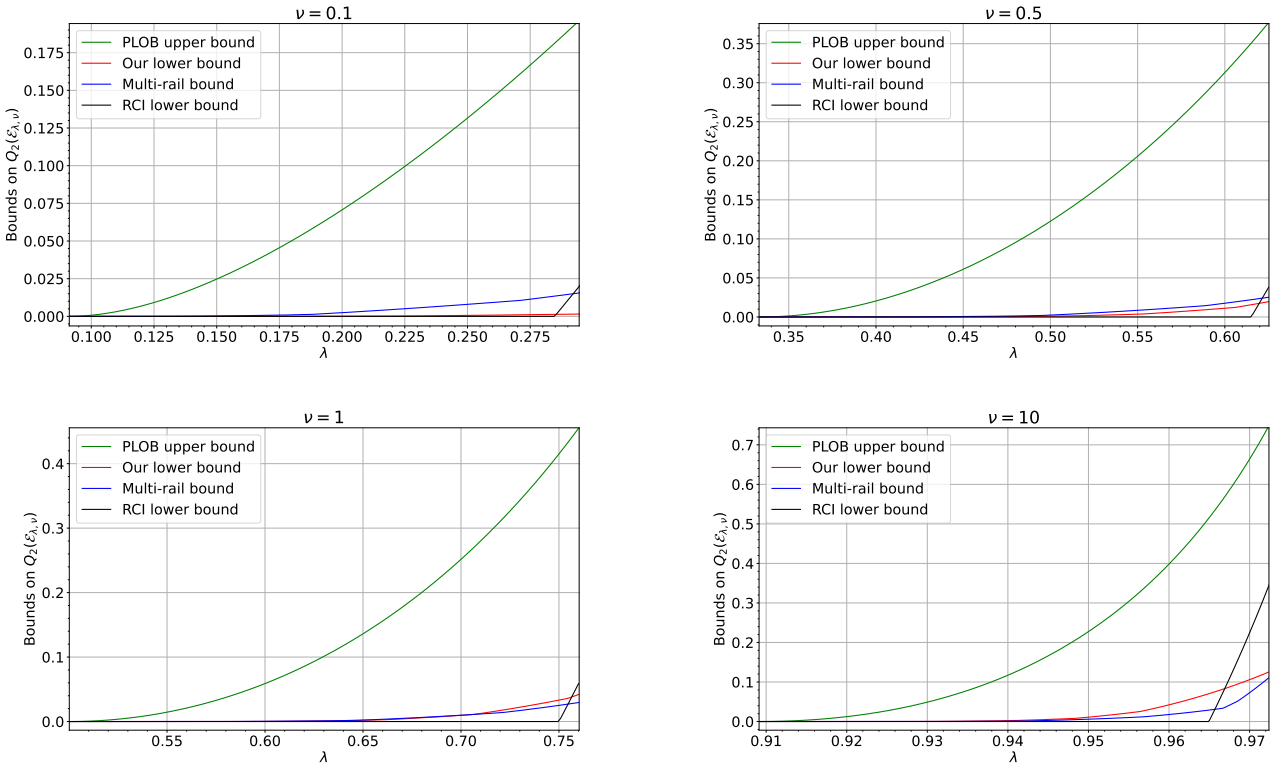
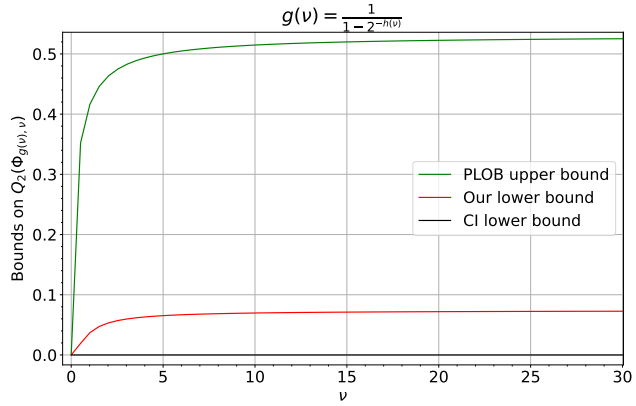
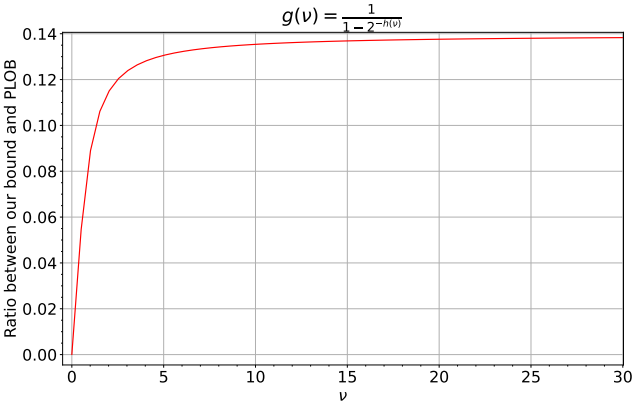


Figure 5.7: Bounds on the two-way quantum capacity of the thermal attenuator $Q_2(\mathscr{E}_{\lambda,\nu})$ plotted with respect to λ. The blue line is our multi-rail lower bound obtained by exploiting (5.272). The red line is our lower bound reported in (5.271). The black line is the best known lower bound on $Q_2(\mathscr{E}_{\lambda,\nu})$, which is the reverse coherent information lower bound reported in (5.64). The green line is the PLOB upper bound reported in (5.62). These bounds are also bounds on the secret-key capacity $K(\mathscr{E}_{\lambda,\nu})$.



(a)



(b)

Figure 5.8: **(a).** Bounds on the two-way quantum capacity of the thermal amplifier $\mathcal{Q}_2(\Phi_{g(\nu),\nu})$ plotted with respect to ν, where the gain is equal to the critical value $g(\nu) = \frac{1}{1-2^{-h(\nu)}}$. The red curve is our new lower bound calculated by exploiting (5.274). The black curve is the best known lower bound, i.e. the coherent information lower bound reported in (5.80) (which is zero since $g(\nu) = \frac{1}{1-2^{-h(\nu)}}$). The green curve is the PLOB upper bound reported in (5.79). These bounds are also bounds on the secret-key capacity $K(\Phi_{g(\nu),\nu})$. **(b).** Ratio between our new lower bound in (5.274) and the PLOB bound in (5.79) as a function of ν where the gain is $g(\nu) := \frac{1}{1-2^{-h(\nu)}}$.

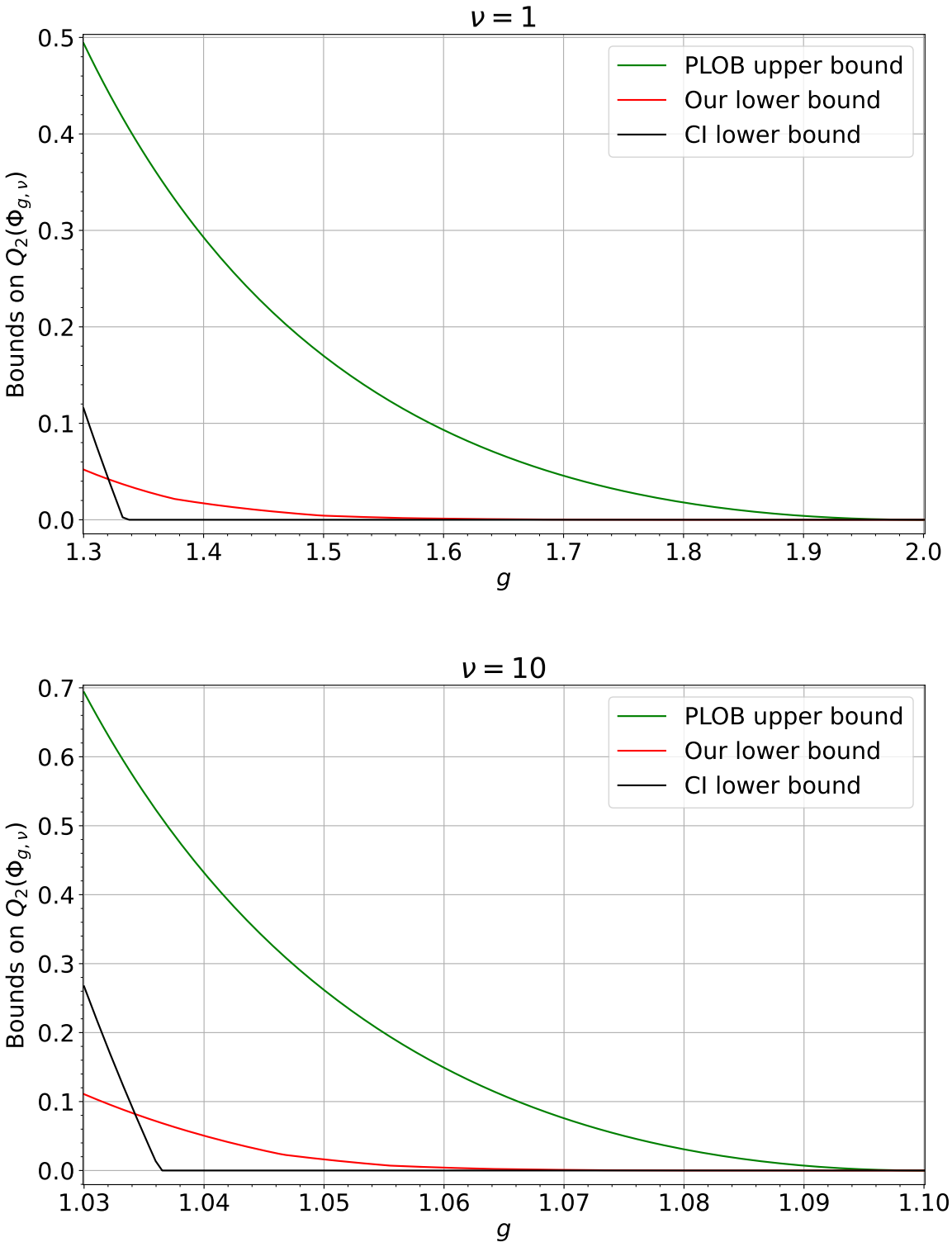


Figure 5.9: Bounds on the two-way quantum capacity of thermal amplifier $Q_2(\Phi_{g,\nu})$ plotted with respect to g. The red line is our new lower bound obtained by exploiting (5.274). The black line is the best known lower bound on $Q_2(\Phi_{g,\nu})$, which is the coherent information lower bound reported in (5.80). The green line is the PLOB upper bound reported in (5.79). These bounds are also bounds on the secret-key capacity $K(\Phi_{g,\nu})$.

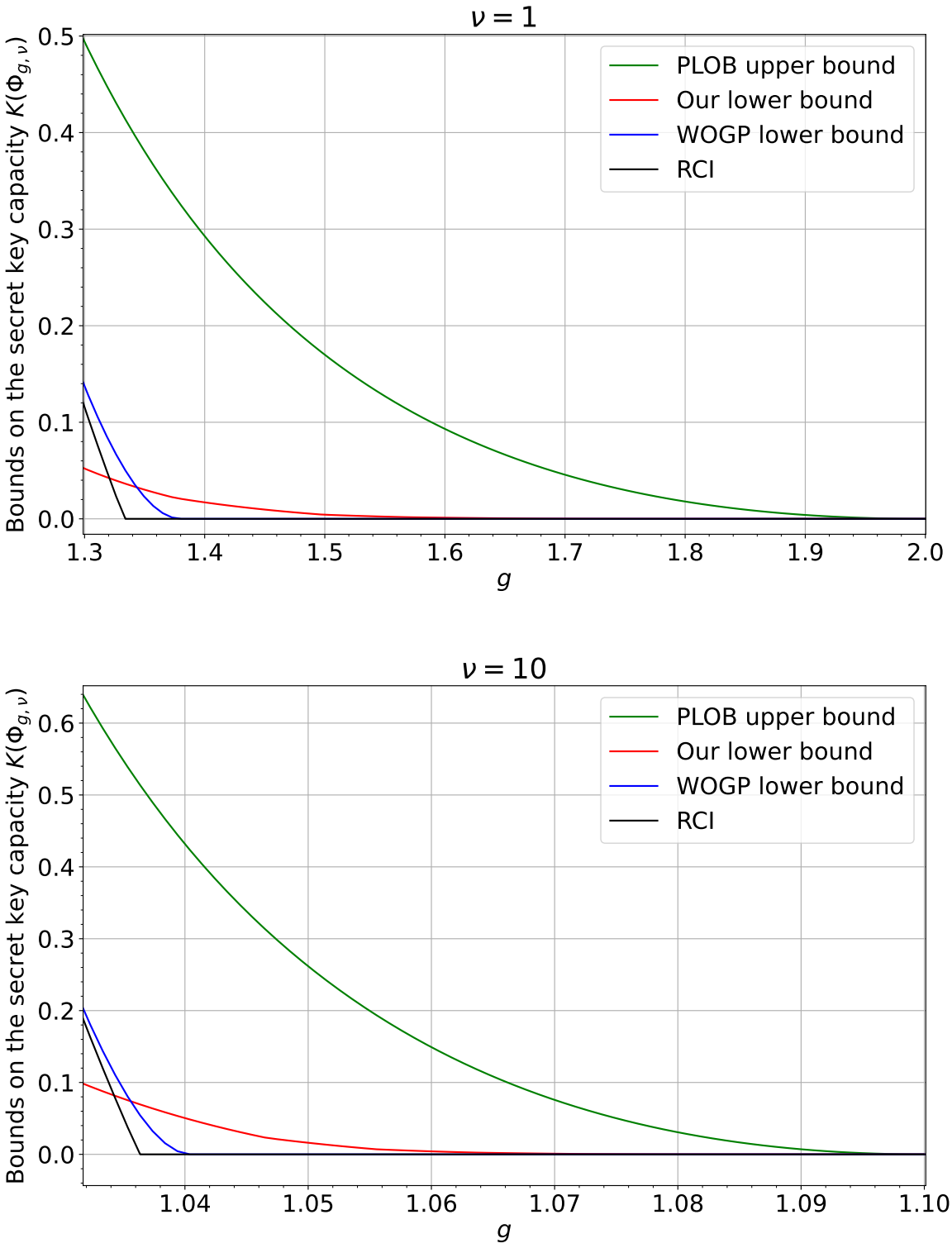


Figure 5.10: Bounds on the secret-key capacity of the thermal amplifier $K(\Phi_{g,\nu})$ plotted with respect to g. The red line is our new lower bound obtained by exploiting (5.274), the black line is the bound in (5.80) calculated by evaluating the coherent information in (5.81), the blue line is the WOGP lower bound [27], and the green line is the PLOB upper bound reported in (5.79).

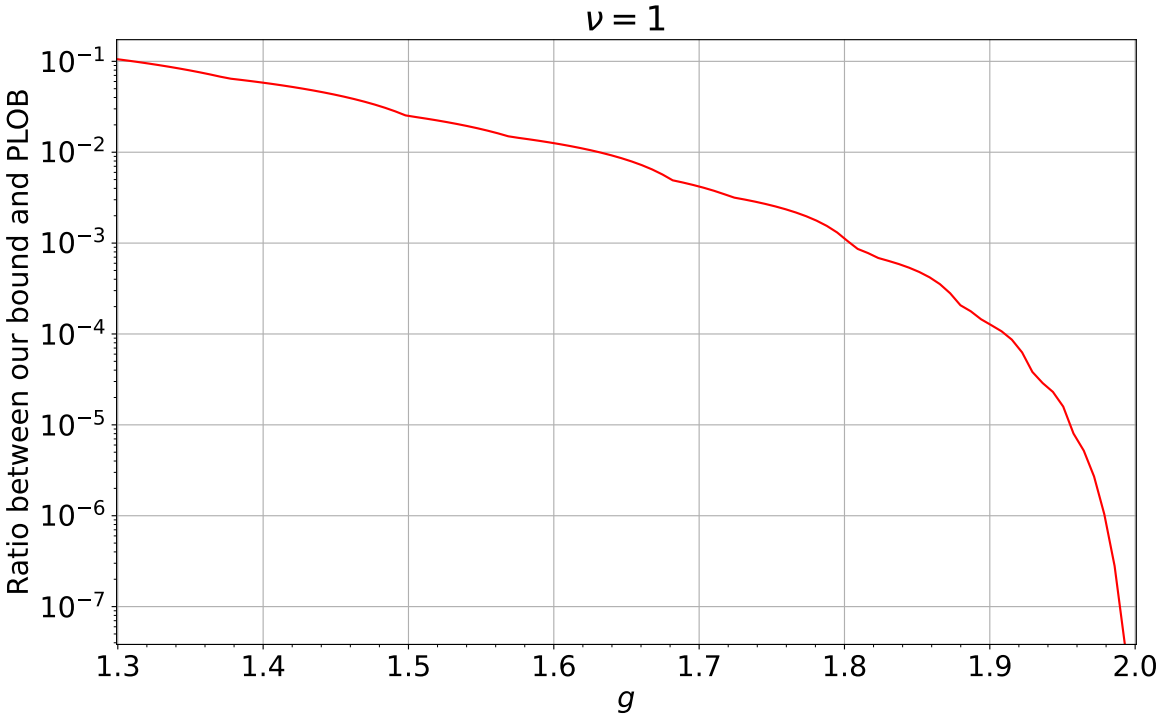


Figure 5.11: Ratio between our new lower bound on the two-way capacities of the thermal amplifier $\Phi_{g,\nu}$ in (5.274) and the PLOB bound in (5.79) as a function of g for $\nu = 1$.

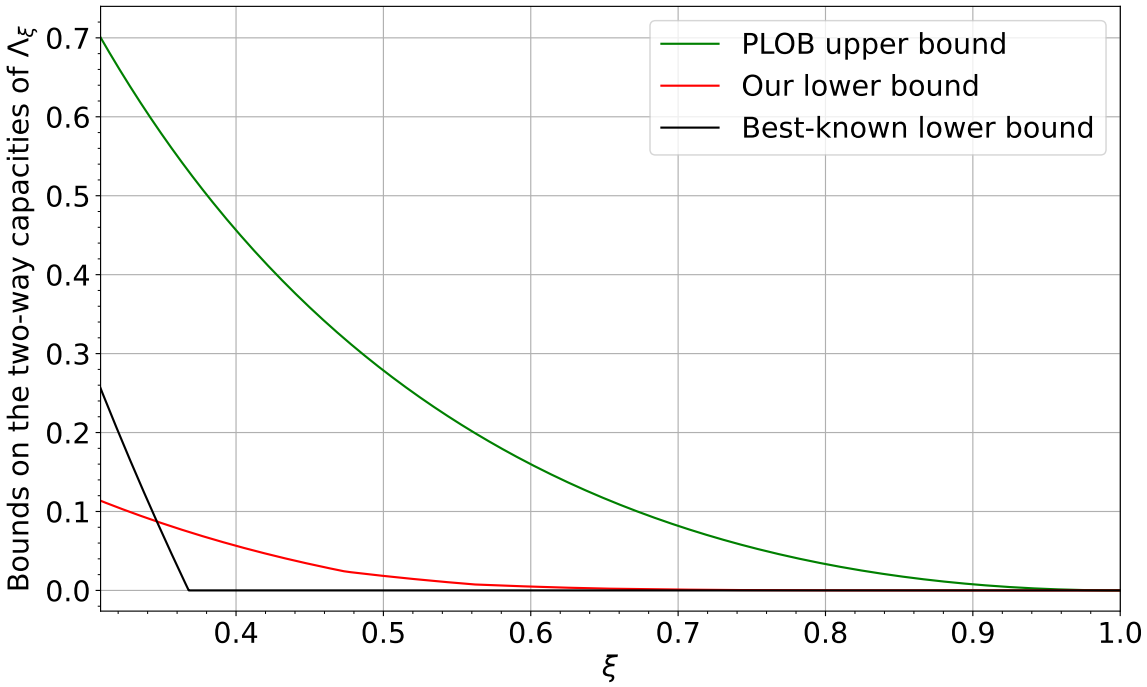


Figure 5.12: Bounds on the two-way quantum capacity $Q_2(\Lambda_\xi)$ and secret-key capacity $K(\Lambda_\xi)$ of the additive Gaussian noise plotted with respect to ξ. The red line is our new lower bound obtained by exploiting (5.276). The black line is the best known lower bound on $Q_2(\Lambda_\xi)$, which is the coherent information lower bound reported in (5.92). The green line is the PLOB bound reported in (5.91).

Bibliography

[1] C. H. Bennett. Quantum cryptography: public key distribution and coin tossing. In *Proc. IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, pages 175–179, 1984.

[2] A. Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, Taylor & Francis Group, Boca Raton, USA, 2017.

[3] F. Grosshans and P. Grangier. Continuous variable quantum cryptography using coherent states. *Phys. Rev. Lett.*, 88(5), 2002. ISSN 1079-7114. doi: 10.1103/physrevlett.88.057902. URL http://dx.doi.org/10.1103/PhysRevLett.88.057902.

[4] S. Pirandola et al. Advances in quantum cryptography. *Advances in Optics and Photonics*, 12(4):1012–1236, 2020. doi: 10.1364/AOP.361502. URL http://opg.optica.org/aop/abstract.cfm?URI=aop-12-4-1012.

[5] F. Laudenbach, C. Pacher, C.-H. F. Fung, A. Poppe, M. Peev, B. Schrenk, M. Hentschel, P. Walther, and H. Hübel. Continuous-variable quantum key distribution with gaussian modulation-the theory of practical implementations. *Advanced Quantum Technologies*, 1(1), 2018. ISSN 2511-9044. doi: 10.1002/qute.201800011. URL http://dx.doi.org/10.1002/qute.201800011.

[6] Y. Zhang, Z. Chen, S. Pirandola, X. Wang, C. Zhou, B. Chu, Y. Zhao, B. Xu, S. Yu, and H. Guo. Long-distance continuous-variable quantum key distribution over 202.81 km of fiber. *Phys. Rev. Lett.*, 125:010502, 2020. doi: 10.1103/PhysRevLett.125.010502. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.010502.

[7] A. A. E. Hajomer, I. Derkach, N. Jain, H.-M. Chin, U. L. Andersen, and T. Gehring. Long-distance continuous-variable quantum key distribution over 100-km fiber with local local oscillator. *Science Advances*, 10(1):eadi9474, 2024. doi: 10.1126/sciadv.adi9474. URL https://www.science.org/doi/abs/10.1126/sciadv.adi9474.

[8] Y. Zhang, Z. Li, Z. Chen, C. Weedbrook, Y. Zhao, X. Wang, Y. Huang, C. Xu, X. Zhang, Z. Wang, M. Li, X. Zhang, Z. Zheng, B. Chu, X. Gao, N. Meng, W. Cai, Z. Wang, G. Wang, S. Yu, and H. Guo. Continuous-variable QKD over 50 km commercial fiber. *Quantum Science and Technology*, 4(3):035006, 2019. ISSN 2058-9565. doi: 10.1088/2058-9565/ab19d1. URL http://dx.doi.org/10.1088/2058-9565/ab19d1.

[9] Y. Pi, H. Wang, Y. Pan, Y. Shao, Y. Li, J. Yang, Y. Zhang, W. Huang, and B. Xu. Sub-mbps key-rate continuous-variable quantum key distribution with local local oscillator over 100-km fiber. *Optics Letters*, 48(7):1766, 2023. ISSN 1539-4794. doi: 10.1364/ol.485913. URL http://dx.doi.org/10.1364/OL.485913.

[10] D. Huang, P. Huang, D. Lin, and G. Zeng. Long-distance continuous-variable quantum key distribution by controlling excess noise. *Scientific Reports*, 6:19201, 01 2016. doi: 10.1038/srep19201.

[11] S. Wehner, D. Elkouss, and R. Hanson. Quantum internet: A vision for the road ahead. *Science*, 362:eaam9288, 10 2018. doi: 10.1126/science.aam9288.

[12] S. Pirandola, S. L. Braunstein, R. Laurenza, C. Ottaviani, T. P. W. Cope, G. Spedalieri, and L. Banchi. Theory of channel simulation and bounds for private communication. *Quantum Science and Technology*, 3(3):035009, 2018.

[13] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2nd edition, 2017. ISBN 9781316813300.

[14] S. Khatri and M. M. Wilde. Principles of quantum communication theory: A modern approach, 2020. URL https://arxiv.org/abs/2011.04672.

[15] N. Davis, M. E. Shirokov, and M. M. Wilde. Energy-constrained two-way assisted private and quantum capacities of quantum channels. *Phys. Rev. A*, 97:062310, 2018. doi: 10.1103/PhysRevA.97.062310.

[16] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, 1991.

[17] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi. Fundamental limits of repeaterless quantum communications. *Nat. Commun.*, 8(1):15043, 2017. doi: 10.1038/ncomms15043.

[18] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.*, 81:5932–5935, 1998. doi: 10.1103/PhysRevLett.81.5932.

[19] W. J. Munro, K. Azuma, K. Tamaki, and K. Nemoto. Inside quantum repeaters. *IEEE Journal of Selected Topics in Quantum Electronics*, 21(3):78–90, 2015. doi: 10.1109/JSTQE.2015.2392076.

[20] K. Goodenough, D. Elkouss, and S. Wehner. Assessing the performance of quantum repeaters for all phase-insensitive Gaussian bosonic channels. *New Journal of Physics*, 18(6):063005, 2016.

[21] M. Takeoka, S. Guha, and M. M. Wilde. Fundamental rate-loss tradeoff for optical quantum key distribution. *Nat. Commun.*, 5(1):5235, 2014.

[22] A. S. Holevo and R. F. Werner. Evaluating capacities of bosonic Gaussian channels. *Phys. Rev. A*, 63:032312, 2001. doi: 10.1103/PhysRevA.63.032312.

[23] M. M. Wilde, M. Tomamichel, and M. Berta. Converse bounds for private communication over quantum channels. *IEEE Transactions on Information Theory*, 63(3):1792–1817, 2017. ISSN 1557-9654. doi: 10.1109/TIT.2017.2648825.

[24] M. Takeoka, S. Guha, and M. M. Wilde. The squashed entanglement of a quantum channel. *IEEE Transactions on Information Theory*, 60(8):4987–4998, 2014. doi: 10.1109/TIT.2014.2330313.

[25] S. Pirandola, R. García-Patrón, S. L. Braunstein, and S. Lloyd. Direct and reverse secret-key capacities of a quantum channel. *Phys. Rev. Lett.*, 102:050503, 2009. doi: 10.1103/PhysRevLett.102.050503.

[26] K. Noh, S. Pirandola, and L. Jiang. Enhanced energy-constrained quantum communication over bosonic Gaussian channels. *Nat. Commun.*, 11(1):457, 2020. doi: 10.1038/s41467-020-14329-6.

[27] C. Ottaviani, R. Laurenza, T. P. W. Cope, G. Spedalieri, S. L. Braunstein, and S. Pirandola. Secret key capacity of the thermal-loss channel: improving the lower bound. In M. T. Gruneisen, M. Dusek, and J. G. Rarity, editors, *Quantum Information Science and Technology II*, volume 9996, page 999609. International Society for Optics and Photonics, SPIE, 2016. doi: 10.1117/12.2244899.

[28] Y. Tamura, H. Sakuma, K. Morita, M. Suzuki, Y. Yamamoto, K. Shimada, Y. Honma, K. Sohma, T. Fujii, and T. Hasegawa. The first 0.14-db/km loss optical fiber and its impact on submarine transmission. *J. Lightwave Technol.*, 36(1):44–49, 2018. URL http://www.osapublishing.org/jlt/abstract.cfm?URI=jlt-36-1-44.

[29] M.-J. Li and T. Hayashi. Chapter 1 – Advances in low-loss, large-area, and multicore fibers. In A. E. Willner, editor, *Optical Fiber Telecommunications VII*, pages 3–50. Academic Press, 2020. ISBN 978-0-12-816502-7. doi: https://doi.org/10.1016/B978-0-12-816502-7.00001-4.

[30] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, 1996.

[31] R. Simon. Peres–Horodecki separability criterion for continuous variable systems. *Phys. Rev. Lett.*, 84:2726–2729, 2000.

[32] G. Giedke, B. Kraus, M. Lewenstein, and J. I. Cirac. Entanglement criteria for all bipartite Gaussian states. *Phys. Rev. Lett.*, 87:167904, 2001.

[33] A. S. Holevo and V. Giovannetti. Quantum channels and their entropic characteristics. *Reports on Progress in Physics*, 75(4):046001, 2012. doi: 10.1088/0034-4885/75/4/046001. URL https://doi.org/10.1088/0034-4885/75/4/046001.

[34] A. S. Holevo. Entanglement-breaking channels in infinite dimensions. *Probl. Pered. Inform.*, 44(3):3–18, 2008. doi: 10.1134/S0032946008030010. (English translation: Probl. Inf. Transm. 44(3):171–184, 2008).

[35] S. Pirandola. Limits and security of free-space quantum communications. *Phys. Rev. Research*, 3:013279, 2021. doi: 10.1103/PhysRevResearch.3.013279.

[36] G. Wang, C. Ottaviani, H. Guo, and S. Pirandola. Improving the lower bound to the secret-key capacity of the thermal amplifier channel. *The European Physical Journal D*, 73(1):17, 2019. doi: 10.1140/epjd/e2018-90351-0. URL https://doi.org/10.1140/epjd/e2018-90351-0.

[37] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. Royal Soc. A*, 461(2053):207–235, 2005. doi: 10.1098/rspa.2004.1372.

[38] H. Dür and Briegel H. J. Entanglement purification and quantum error correction. *Reports on Progress in Physics*, 70(8):1381–1424, 2007. doi: 10.1088/0034-4885/70/8/r03. URL https://doi.org/10.1088/0034-4885/70/8/r03.

[39] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, 1996. doi: 10.1103/PhysRevA.54.3824.

[40] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–725, 1996.

[41] J. Miguel-Ramiro and W. Dür. Efficient entanglement purification protocols for d-level systems. *Phys. Rev. A*, 98:042309, 2018. doi: 10.1103/PhysRevA.98.042309. URL https://link.aps.org/doi/10.1103/PhysRevA.98.042309.

[42] K. G. H. Vollbrecht and F. Verstraete. Interpolation of recurrence and hashing entanglement distillation protocols. *Phys. Rev. A*, 71:062325, 2005. doi: 10.1103/PhysRevA.71.062325. URL https://link.aps.org/doi/10.1103/PhysRevA.71.062325.

[43] B. C. Sanders. Quantum dynamics of the nonlinear rotator and the effects of continual spin measurement. *Phys. Rev. A*, 40:2417–2427, 1989. doi: 10.1103/PhysRevA.40.2417.

[44] M. S. Winnel, J. J. Guanzon, N. Hosseinidehaj, and T. C. Ralph. Achieving the ultimate end-to-end rates of lossy quantum communication networks. *arXiv:2203.13924*, 2022. doi: 10.48550/ARXIV.2203.13924. URL https://arxiv.org/abs/2203.13924.

[45] N. N. Kalb, A. A. A. A. Reiserer, P. C. Humphreys, J. J. W. Bakermans, S. J. Kamerling, N. H. Nickerson, S. C. Benjamin, D. J. Twitchen, M. Markham, and R. Hanson. Entanglement distillation between solid-state quantum network nodes. *Science*, 356(6341):928–932, 2017. doi: 10.1126/science.aan0070. URL https://www.science.org/doi/abs/10.1126/science.aan0070.

[46] X.-M. Hu, C.-X. Huang, Y.-B. Sheng, L. Zhou, B.-H. Liu, Y. Guo, C. Zhang, W.-B. Xing, Y.-F. Huang, C.-F. Li, and G.-C. Guo. Long-distance entanglement purification for quantum communication. *Phys. Rev. Lett.*, 126(1), 2021. ISSN 1079-7114. doi: 10.1103/physrevlett.126.010503. URL http://dx.doi.org/10.1103/PhysRevLett.126.010503.

[47] S. Ecker, P. Sohr, L. Bulla, M. Huber, M. Bohmann, and R. Ursin. Experimental single-copy entanglement distillation. *Phys. Rev. Lett.*, 127(4), 2021. ISSN 1079-7114. doi: 10.1103/physrevlett.127.040506. URL http://dx.doi.org/10.1103/PhysRevLett.127.040506.

[48] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera. Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Phys. Rev. Lett.*, 77:2818–2821, 1996. doi: 10.1103/PhysRevLett.77.2818. URL https://link.aps.org/doi/10.1103/PhysRevLett.77.2818.

[49] J. Dehaene, M. Van den Nest, B. De Moor, and F. Verstraete. Local permutations of products of bell states and entanglement distillation. *Phys. Rev. A*, 67:022310, 2003. doi: 10.1103/PhysRevA.67.022310. URL https://link.aps.org/doi/10.1103/PhysRevA.67.022310.

[50] M. Horodecki and P. Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Phys. Rev. A*, 59:4206–4216, 1999.

[51] G. Alber, A. Delgado, N. Gisin, and I. Jex. Efficient bipartite quantum state purification in arbitrary dimensional hilbert spaces. *Journal of Physics A: Mathematical and General*, 34 (42):8821–8833, 2001. doi: 10.1088/0305-4470/34/42/307. URL https://doi.org/10.1088/0305-4470/34/42/307.

[52] H. Bombin and M. A. Martin-Delgado. Entanglement distillation protocols and number theory. *Phys. Rev. A*, 72:032313, 2005. doi: 10.1103/PhysRevA.72.032313. URL https://link.aps.org/doi/10.1103/PhysRevA.72.032313.

[53] E. Hostens, J. Dehaene, and B. De Moor. Asymptotic adaptive bipartite entanglement-distillation protocol. *Phys. Rev. A*, 73:062337, 2006. doi: 10.1103/PhysRevA.73.062337. URL https://link.aps.org/doi/10.1103/PhysRevA.73.062337.

[54] S. Barnett and P. M. Radmore. *Methods in Theoretical Quantum Optics*. Oxford Series in Optical and Imaging Sciences. Clarendon Press, 2002. ISBN 9780198563617.

[55] C. D. Cushen and R. L. Hudson. A quantum-mechanical central limit theorem. *Journal of Applied Probability*, 8(3):454–469, 1971. doi: 10.2307/3212170.

[56] F. A. Mele, L. Lami, and V. Giovannetti. Quantum optical communication in the presence of strong attenuation noise. *Phys. Rev. A*, 106:042437, 2022. doi: 10.1103/PhysRevA.106.042437. URL https://link.aps.org/doi/10.1103/PhysRevA.106.042437.

[57] A. S. Holevo. The Choi–Jamiolkowski forms of quantum Gaussian channels. *J. Math. Phys.*, 52(4):042202, 2011. doi: 10.1063/1.3581879.

[58] A. S. Holevo. On the Choi–Jamiolkowski correspondence in infinite dimensions. *Preprint arXiv:1004.0196*, 2010.

[59] M. Horodecki, P. Horodecki, and R. Horodecki. Inseparable two spin-$\frac{1}{2}$ density matrices can be distilled to a singlet form. *Phys. Rev. Lett.*, 78:574–577, 1997. doi: 10.1103/PhysRevLett.78.574.

[60] E. Bersin, M. Sutula, Y. Q. Huan, A. Suleymanzade, D. R. Assumpcao, Y.-C. Wei, P.-J. Stas, C. M. Knaut, E. N. Knall, C. Langrock, N. Sinclair, R. Murphy, R. Riedinger, M. Yeh, C. J. Xin, S. Bandyopadhyay, D. D. Sukachev, B. Machielse, D. S. Levonian, M. K. Bhaskar, S. Hamilton, H. Park, M. Lončar, M. M. Fejer, P. B. Dixon, D. R. Englund, and M. D. Lukin. Telecom networking with a diamond quantum memory. *PRX Quantum*, 5:010303, 2024. doi: 10.1103/PRXQuantum.5.010303. URL https://link.aps.org/doi/10.1103/PRXQuantum.5.010303.

[61] A. Wallucks, I. Marinkovic, B. Hensen, R. Stockill, and S. Groblacher. A quantum memory at telecom wavelengths. *Nat. Phys.*, 16(7):772–777, 2020. ISSN 1745-2481. doi: 10.1038/s41567-020-0891-z. URL http://dx.doi.org/10.1038/s41567-020-0891-z.

[62] G. J. Pryde and A. G. White. Creation of maximally entangled photon-number states using optical fiber multiports. *Phys. Rev. A*, 68(5), 2003. ISSN 1094-1622. doi: 10.1103/physreva.68.052315. URL http://dx.doi.org/10.1103/PhysRevA.68.052315.

[63] I. Afek, O. Ambar, and Y. Silberberg. High-NOON states by mixing quantum and classical light. *Science*, 328(5980):879–881, 2010. doi: 10.1126/science.1188172. URL https://www.science.org/doi/abs/10.1126/science.1188172.

[64] L. Liang, G. W. Lin, Y. M. Hao, Y. P. Niu, and S. Q. Gong. Quantum nondemolition measurement of small photon numbers using stored light. *Phys. Rev. A*, 90:055801, Nov 2014. doi: 10.1103/PhysRevA.90.055801. URL https://link.aps.org/doi/10.1103/PhysRevA.90.055801.

[65] A. Cabello and F. Sciarrino. Loophole-free Bell test based on local precertification of photon's presence. *Phys. Rev. X*, 2:021010, 2012. doi: 10.1103/PhysRevX.2.021010. URL https://link.aps.org/doi/10.1103/PhysRevX.2.021010.

[66] A. V. Gorshkov, R. Nath, and T. Pohl. Dissipative many-body quantum optics in Rydberg media. *Phys. Rev. Lett.*, 110:153601, 2013. doi: 10.1103/PhysRevLett.110.153601. URL https://link.aps.org/doi/10.1103/PhysRevLett.110.153601.

[67] G.-S. Ye, B. Xu, Y. Chang, S. Shi, T. Shi, and L. Li. A photonic entanglement filter with Rydberg atoms. *Nat. Photonics*, 17(6):538–543, 2023. doi: 10.1038/s41566-023-01194-0.

[68] J. Honer, R. Löw, H. Weimer, T. Pfau, and H. P. Büchler. Artificial atoms can do more than atoms: Deterministic single photon subtraction from arbitrary light fields. *Phys. Rev. Lett.*, 107:093601, 2011. doi: 10.1103/PhysRevLett.107.093601. URL https://link.aps.org/doi/10.1103/PhysRevLett.107.093601.

[69] N. Stiesdal, H. Busche, K. Kleinbeck, J. Kumlin, M. G. Hansen, H. P. Büchler, and S. Hofferberth. Controlled multi-photon subtraction with cascaded Rydberg superatoms as single-photon absorbers. *Nat. Commun.*, 12(1), 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-24522-w. URL http://dx.doi.org/10.1038/s41467-021-24522-w.

[70] M. Mücke, E. Figueroa, J. Bochmann, C. Hahn, K. Murr, S. Ritter, C. Villas-Boas, and G. Rempe. Electromagnetically induced transparency with single atoms in a cavity. *Nature*, 465:755–8, 06 2010. doi: 10.1038/nature09093.

[71] S. Rosenblum, O. Bechler, I. Shomroni, Y. Lovsky, G. Guendelman, and B. Dayan. Extraction of a single photon from an optical pulse. *Nat. Photonics*, 10(1):19–22, 2015. ISSN 1749-4893. doi: 10.1038/nphoton.2015.227. URL http://dx.doi.org/10.1038/nphoton.2015.227.

[72] A. Kumar, A. Suleymanzade, M. Stone, L. Taneja, A. Anferov, D. I. Schuster, and J. Simon. Quantum-enabled millimetre wave to optical transduction using neutral atoms. *Nature*, 615 (7953):614–619, 2023. doi: 10.1038/s41586-023-05740-2.

[73] J. Rochman, T. Xie, J. G. Bartholomew, K. C. Schwab, and A. Faraon. Microwave-to-optical transduction with erbium ions coupled to planar photonic and superconducting resonators. *Nature Commun.*, 14(1):1153, 2023. doi: 10.1038/s41467-023-36799-0.

[74] Y. Xu, A. Al Sayem, L. Fan, C.-L. Zou, S. Wang, R. Cheng, W. Fu, L. Yang, M. Xu, and H. X. Tang. Bidirectional interconversion of microwave and light with thin-film lithium niobate. *Nat. Commun.*, 12, 2021. URL https://api.semanticscholar.org/CorpusID:236197889.

Chapter 6

Quantum communication on the bosonic loss-dephasing channel

Quantum optical systems are typically affected by two types of noise: photon loss and dephasing. Despite extensive research on each noise process individually, a comprehensive understanding of their combined effect is still lacking. A crucial problem lies in determining the values of loss and dephasing for which the resulting loss-dephasing channel is anti-degradable, implying the absence of codes capable of correcting its effect or, alternatively, capable of enabling quantum communication. A conjecture in Ref. [1] suggested that the bosonic loss-dephasing channel is not anti-degradable if the loss is below 50%. In this Chapter we refute this conjecture, specifically proving that for any value of the loss, if the dephasing is above a critical value, then the bosonic loss-dephasing channel is anti-degradable. While our result identifies a large parameter region where quantum communication is not possible, we also prove that if two-way classical communication is available, then quantum communication — and thus quantum key distribution — is always achievable, even for high values of loss and dephasing.

[This chapter is based on F. A. Mele, F. Salek, V. Giovannetti, L. Lami, Quantum communication on the bosonic loss-dephasing channel, Physical Review A 110, 012460 (2024)]

6.1 Introduction

Quantum optical platforms are key elements of quantum technologies, contributing significantly to both quantum communication and quantum computation [2–10]. Since the potential benefits of quantum technologies are hindered by the presence of decoherence [11], the investigation of decoherence sources affecting bosonic systems and the development of bosonic quantum error-correcting codes have been extensively analysed in recent years [12–23]. The primary noise processes in bosonic systems that act as dominant sources of decoherence are *photon loss* and *bosonic dephasing* [24–26], which have been both extensively analysed [14, 27–29]. Loss affects the system by causing it to dissipate some of its energy, whereas dephasing works to transform coherent superpositions into probabilistic mixtures. Although loss and dephasing sources can simultaneously affect bosonic systems [16, 30], such as in superconducting systems [31, 32], the existing literature provides only partial results about their combined effect [1]. On a technical level, understanding the combined effect of loss and dephasing is challenging due to the conflicting behaviours they exhibit: the action of loss takes a simple form when written in the coherent state basis but is complicated to analyse in the Fock basis [3], whereas dephasing demonstrates the opposite pattern, making the analysis of their combined effect quite intricate.

Consider an optical link (e.g. an optical fibre or a free-space link) or a quantum memory affected by both loss and dephasing, where the link is used for quantum communication and the memory for quantum computation. A crucial challenge is to determine the conditions under which

there exist protocols capable of enabling reliable quantum communication across the optical link or capable of mitigating the combined noise affecting the quantum memory. This problem is closely related to the anti-degradability condition in quantum Shannon theory: if a noise channel is *anti-degradable* [33, 34], there are no quantum communication protocols for reliable information transmission or quantum error-correcting codes capable of overcoming it. Consequently, it is crucial to understand whether the combined effect of loss and dephasing results in an anti-degradable channel. This has been a puzzling problem, to the point that in [1] it was conjectured that the combined loss-dephasing noise does not result in an anti-degradable channel if the loss is below 50%.

In this paper we refute the above conjecture; specifically, we prove that for any value of the photon loss there exists a *critical value* of the dephasing above which the resulting *bosonic loss-dephasing channel* is anti-degradable. Our discovery thus identifies a large region of the loss-dephasing parameter space where correcting the noise and achieving reliable quantum communication is impossible. On the more positive side, however, we also prove that if the sender and the receiver are assisted by two-way classical communication, then reliable quantum communication — and thus quantum key distribution — is always possible, even in scenarios characterised by arbitrarily high levels of loss and dephasing. Ours are the first analytical results to characterise the transmission of quantum information in the presence of both loss and dephasing noise.

The structure of the paper is as follows. In Section 6.2, we briefly review some preliminary notions necessary for stating our results. In Section 6.3, we present our results on the anti-degradability of the bosonic loss-dephasing channel, refuting a conjecture put forth in [1]. In Section 6.4, we analyse quantum communication across the bosonic loss-dephasing channel assisted by two-way classical communication. In Section 6.5, we discuss the degradability of the loss-dephasing channel. In the Appendix, we provide detailed derivations and additional results concerning the bosonic loss-dephasing channel.

6.2 Preliminaries

The *quantum capacity* $Q(\mathcal{N})$ of a quantum channel $\mathcal{N}$ quantifies the efficiency in transmitting qubits reliably across $\mathcal{N}$ [33, 34]. The condition $Q(\mathcal{N}) = 0$ implies that there exist neither reliable quantum communication protocols across $\mathcal{N}$ nor codes capable of correcting the errors induced by $\mathcal{N}$. Accordingly, if $\mathcal{N}$ is *anti-degradable* its quantum capacity vanishes [33]. This underscores the significance of determining whether a channel is anti-degradable, as the noise associated with such a channel cannot be corrected. By definition, a channel $\mathcal{N}$ is anti-degradable if there exists a channel $\mathcal{A}$ — called the *anti-degrading map* — such that $\mathcal{A} \circ \mathcal{N}^{\mathrm{c}} = \mathcal{N}$, where $\mathcal{N}^{\mathrm{c}}$ denotes a complementary channel of $\mathcal{N}$ [33]. Conversely, a channel $\mathcal{N}$ is degradable if there exists a channel $\mathcal{D}$ such that $\mathcal{D} \circ \mathcal{N} = \mathcal{N}^{\mathrm{c}}$. Degradable channels are theoretically important because their quantum capacity can be calculated as the single-letter coherent information of the channel [33–35].

The phenomenon of photon loss is mathematically modelled by the *pure-loss channel* $\mathcal{E}_\lambda$ [3, 5], a single-mode continuous-variable channel that acts on the input state ρ by mixing it with an environmental vacuum state in a beam splitter of transmissivity $\lambda \in [0, 1]$:

$$\mathcal{E}_\lambda(\rho) \coloneqq \mathrm{Tr}_E \left[U_\lambda \left(\rho_S \otimes |0\rangle\langle 0|_E \right) U_\lambda^\dagger \right] , \tag{6.1}$$

where $U_\lambda \coloneqq \exp\left[\arccos(\sqrt{\lambda})(\hat{a}^\dagger \hat{e} - \hat{a}\, \hat{e}^\dagger)\right]$ is the beam splitter unitary, $\hat{a}$ and $\hat{e}$ are the annihilation operators of the input system S and of the environment E, and Tr_E is the partial trace w.r.t. E. When a single photon is fed into $\mathcal{E}_\lambda$, it is transmitted to the output with probability λ, while it is lost to the environment with probability $1-\lambda$. More generally, if n photons are fed into the channel, the output is given by the binomial probability mixture $\mathcal{E}_\lambda(|n\rangle\langle n|) = \sum_{\ell=0}^{n} \binom{n}{l} (1-\lambda)^\ell \lambda^{n-\ell} |n-\ell\rangle\langle n-\ell|$, where $|n\rangle$ denotes the Fock state with n photons [3]. When $\lambda = 1$ the pure-loss channel is noiseless,

while when $\lambda = 0$ it is completely noisy — it maps any state into the vacuum. It is known that the pure-loss channel is anti-degradable for $\lambda \in [0, \frac{1}{2}]$ and degradable for $\lambda \in [\frac{1}{2}, 1]$ [36–39].

The phenomenon of bosonic dephasing is mathematically described by the *bosonic dephasing channel* $\mathcal{D}_\gamma$ [1, 29, 40], which maps the state $\rho = \sum_{m,n=0}^{\infty} \rho_{mn} |m\rangle\langle m| n$, written in the Fock basis, to

$$\mathcal{D}_\gamma(\rho) \coloneqq \sum_{m,n=0}^{\infty} \rho_{mn} e^{-\frac{\gamma}{2}(m-n)^2} |m\rangle\langle m| n \,, \tag{6.2}$$

resulting in a reduction in magnitude of the off-diagonal elements. When $\gamma = 0$, the bosonic dephasing channel is noiseless. In contrast, when $\gamma \to \infty$, it completely annihilates all off-diagonal components of the input density matrix, reducing it to an incoherent probabilistic mixture of Fock states. Moreover, the bosonic dephasing channel is never anti-degradable and it is always degradable [40].

Consider an optical system undergoing simultaneous loss and dephasing over a finite time interval. At each instant, the system is susceptible to both an infinitesimal pure-loss channel and an infinitesimal bosonic dephasing channel. Hence, the overall channel, which describes the simultaneous effect of loss and dephasing, results in a suitable composition of numerous concatenations between infinitesimal pure-loss and bosonic dephasing channels. However, given that (i) the pure-loss channel and the bosonic dephasing channel commute, $\mathcal{E}_\lambda \circ \mathcal{D}_\gamma = \mathcal{D}_\gamma \circ \mathcal{E}_\lambda$; (ii) the composition of pure-loss channels is a pure-loss channel, $\mathcal{E}_{\lambda_1} \circ \mathcal{E}_{\lambda_2} = \mathcal{E}_{\lambda_1 \lambda_2}$; and (iii) the composition of bosonic dephasing channels is a bosonic dephasing channel, $\mathcal{D}_{\gamma_1} \circ \mathcal{D}_{\gamma_2} = \mathcal{D}_{\gamma_1 + \gamma_2}$; it follows that the combined effect of loss and dephasing can be modelled by the composition

$$\mathcal{N}_{\lambda,\gamma} \coloneqq \mathcal{E}_\lambda \circ \mathcal{D}_\gamma \,, \tag{6.3}$$

which we will refer to as the *bosonic loss-dephasing channel*.

6.3 Anti-degradability

Prior to this work, the only result on the anti-degradability of the bosonic loss-dephasing channel was that it is anti-degradable if the transmissivity is below $\frac{1}{2}$ [1]. This result trivially follows from the anti-degradability of the pure-loss channel for transmissivities below $\frac{1}{2}$, and the fact that the composition of an anti-degradable channel with another channel inherits the property of being anti-degradable (see Lemma 71 in the Appendix). Notably, in the regime $\lambda > \frac{1}{2}$, it was an open question to understand whether or not the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable for some values of the dephasing γ, and in [1] the answer was conjectured to be negative. However, in the forthcoming Theorem 1, we show that the latter conjecture is incorrect, specifically, we prove that for all $\lambda \in [0, 1)$, if γ is sufficiently large, then $\mathcal{N}_{\lambda,\gamma}$ becomes anti-degradable.

Theorem 1. The bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable if the transmissivity λ and the dephasing γ fall within one of the following regions: (i) $\lambda \in [0, \frac{1}{2}]$ and $\gamma \geq 0$; (ii) $\lambda \in (\frac{1}{2}, 1)$ and γ such that $\theta\Big(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\Big) \leq \frac{3}{2}$, where $\theta(x, y) \coloneqq \sum_{n=0}^{\infty} x^{n^2} y^n$. A weaker but simpler sufficient condition that implies anti-degradability is given by $\lambda \leq \max\Big(\frac{1}{2}, \frac{1}{1+9e^{-\gamma}}\Big)$.

Here, we present a sketch of the proof, with a detailed version of the proof provided in Theorem 4 in the Appendix.

Proof sketch. Any finite dimensional channel $\mathcal{N}$ is anti-degradable if and only if its Choi state is *two-extendible* [41], meaning that there exists a tripartite state $\rho_{AB_1B_2}$ such that the reduced states on AB_1 and AB_2 both coincide with the Choi state:

$$\begin{aligned} \mathrm{Tr}_{B_2}\left[\rho_{AB_1B_2}\right] &= C_{AB_1}(\mathcal{N}) \,, \\ \mathrm{Tr}_{B_1}\left[\rho_{AB_1B_2}\right] &= C_{AB_2}(\mathcal{N}) \,, \end{aligned} \tag{6.4}$$

where the Choi state is defined as $C_{AB}(\mathcal{N}) \coloneqq \mathrm{Id}_A \otimes \mathcal{N}_{A'\to B}(\Phi_{AA'})$, with $\Phi_{AA'}$ being the maximally entangled state. Such a characterisation extends to infinite dimension by considering the generalised Choi state $C^{(r)}_{AB}(\mathcal{N}) \coloneqq \mathrm{Id}_A \otimes \mathcal{N}_{A'\to B}\left(\Psi^{(r)}_{AA'}\right)$ [42], obtained by replacing $\Phi_{AA'}$ by the two-mode squeezed vacuum state $\Psi^{(r)}_{AA'}$ with squeezing parameter $r > 0$ [3]. The crux of our proof is to find a two-extension of $C^{(r)}_{AB}(\mathcal{N}_{\lambda,\gamma})$ in the region identified by condition (ii). We do this in two steps.

First, after scrutinising the matrix $C^{(r)}_{AB}(\mathcal{N}_{\lambda,\gamma})$ written in the Fock basis, we construct a tripartite state $\tau_{AB_1B_2}$ such that the reduced states on AB_1 and AB_2 have the same diagonal as $C^{(r)}_{AB}(\mathcal{N}_{\lambda,\gamma})$, and the same pattern of vanishing off-diagonal entries. The construction of this tripartite state involves applying several channels — namely, beam splitter unitaries, squeezing unitary, partial trace, and a three mode controlled-add-add isometry — to a 4-mode vacuum state.

The second step consists in transforming $\tau_{AB_1B_2}$ into a two-extension of $C^{(r)}_{AB}(\mathcal{N}_{\lambda,\gamma})$ by tweaking its off-diagonal entries. This is done by using the toolbox of *Hadamard maps* [34]. For any matrix $A \coloneqq (a_{mn})_{m,n\in\mathbb{N}}$, the associated Hadamard map $H^{(A)}$ is defined by

$$H^{(A)}(|m\rangle\langle m|\, n) = a_{mn}\, |m\rangle\langle m|\, n \tag{6.5}$$

for all m, n [34]. In practice, $H^{(A)}$ acts on the input density matrix by multiplying each (m, n) entry by the corresponding coefficient a_{mn}. Importantly, $H^{(A)}$ is a quantum channel if and only if A is Hermitian, positive semi-definite, and has all 1's on the main diagonal [34]. The crucial observation is that it is always possible to find an infinite matrix $A_{\lambda,\gamma}$ (possibly not positive semi-definite), which is real, symmetric, and has all 1's on the main diagonal, such that the operator $\mathrm{Id}_A \otimes H^{(A_{\lambda,\gamma})}_{B_1} \otimes H^{(A_{\lambda,\gamma})}_{B_2}(\tau_{AB_1B_2})$ coincides with $C^{(r)}_{AB}(\mathcal{N}_{\lambda,\gamma})$ when tracing out either B_1 or B_2.

This however does not mean that we have found a two-extension of $C^{(r)}_{AB}$, because the above operator is not necessarily a state — it may fail to be positive semi-definite. It *is* a state, however, whenever $H^{(A_{\lambda,\gamma})}$ is a quantum channel, i.e. when the infinite matrix $A_{\lambda,\gamma}$ is positive semi-definite, in formula $A_{\lambda,\gamma} \geq 0$. Therefore, a sufficient condition on the anti-degradability of $\mathcal{N}_{\lambda,\gamma}$ is that $A_{\lambda,\gamma} \geq 0$.

The rest of the proof consists in showing that under condition (ii) one indeed finds $A_{\lambda,\gamma} \geq 0$. This is not straightforward to check, because $A_{\lambda,\gamma}$ is an *infinite* matrix, and it cannot be diagonalised analytically nor numerically. To by-pass this last hurdle we employ the theory of diagonally dominant matrices, and in particular the statement that if a matrix A is such that $a_{nn} - \sum_{m:\, m\neq n} |a_{mn}| \geq 0$ for all n, then necessarily $A \geq 0$ [43, Chapter 6]. We demonstrate that if λ and γ satisfy condition (ii), then $A_{\lambda,\gamma}$ satisfies this condition, which establishes that $A_{\lambda,\gamma} \geq 0$ and hence concludes the proof. For a more detailed proof, please refer to Theorem 4 in the Appendix. □

Theorem 1 identifies a region of the parameter space (λ, γ), with λ identifying the transmissivity and γ the dephasing, where the channel is anti-degradable and thus its quantum capacity vanishes, thereby implying the absence of viable error correcting codes for quantum data transfer and storage. This region is illustrated in Fig. 6.1. Interestingly, Theorem 1 implies that even if $\lambda > \frac{1}{2}$ one can pick γ large enough so that there exists an anti-degrading map achieving the transformation $\mathcal{N}^{\mathrm{c}}_{\lambda,\gamma}(|n\rangle\langle n|_F) \longrightarrow \mathcal{N}_{\lambda,\gamma}(|n\rangle\langle n|_F)$, which can be expressed as (see Lemma 52 in the Appendix)

$$\mathcal{E}_{1-\lambda}(|n\rangle\langle n|_F) \otimes |\sqrt{\gamma}n\rangle\langle\sqrt{\gamma}n|_C \longrightarrow \mathcal{E}_\lambda(|n\rangle\langle n|_F)\,, \tag{6.6}$$

where $|n\rangle_F$ denotes the nth Fock state and $|\sqrt{\gamma}n\rangle_C$ denotes a coherent state [3]. In Theorem 6 in the Appendix, we provide an explicit construction of such anti-degrading map. This entails the following remarkable fact: for $\lambda > 1/2$ and large enough γ there exists an *n-independent* strategy to convert the lossy Fock state $\mathcal{E}_{1-\lambda}(|n\rangle\langle n|_F)$ into the less lossy Fock state $\mathcal{E}_\lambda(|n\rangle\langle n|_F)$ using the coherent state $|\sqrt{\gamma}n\rangle_C$ as a resource. In other words, one can undo part of the loss on $|n\rangle_F$ if one has a coherent state that contains some information on n, sufficiently amplified so

that that information is accessible enough. The nontrivial and somewhat surprising nature of this exact conversion strategy arises from the fact that the coherent states $\{|\sqrt{\gamma}n\rangle_C\}_{n\in\mathbb{N}}$ are not orthogonal, meaning that the strategy that consists in measuring the coherent state, guessing n, and re-preparing $\mathcal{E}_\lambda(|n\rangle\langle n|_F)$ cannot succeed with probability 1.

Theorem 1 does not identify the entire anti-degradability region of $\mathcal{N}_{\lambda,\gamma}$, but only a subset of it. One way to improve this approximation is to determine numerically the region where the infinite matrix $A_{\lambda,\gamma}$ introduced in the proof sketch of Theorem 1 is positive semi-definite. In Fig. 6.1 we depict a numerical estimate of this region (see the crossed red part of the plot).

So far we have been concerned with inner approximations of the anti-degradability region. To obtain outer approximations, instead, one can start by observing that the action of $\mathcal{N}_{\lambda,\gamma}$ can only subtract and never add any photons. Mathematically, if the input state to $\mathcal{N}_{\lambda,\gamma}$ is supported on the span of the first d Fock states, so is the output state. This restriction defines a qu*d*it-to-qu*d*it channel $\mathcal{N}^{(d)}_{\lambda,\gamma}$, analysing which can yield some insights into $\mathcal{N}_{\lambda,\gamma}$ itself. First, if $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is *not* anti-degradable then the same is true of $\mathcal{N}_{\lambda,\gamma}$ (see Corollary 57 in the Appendix); secondly, as discussed above the anti-degradability of $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is equivalent to the two-extendibility of the corresponding Choi state [41], and for moderate values of d this latter condition can be efficiently checked numerically via *semi-definite programming* [34, 44]. In this way, as discussed in Section 6.7.2 of the Appendix, we can numerically determine a parameter region (see green region of Fig. 6.1) where $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable.

Interestingly, already the qubit restriction $\mathcal{N}^{(2)}_{\lambda,\gamma}$, which coincides with the composition between the amplitude damping channel and the qubit dephasing channel [34], yields the necessary condition $\lambda \leq \frac{1}{1+e^{-\gamma}}$ on the anti-degradability of $\mathcal{N}_{\lambda,\gamma}$, as shown in the forthcoming Theorem 2. Based on the analysis of the qudit restrictions, we conjecture that, if γ is sufficiently large, the latter condition $\lambda \leq \frac{1}{1+e^{-\gamma}}$ is not only necessary but also sufficient.

Theorem 2. If $\lambda > \frac{1}{1+e^{-\gamma}}$ then $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable.

Proof. A qubit channel $\mathcal{N}$ is anti-degradable if and only if

$$\frac{1}{4}\operatorname{Tr}\left[\mathcal{N}(\mathbb{1}_2)^2\right] \geq \operatorname{Tr}\left[C(\mathcal{N})^2\right] - 4\sqrt{\det[C(\mathcal{N})]}, \tag{6.7}$$

where $C(\mathcal{N})$ is the Choi state [41, 45, 46]. By employing the condition in (6.7) together with the expression of the Choi state of the qubit restriction $\mathcal{N}^{(2)}_{\lambda,\gamma}$ provided in (6.62) in the Appendix, one can show that $\mathcal{N}^{(2)}_{\lambda,\gamma}$ is anti-degradable if and only if $\lambda \leq \frac{1}{1+e^{-\gamma}}$. Hence, we conclude that the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable if $\lambda > \frac{1}{1+e^{-\gamma}}$. □

We are interested in the anti-degradability of the loss-dephasing channel because it implies that the quantum capacity vanishes, entailing the impossibility of quantum communication. We now look at the complementary question: when is the quantum capacity $Q(\mathcal{N}_{\lambda,\gamma})$ strictly positive? A simple sufficient condition can be obtained by optimising the coherent information [33, 34] of $\mathcal{N}_{\lambda,\gamma}$ over input states of the form $\rho_p := p\,|0\rangle\langle 0| + (1-p)\,|1\rangle\langle 1|$. By doing so we identify a region of the (λ,γ) parameter space where $Q(\mathcal{N}_{\lambda,\gamma}) > 0$ (see the crossed green region in Fig. 6.1). In this region quantum communication and quantum error correction become feasible.

6.4 Two-way quantum communication

As we have just seen, (unassisted) quantum communication is not possible when the combined effects of loss and dephasing are too strong. However, in the forthcoming Theorem 3 we show that if Alice (the sender) and Bob (the receiver) have access to a *two-way* classical communication line, then quantum communication, entanglement distribution, and quantum-key distribution [47]

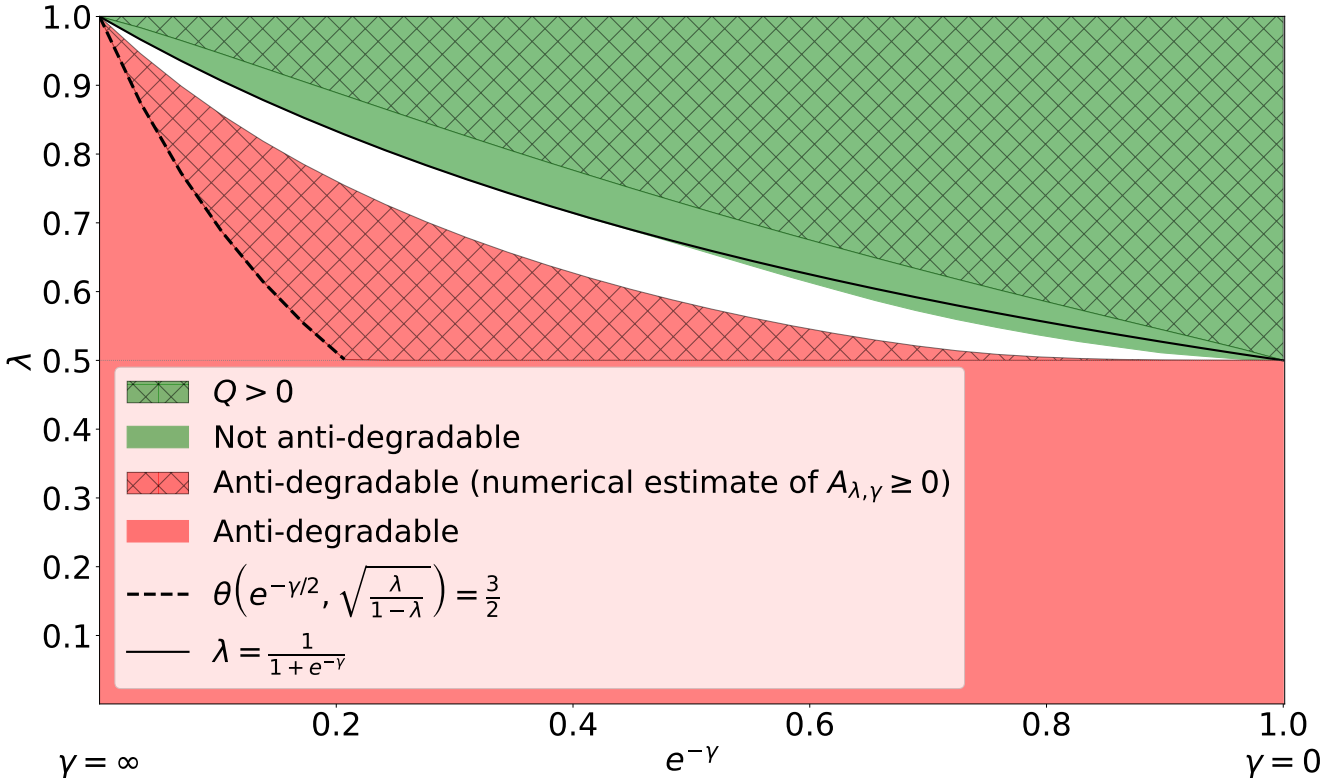


Figure 6.1: Summary of results on the anti-degradability of the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$. The vertical axis represents the transmissivity λ, while the horizontal axis corresponds to $e^{-\gamma}$, where γ is the dephasing parameter. In the green region $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable, while in the red region it is anti-degradable. In the crossed green region, the quantum capacity of $\mathcal{N}_{\lambda,\gamma}$ is strictly positive. The crossed red region is a numerical estimate of the region where the infinite matrix $A_{\lambda,\gamma}$ is positive semi-definite, a condition implying that $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable, as explained in the proof sketch of Theorem 1. Such an estimate can be obtained by examining the positive semi-definiteness of the $d \times d$ top-left corner of $A_{\lambda,\gamma}$ for large values of d (here we employ $d = 30$, but increasing d already beyond $d \geq 20$ yields no discernible change in the plot). The restriction $\mathcal{N}^{(6)}_{\lambda,\gamma}$ is anti-degradable if and only if λ and γ fall within the green region, and this is the reason why $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable in the green region. Below the curve $\theta\left(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\right) = \frac{3}{2}$, the channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable, as stated in Theorem 1. Above the curve $\lambda = \frac{1}{1+e^{-\gamma}}$, the channel is not anti-degradable, as guaranteed by Theorem 2.

become again achievable for any value of loss and dephasing, even when Alice's input signals are constrained to have limited energy.

In this two-way communication setting the relevant notion of capacity is the *two-way quantum capacity* $Q_2(\mathcal{N})$ [33, 34], defined as the maximum achievable rate of qubits that can be reliably transmitted across $\mathcal{N}$ with the aid of two-way classical communication. Since in practice Alice has only a limited amount of energy to produce her input signals, one usually defines the so-called *energy-constrained* two-way quantum capacity [48, 49], denoted as $Q_2(\mathcal{N}, N_s)$. Here, N_s denotes the mean photon number constraint at the input of the channel.

Theorem 3. For all $N_s > 0$, $\lambda \in (0, 1]$, and $\gamma \geq 0$, the energy-constrained two-way quantum capacity of the bosonic loss-dephasing channel is strictly positive, i.e. $Q_2(\mathcal{N}_{\lambda,\gamma}, N_s) > 0$. In particular, $\mathcal{N}_{\lambda,\gamma}$ is not entanglement breaking. An explicit lower bound is

$$Q_2(\mathcal{N}_{\lambda,\gamma}, N_s) \geq \frac{\lambda^N}{k}\left[\log_2\binom{N+k-1}{N} - S(\rho_{N,k,\gamma})\right] \tag{6.8}$$

for any $N, k \in \mathbb{N}_+$ satisfying $\frac{N}{k} \leq N_s$. Here, $S(\cdot)$ is the von Neumann entropy, $\rho_{N,k,\gamma}$ is a

$\binom{N+k-1}{N}$-dimensional state defined by

$$\rho_{N,k,\gamma} := \binom{N+k-1}{N}^{-1} \sum_{p,q\in\Pi(N,k)} e^{-\frac{\gamma}{2}\|p-q\|_2^2} |p\rangle\langle p| \, q \,, \tag{6.9}$$

where $\Pi(N,k) := \left\{p \in \mathbb{N}^k : \sum_{i=1}^k p_i = N\right\}$ represents the set of partitions of a set of N elements into k parts, and the vectors $\{|p\rangle\}_{p\in\Pi(N,k)}$ are orthonormal.

Here, we provide a sketch of the proof, while in Theorem 12 in the Appendix we provide a detailed proof.

Proof sketch. The proof exploits entanglement transmission protected by a particular error correction technique, *rail encoding*. In a k-mode bosonic system, consider the subspace $V_{N,k}$ corresponding to a total photon number N. We can use this subspace, whose dimension is $d_{N,k} := \dim V_{N,k} = \binom{N+k-1}{N}$, as an error correction code that protects against the detrimental action of $\mathcal{N}_{\lambda,\gamma}$. To this end, we prepare a maximally entangled state of dimension $d_{N,k}$ and send one share of it through k copies of the channel $\mathcal{N}_{\lambda,\gamma}$, one per mode. Since under the action of $\mathcal{N}_{\lambda,\gamma}$ photons can only be lost and never added, and each photon has a probability λ of being transmitted, the probability that an N-photon state will retain all of its photons at the output of the channel is exactly λ^N. If this happens to be the case, which — crucially — can be certified by a total photon number measurement at the output, then the input state has been subjected to no loss and only dephasing. The entanglement of the resulting, maximally correlated state can be distilled via an explicit protocol known as the hashing protocol [34, 50], resulting in $\log_2 d_{N,k} - S(\rho_{N,k,\gamma}) > 0$ singlet (a.k.a. ebit, i.e. unit of entanglement) yield. The strict positivity of this yield follows by observing that $\rho_{N,k,\gamma}$ is a $d_{N,k}$-dimensional mixed state that is not maximally mixed. □

6.5 Degradability

The bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is never degradable, except in the simple cases when either $\lambda = 1$ or $\gamma = 0$ and $\lambda \geq 1/2$ [1]. This in turn implies that no single-letter formula for its quantum capacity is known outside of the anti-degradability region studied here, where the capacity vanishes. The failure of degradability has been demonstrated in [1] through a lengthy proof; we are now in position to provide an alternative, much simpler argument. The key ideas are as follows: (i) If the qubit restriction $\mathcal{N}^{(2)}_{\lambda,\gamma}$ is not degradable, then $\mathcal{N}_{\lambda,\gamma}$ is not degradable either (see Corollary 57 in the Appendix); and (ii) if the rank of the Choi state of a qubit channel is greater or equal to 3 than such channel is not degradable [51, Theorem 4]. The result then follows by observing that the Choi state of the qubit channel $\mathcal{N}^{(2)}_{\lambda,\gamma}$, as provided in (6.62) in the Appendix, has a rank exactly equal to 3. For a detailed proof see Theorem 7 in the Appendix.

6.6 Conclusion

In this paper we have provided the first analytical investigation of the quantum communication capabilities of the bosonic loss-dephasing channel, a much more realistic model of noise than dephasing and loss treated separately. Refuting a conjecture put forth in [1], we showed that the bosonic loss-dephasing channel is anti-degradable in a large region of the loss-dephasing parameter space, entailing that neither quantum communication nor quantum error correcting codes are possible in this region. On the positive side, we also showed that if two-way classical communication is suitably exploited, then quantum communication is always achievable, even in scenarios characterised by high levels of loss and dephasing, and even in the presence of stringent energy constraints.

Two fundamental technical innovations are key to our approach. First, a new method to analyse anti-degradability of bosonic channels, based on a two-stage construction of a symmetric extension of the Choi state. The introduction of this technique is crucial here also on the conceptual level, as *all* other known tools to analyse quantum capacities (e.g. degradability [34], PPT-ness [52], or teleportation simulability [53, 54]) fail completely for the loss-dephasing channel [1]. In Section 6.7.3 in the Appendix, we envision that our technique could also be applied to other cases, e.g. to analyse the anti-degradability of the composition between the pure-loss channel and a *general* bosonic dephasing channel. The second innovation we introduce is based on the use of rail encoding to investigate two-way assisted entanglement generation on the loss-dephasing channel. This technique, which we anticipate may be used to study general processes where photon loss is involved, has the additional benefit of yielding an explicit entanglement generation protocol.

Although the capacities of the dephasing channel and the pure-loss channel (separately) have been determined [29, 38, 39, 53, 55–58], the capacities of the channel resulting from their combined action remain unknown. An intriguing open problem is to calculate or approximate these capacities.

6.7 Supplementary material

6.7.1 Preliminaries and notation

Quantum states and Channels

In this subsection, we present a summary of the notation and fundamental properties used in the paper, drawing from the conventions established in standard quantum information theory textbooks [33, 34, 44, 59]. Every quantum system is associated with a separable complex Hilbert space $\mathcal{H}$ whose dimension is denoted by $|\mathcal{H}|$. We use subscripts to denote the system associated to a Hilbert space and also systems on which the operators act. The composite quantum systems A and B exist within the tensor product of their individual Hilbert spaces $\mathcal{H}_A \otimes \mathcal{H}_B$ which is also denoted by $\mathcal{H}_{AB}$.

We shall use $\mathbb{1}$ to denote the identity operator on $\mathcal{H}$. The operator norm of a linear operator $\Theta : \mathcal{H} \to \mathcal{H}$ is defined as

$$\|\Theta\|_\infty := \sup_{|\psi\rangle \in \mathcal{H}:\, \langle\psi|\psi\rangle = 1} \sqrt{\langle\psi|\Theta^\dagger\Theta|\psi\rangle}\,. \tag{6.10}$$

An alternative (but equivalent) definition of the operator norm is as follows:

$$\|\Theta\|_\infty := \sup_{|v\rangle, |w\rangle \in \mathcal{H},\, \langle v|v\rangle = \langle w|w\rangle = 1} |\langle v|\,\Theta\,|w\rangle|\,. \tag{6.11}$$

An operator is called bounded if its operator norm is bounded, i.e. $\|\Theta\|_\infty < \infty$. A bounded operator Θ is positive semi-definite if $\langle\psi|\,\Theta\,|\psi\rangle \geq 0, \forall\ |\psi\rangle \in \mathcal{H}$, while it is positive definite if $\langle\psi|\,\Theta\,|\psi\rangle > 0, \forall\ |\psi\rangle \in \mathcal{H}$. The trace norm of a linear operator $\Theta : \mathcal{H} \to \mathcal{H}$ is defined as $\|\Theta\|_1 := \mathrm{Tr}\sqrt{\Theta^\dagger\Theta}$. The set of trace class operators, denoted as $\mathcal{T}(\mathcal{H})$, is the set of all the linear operators on $\mathcal{H}$ such that their trace norm is finite, i.e. $\|\Theta\|_1 < \infty$. The operator and trace norm satisfy $\|\Theta\|_\infty \leq \|\Theta\|_1$. The set of quantum states (density operators), denoted as $\mathcal{P}(\mathcal{H})$, is the set of positive semi-definite trace class operators on $\mathcal{H}$ with unit trace. The *fidelity* between two quantum states $\rho, \sigma \in \mathcal{P}(\mathcal{H})$ is defined as $F(\rho, \sigma) := \mathrm{Tr}[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}]$.

A superoperator is a linear map between spaces of linear operators. The identity superoperator will be denoted as Id. Quantum channels are completely-positive trace-preserving (cptp) superoperators. In this paper, we will use two different representations of a quantum channel that are known as Stinespring and Choi–Jamiołkowski representation. A quantum channel $\mathcal{N}_{A'\to B}$ can be represented in Stinespring representation as

$$\mathcal{N}_{A'\to B}(\cdot) = \mathrm{Tr}_E\left[U_{A'E\to BE}(\cdot \otimes |0\rangle\langle 0|_E)U^\dagger_{A'E\to BE}\right].$$

Here, E is an environment system, $|0\rangle_E$ is a pure state of the environment, and $U_{A'E\to BE}$ is an isometry that takes as input the systems A' and E and outputs the systems B, E. The associated complementary channel $\mathcal{N}^{\mathrm{c}}_{A'\to B}$ is given by

$$\mathcal{N}^{\mathrm{c}}_{A'\to E}(\cdot) = \mathrm{Tr}_B\left[U_{A'E\to BE}(\cdot\otimes|0\rangle\langle 0|_E)U^\dagger_{A'E\to BE}\right].$$

A channel $\mathcal{N}$ is called degradable if there exist a quantum channel $\mathcal{J}$, such that when is used after $\mathcal{N}$, we get back to the complementary channel $\mathcal{N}^{\mathrm{c}}$, i.e. $\mathcal{J}\circ\mathcal{N}=\mathcal{N}^{\mathrm{c}}$. On the other hand, a channel is called anti-degradable if there is another quantum channel $\mathcal{A}$, such that using it after the complementary channel, gives back the original channel, i.e. $\mathcal{A}\circ\mathcal{N}^{\mathrm{c}}=\mathcal{N}$. The channels $\mathcal{J}$ and $\mathcal{A}$ are usually called the degrading map and the anti-degrading map of $\mathcal{N}$, respectively.

The Choi–Jamiołkowski representation of the channel $\mathcal{N}_{A'\to B}$ is the operator $C(\mathcal{N})\in\mathcal{T}(\mathcal{H}_A\otimes\mathcal{H}_B)$ that is defined as

$$C(\mathcal{N}) \coloneqq \mathrm{Id}_A\otimes\mathcal{N}_{A'\to B}(|\Phi\rangle\langle\Phi|_{AA'}), \tag{6.12}$$

where $|\Phi\rangle=\frac{1}{\sqrt{|\mathcal{H}_A|}}\sum_{i=0}^{|\mathcal{H}_A|-1}|i\rangle_A\otimes|i\rangle_{A'}$ is a maximally entangled state of schmidt rank $|\mathcal{H}_A|$, the set of states $\{|i\rangle\}_{i=0,\ldots,|\mathcal{H}_A|-1}$ forms a basis for $\mathcal{H}_A$, and $\mathcal{H}_A=\mathcal{H}_{A'}$. It is a well-established fact that the superoperator $\mathcal{N}$ is a quantum channel if and only if $C(\mathcal{N})\geq 0$ and $\mathrm{Tr}_B\,C(\mathcal{N})=\mathbb{1}_A/|\mathcal{H}_A|$ [33].

Definition 29. *A bipartite state ρ_{AB} is symmetric two-extendible on B if there exists a tripartite state $\tau_{AB_1B_2}$ such that*

- $F_{B_1B_2}\tau_{AB_1B_2}F^\dagger_{B_1B_2}=\tau_{AB_1B_2}$
- $\mathrm{Tr}_{B_1}\,\tau_{AB_1B_2}=\rho_{AB}$,

where B_1 and B_2 are two copies of the system B, the operator $F_{B_1B_2}\coloneqq\sum_{i,j}|i\rangle\langle i|_{B_1}\otimes|j\rangle\langle j|_{B_2}$ denotes the swap unitary, and $\{|i\rangle_{B_1}\}_i$ and $\{|i\rangle_{B_2}\}_i$ form an orthonormal basis. The state $\tau_{AB_1B_2}$ is called a symmetric two-extension of ρ_{AB} on B.

Definition 30. *A bipartite state ρ_{AB} is called two-extendible on B if there exists a tripartite state $\tau_{AB_1B_2}$ such that*

$$\mathrm{Tr}_{B_1}\,\tau_{AB_1B_2}=\mathrm{Tr}_{B_2}\,\tau_{AB_1B_2}=\rho_{AB}, \tag{6.13}$$

where B_1 and B_2 are two copies of the system B.

Lemma 31 [60]**.** *A bipartite state ρ_{AB} is two-extendible on B if and only if it is symmetric two-extendible on B .*

Proof. First, assume ρ_{AB} is symmetric two-extendible on B. Since $F_{B_1B_2}\tau_{AB_1B_2}F^\dagger_{B_1B_2}=\tau_{AB_1B_2}$, it holds that $\mathrm{Tr}_{B_2}\,\tau_{AB_1B_2}=\mathrm{Tr}_{B_1}\,\tau_{AB_1B_2}$. This implies that ρ_{AB} is two-extendible on B. Second, let ρ_{AB} be two-extendible on B. One can easily check that the state $1/2(\tau_{AB_1B_2}+F_{B_1B_2}\tau_{AB_1B_2}F^\dagger_{B_1B_2})$ is a symmetric two-extension of ρ_{AB}. □

It has been demonstrated that a quantum channel is anti-degradable if and only if its Choi state is two-extendible on the output system [41]. This equivalence leads to a simple necessary and sufficient condition for the anti-degradability of qubit channels:

Lemma 32. *[45, Corollary 4](See also [41, 46]) Any qubit quantum channel $\mathcal{N}$ is anti-degradable if and only if it satisfies*

$$\mathrm{Tr}\left[\left(\mathcal{N}\left(\frac{\mathbb{1}_2}{2}\right)\right)^2\right]\geq\mathrm{Tr}\left[(C(\mathcal{N}))^2\right]-4\sqrt{\det(C(\mathcal{N}))},$$

where $\mathbb{1}_2$ denotes the identity operator on the qubit Hilbert space.

Bosonic systems

In this subsection, we will provide an overview of relevant definitions and properties concerning quantum information with continuous variable systems; refer to [3] for detailed explanations. A single-mode of electromagnetic radiation with definite frequency and polarisation is represented by the Hilbert space $L^2(\mathbb{R})$, which comprises all square-integrable complex-valued functions over $\mathbb{R}$. Let $\mathbb{N}_+$ be the set of strictly positive integers and let $\mathbb{N} \coloneqq \{0\} \cup \mathbb{N}_+$. For any $n \in \mathbb{N}$, the construction of the *Fock state* $|n\rangle$ (the quantum state with n photons) involves the application of the bosonic creation operator $\hat{a}^\dagger$ to the *vacuum state* $|0\rangle$:

$$|n\rangle \coloneqq \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n |0\rangle \ . \tag{6.14}$$

The Fock states $\{|n\rangle\}_{n\in\mathbb{N}}$ form an orthonormal basis of $L^2(\mathbb{R})$. The bosonic annihilation operator $\hat{a}$ and creation operator $\hat{a}^\dagger$ satisfy the well-known canonical commutation relation $[\hat{a}, \hat{a}^\dagger] = \mathbb{1}$.

Let $\mathbb{C}$ be the set of complex numbers. For any $\alpha \in \mathbb{C}$, let $D(\alpha) \coloneqq e^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}}$ be the displacement operator. A coherent state of parameter α, denoted by $|\alpha\rangle$, is defined by applying the displacement operator $D(\alpha)$ to the vacuum state, i.e. $|\alpha\rangle \coloneqq D(\alpha)|0\rangle$. The overlap between coherent states is given by

$$\langle\alpha|\beta\rangle = e^{-\frac{1}{2}(|\alpha|^2+|\beta|^2-2\alpha^*\beta)} \ . \tag{6.15}$$

Quantum channels acting on bosonic systems are sometimes called bosonic channels. Similar to finite-dimensional channels, bosonic channels admit a Choi–Jamiołkowski representation, usually referred to as *generalised* Choi–Jamiołkowski representation [61, 62]. Consider isomorphic Hilbert spaces $\mathcal{H}_A, \mathcal{H}_{A'}$ which are possibly infinite dimensional. Let $|\psi\rangle_{AA'}$ be a pure state satisfying $\mathrm{Tr}_{A'}[|\psi\rangle\langle\psi|_{AA'}] > 0$ (See [63, Lemma 26]). The generalised Choi state is constructed by applying the channel to the subsystem A' of $|\psi\rangle_{A'A}$ (see Lemma 69 in the Appendix): $\mathrm{Id}_A \otimes \mathcal{N}_{A'\to B}(|\psi\rangle\langle\psi|_{AA'})$. The construction of the generalised Choi state usually utilises the two-mode squeezed vacuum state with squeezing parameter $r > 0$, defined as follows [3]:

$$|\psi(r)\rangle_{AA'} \coloneqq \frac{1}{\cosh(r)} \sum_{n=0}^{\infty} \tanh^n(r) |n\rangle_A |n\rangle_{A'} \ . \tag{6.16}$$

The equivalence between anti-degradability of a channel and two-extendibility of its Choi state extends to the infinite-dimensional channels [42]. We provide a detailed proof of this equivalence in Lemma 70 in the Appendix as it helps us in developing our intuition in inventing an explicit example of an anti-degrading map of the bosonic loss-dephasing channel.

Hadamard maps

Let $A = (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$, be an infinite matrix of complex numbers. Consider the superoperator H on $\mathcal{T}(L^2(\mathbb{R}))$, recognised as the *Hadamard map*, whose action on rank one operator $|m\rangle\langle m|\, n$ is defined as follows:

$$H(|m\rangle\langle m|\, n) = a_{mn} |m\rangle\langle m|\, n, \quad \forall\, m, n \in \mathbb{N} \, .$$

In Section 6.7.6 in the Appendix, we provide an overview of relevant properties of Hadamard maps. In particular, by combining known results about Hadamard maps and matrix analysis, in Lemma 67 in the Appendix, we show that given an infinite matrix $A = (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$, the associated Hadamard map is a quantum channel if

- A is Hermitian
- $a_{nn} = 1$, $\forall n \in \mathbb{N}$
- A is diagonally dominant, i.e. $\sum_{\substack{m=0 \\ m\neq n}}^{\infty} |a_{mn}| \leq 1$, $\forall\, n \in \mathbb{N}$.

Beam splitter

A beam splitter serves as a linear optical tool employed for creating quantum entanglement between two modes, referred to as the *system* mode (denoted as S) and the *environment* mode (denoted as E). A depiction of a beam splitter is reported in Fig 6.2.

Definition 33. *Let $\mathcal{H}_S, \mathcal{H}_E := L^2(\mathbb{R})$. Let $\hat{a}$ and $\hat{e}$ denote the annihilation operator of $\mathcal{H}_S$ and $\mathcal{H}_E$, respectively. For all $\lambda \in [0,1]$, the beam splitter unitary of transmissivity λ is given by*

$$U_\lambda^{SE} := \exp\left[\arccos\sqrt{\lambda}\left(\hat{a}^\dagger \hat{e} - \hat{a}\,\hat{e}^\dagger\right)\right] . \tag{6.17}$$

Lemma 34. *For all $\lambda \in [0,1]$, it holds that*

$$\begin{aligned}
(U_\lambda^{SE})^\dagger \hat{a}\, U_\lambda^{SE} &= \sqrt{\lambda}\,\hat{a} + \sqrt{1-\lambda}\,\hat{e}\,,\\
U_\lambda^{SE} \hat{a}\, (U_\lambda^{SE})^\dagger &= \sqrt{\lambda}\,\hat{a} - \sqrt{1-\lambda}\,\hat{e}\,,\\
(U_\lambda^{SE})^\dagger \hat{e}\, U_\lambda^{SE} &= -\sqrt{1-\lambda}\,\hat{a} + \sqrt{\lambda}\,\hat{e}\,,\\
U_\lambda^{SE} \hat{e}\, (U_\lambda^{SE})^\dagger &= \sqrt{1-\lambda}\,\hat{a} + \sqrt{\lambda}\,\hat{e}\,.
\end{aligned} \tag{6.18}$$

Proof. These identities can be readily proved by applying the Baker-Campbell-Hausdorff formula. For an alternative proof see [22, Lemma A.2]. □

Lemma 35. *For all $\lambda \in [0,1]$ and all $n \in \mathbb{N}$, it holds that*

$$U_\lambda^{SE} |n\rangle_S \otimes |0\rangle_E = \sum_{l=0}^{n} (-1)^l \sqrt{\binom{n}{l}} \lambda^{\frac{n-l}{2}} (1-\lambda)^{\frac{l}{2}} |n-l\rangle_S \otimes |l\rangle_E \,, \tag{6.19}$$

$$U_\lambda^{SE} |0\rangle_S \otimes |n\rangle_E = \sum_{l=0}^{n} \sqrt{\binom{n}{l}} (1-\lambda)^{\frac{l}{2}} \lambda^{\frac{n-l}{2}} |l\rangle_S \otimes |n-l\rangle_E \,. \tag{6.20}$$

Proof. Thanks to Lemma (34), we have that $U_\lambda^{SE} \hat{a} \left(U_\lambda^{SE}\right)^\dagger = \sqrt{\lambda}\hat{a} - \sqrt{1-\lambda}\hat{e}$. Consequently, it holds that

$$\begin{aligned}
U_\lambda^{SE} |n\rangle_S \otimes |0\rangle_E &= \frac{1}{\sqrt{n!}} U_\lambda^{SE} (a^\dagger)^n |0\rangle_S \otimes |0\rangle_E\\
&= \frac{1}{\sqrt{n!}} \left(U_\lambda^{SE} a^\dagger \left(U_\lambda^{SE}\right)^\dagger\right)^n |0\rangle_S \otimes |0\rangle_E\\
&= \frac{1}{\sqrt{n!}} \left(\sqrt{\lambda} a^\dagger - \sqrt{1-\lambda} b^\dagger\right)^n |0\rangle_S \otimes |0\rangle_E\\
&= \frac{1}{\sqrt{n!}} \sum_{l=0}^{n} (-1)^l \binom{n}{l} \lambda^{\frac{n-l}{2}} (1-\lambda)^{\frac{l}{2}} (a^\dagger)^{n-l} |0\rangle_S \otimes (b^\dagger)^l |0\rangle_E\\
&= \sum_{l=0}^{n} (-1)^l \sqrt{\binom{n}{l}} \lambda^{\frac{n-l}{2}} (1-\lambda)^{\frac{l}{2}} |n-l\rangle_S \otimes |l\rangle_E \,.
\end{aligned} \tag{6.21}$$

Analogously, one can show the validity of (6.20) by exploiting $U_\lambda^{SE} \hat{e}\, (U_\lambda^{SE})^\dagger = \sqrt{1-\lambda}\,\hat{a} + \sqrt{\lambda}\,\hat{e}$. □

Remark 36. *It is easily seen that Eq. (6.19) is equivalent to*

$$U_\lambda^{SE} |n\rangle_S \otimes |0\rangle_E = (-1)^n \sum_{l=0}^{n} (-1)^l \sqrt{\binom{n}{l}} \lambda^{\frac{l}{2}} (1-\lambda)^{\frac{n-l}{2}} |l\rangle_S \otimes |n-l\rangle_E \,.$$

Pure-loss channel

In optical platforms, the most common source of noise is photon loss, which is modelled by the *pure-loss channel* [3]. For any $\lambda \in [0,1]$, the pure-loss channel $\mathcal{E}_\lambda$ of transmissivity λ is a single-mode bosonic channel which acts on the input system by mixing it in a beam splitter of transmissivity λ with an environmental vacuum state; see Fig 6.2. In this model, the input signal is partially transmitted and partially reflected by the beam splitter, representing the loss of photons (or energy) in the channel. When $\lambda = 1$, the pure-loss channel is noiseless (it equals the identity superoperator). Conversely, when $\lambda = 0$, the pure-loss channel is completely noisy (specifically, it is a constant channel that maps any state in $|0\rangle\langle 0|$).

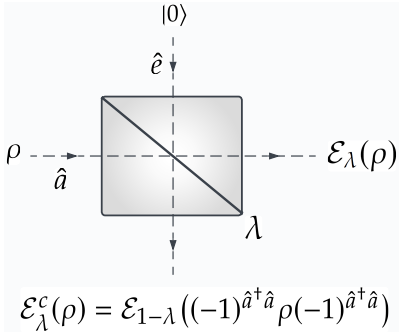


Figure 6.2: Stinespring representation of the pure-loss channel $\mathcal{E}_\lambda$. The pure-loss channel $\mathcal{E}_\lambda$ acts on the input state ρ by mixing it in a beam splitter of transmissivity λ (represented by the grey box) with an environmental vacuum state $|0\rangle$. Moreover, $\hat{a}$ and $\hat{e}$ are the annihilation operators of the input mode and the environmental mode, respectively. The complementary channel of the pure-loss channel is given by $\mathcal{E}^{\mathrm{c}}_\lambda(\rho) = \mathcal{E}_{1-\lambda}\Big((-1)^{\hat{a}^\dagger\hat{a}}\rho(-1)^{\hat{a}^\dagger\hat{a}}\Big)$.

Definition 37. *Let $\mathcal{H}_S, \mathcal{H}_E := L^2(\mathbb{R})$. For all $\lambda \in [0,1]$, the pure-loss channel of transmissivity λ is a quantum channel $\mathcal{E}_\lambda : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)$ defined as follows:*

$$\mathcal{E}_\lambda(\rho) := \mathrm{Tr}_E\left[U^{SE}_\lambda\Big(\rho_S \otimes |0\rangle\langle 0|_E\Big)(U^{SE}_\lambda)^\dagger\right], \quad \forall\, \rho \in \mathcal{T}(\mathcal{H}_S)\,,$$

where $|0\rangle\langle 0|_E$ denotes the vacuum state of $\mathcal{H}_E$ and U^{SE}_λ denotes the beam splitter unitary defined in (6.17).

Lemma 38. *For all $\lambda \in [0,1]$ and all $n, m \in \mathbb{N}$, it holds that*

$$\mathcal{E}_\lambda(|m\rangle\langle m|\, n) = \sum_{\ell=0}^{\min(n,m)} \sqrt{\binom{m}{\ell}\binom{n}{\ell}}\,(1-\lambda)^\ell \lambda^{\frac{n+m}{2}-\ell}\,|m-\ell\rangle\langle m-\ell|\, n-\ell\,.$$

Proof. This is a direct consequence of (6.19) and of the definition of pure-loss channel. □

Lemma 39. *For all $\lambda \in [0,1]$, a complementary channel $\mathcal{E}^c_\lambda : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_E)$ of the pure-loss channel $\mathcal{E}_\lambda : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)$ is given by*

$$\mathcal{E}^c_\lambda(\rho) := \mathrm{Tr}_S\left[U^{SE}_\lambda\Big(\rho_S \otimes |0\rangle\langle 0|_E\Big)\left(U^{SE}_\lambda\right)^\dagger\right] = \mathcal{E}_{1-\lambda} \circ \mathcal{R}(\rho), \qquad \forall\, \rho \in \mathcal{T}(\mathcal{H}_S)\,, \tag{6.22}$$

where $\mathcal{R}(\cdot) := V \cdot V^\dagger$ with $V := (-1)^{\hat{a}^\dagger\hat{a}}$.

Proof. By linearity, it suffices to show the identity in (6.22) for rank-one Fock operators of the form $|m\rangle\langle m|\, n$ for any $n, m \in \mathbb{N}$, i.e.

$$\mathrm{Tr}_S\left[U^{SE}_\lambda\Big(\,|m\rangle\langle m|\, n_S \otimes |0\rangle\langle 0|_E\Big)\left(U^{SE}_\lambda\right)^\dagger\right] = \mathcal{E}_{1-\lambda} \circ \mathcal{R}(|m\rangle\langle m|\, n)\,. \tag{6.23}$$

This follows directly from (6.19). □

Proposition 40. *[38, 39] The pure-loss channel $\mathcal{E}_\lambda$ is degradable if and only if $\lambda \in [\frac{1}{2}, 1]$, and it is anti-degradable if and only if $\lambda \in [0, \frac{1}{2}]$.*

Lemma 41. *[22, Lemma A.7] For all $\lambda_1, \lambda_2 \in [0,1]$ it holds that $\mathcal{E}_{\lambda_1} \circ \mathcal{E}_{\lambda_2} = \mathcal{E}_{\lambda_1 \lambda_2}$.*

Bosonic dephasing channel

Another main source of noise in optical platforms is bosonic dephasing, which serves as a prominent example of a non-Gaussian channel [29, 40].

Definition 42. *Let $\mathcal{H}_S := L^2(\mathbb{R})$ and let $\hat{a}$ be the corresponding annihilation operator. For all $\gamma \geq 0$, the bosonic dephasing channel $\mathcal{D}_\gamma : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)$ is a quantum channel defined as follows:*

$$\mathcal{D}_\gamma(\rho) := \frac{1}{\sqrt{2\pi\gamma}} \int_{-\infty}^{\infty} \mathrm{d}\phi \, e^{-\frac{\phi^2}{2\gamma}} \, e^{i\phi \hat{a}^\dagger \hat{a}} \, \rho \, e^{-i\phi \hat{a}^\dagger \hat{a}}, \quad \forall \, \rho \in \mathcal{T}(\mathcal{H}_S) \, .$$

In other words, $\mathcal{D}_\gamma$ is a convex combination of phase space rotations $\rho \longmapsto e^{i\phi \hat{a}^\dagger \hat{a}} \, \rho \, e^{-i\phi \hat{a}^\dagger \hat{a}}$, where the random variable ϕ follows a centered Gaussian distribution with a variance of γ.

Lemma 43. *For all $\gamma \geq 0$ and all $n, m \in \mathbb{N}$, it holds that*

$$\mathcal{D}_\gamma(|m\rangle\langle m| \, n) = e^{-\frac{1}{2}\gamma(n-m)^2} \, |m\rangle\langle m| \, n \, .$$

Proof. This result follows from the Fourier transform of the Gaussian function:

$$\frac{1}{\sqrt{2\pi\gamma}} \int_{-\infty}^{\infty} \mathrm{d}\phi \, e^{-\frac{\phi^2}{2\gamma}} e^{i\phi k} = e^{-\frac{1}{2}\gamma k^2}, \quad \forall \, k \in \mathbb{R} \, . \tag{6.24}$$

□

When $\gamma = 0$, the bosonic dephasing channel is noiseless. In contrast, when $\gamma \to \infty$ it annihilates all off-diagonal components of the input density matrix, reducing it to an incoherent probabilistic mixture of Fock states. We now present a Stinespring dilation of the bosonic dephasing channel.

Lemma 44. *Let $\mathcal{H}_S = \mathcal{H}_E := L^2(\mathbb{R})$ and let $\hat{a}$ and $\hat{e}$ be annihilation operators on $\mathcal{H}_S$ and $\mathcal{H}_E$, respectively. For all $\gamma > 0$, the bosonic dephasing channel $\mathcal{D}_\gamma : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)$ can be expressed in Stinespring representation as*

$$\mathcal{D}_\gamma(\rho) = \mathrm{Tr}_E \left[V_\gamma^{SE} \Big(\rho_S \otimes |0\rangle\langle 0|_E \Big) (V_\gamma^{SE})^\dagger \right], \qquad \forall \, \rho \in \mathcal{T}(\mathcal{H}_S) \, , \tag{6.25}$$

where V_γ^{SE} denotes the conditional displacement unitary defined by

$$V_\gamma^{SE} := \exp\left[\sqrt{\gamma} \, \hat{a}^\dagger \hat{a} \otimes \, (\hat{e}^\dagger - \hat{e}) \right] = \sum_{n=0}^{\infty} |n\rangle\langle n|_S \otimes D(\sqrt{\gamma} n). \tag{6.26}$$

Proof. For any $n, m \in \mathbb{N}$ it holds that

$$\begin{aligned} & \mathrm{Tr}_E \left[V_\gamma^{SE} \Big(\, |m\rangle\langle m| \, n_S \otimes |0\rangle\langle 0|_E \Big) (V_\gamma^{SE})^\dagger \right] \\ & \overset{\text{(i)}}{=} \mathrm{Tr}_E \left[|m\rangle\langle m| \, n_S \otimes |\sqrt{\gamma} n\rangle\langle \sqrt{\gamma} n| \, \sqrt{\gamma} m_E \right] \\ & \overset{\text{(ii)}}{=} e^{-\frac{\gamma}{2}(n-m)^2} \, |m\rangle\langle m| \, n \\ & \overset{\text{(iii)}}{=} \mathcal{D}_\gamma(|m\rangle\langle m| \, n) \, . \end{aligned} \tag{6.27}$$

Here, in (i) we used that $V_\gamma^{SE} \, |n\rangle_S \otimes |0\rangle_E = |n\rangle_S \otimes |\sqrt{\gamma} n\rangle_E$, where $|\sqrt{\gamma} n\rangle_E$ denotes the coherent state with parameter $\sqrt{\gamma} n$. In (ii), we exploited the formula for the overlap between coherent states provided in (6.15), and in (iii) we used Lemma 43. The proof is completed by linearity. □

Remark 45. *A different approach to dilating the bosonic dephasing channel, as outlined in the existing literature [40, 64, 65], is as follows:*

$$\tilde{V}_\gamma^{SE} = \exp\left[-i\sqrt{\gamma}\,\hat{a}^\dagger\hat{a}\,(\hat{e}^\dagger + \hat{e})\right].$$

This unitary transformation is achieved by rotating the environmental mode of the unitary operator V_γ^{SE} in (6.26) by $\frac{\pi}{2}$, that is $\tilde{V}_\gamma^{SE} = e^{i\frac{\pi}{2}\hat{e}^\dagger\hat{e}} V_\gamma^{SE} e^{-i\frac{\pi}{2}\hat{e}^\dagger\hat{e}}$. These dilations yield the same dephasing channel, as all dilations are equivalent up to unitary transformations.

Proposition 46 [1]**.** *The bosonic dephasing channel $\mathcal{D}_\gamma$ is degradable for all $\gamma \geq 0$.*

Proposition 47 [1]**.** *The bosonic dephasing channel $\mathcal{D}_\gamma$ is never anti-degradable.*

Lemma 48. *For all $\gamma_1, \gamma_2 \geq 0$, the composition of bosonic dephasing channels is given by $\mathcal{D}_{\gamma_1} \circ \mathcal{D}_{\gamma_2} = \mathcal{D}_{\gamma_1+\gamma_2}$.*

Proof. This can be shown by leveraging Lemma 43. □

Bosonic loss-dephasing channel

Consider an optical system undergoing simultaneous loss and dephasing over a finite time interval. At each instant, the system is susceptible to both an infinitesimal pure-loss channel and an infinitesimal bosonic dephasing channel. Hence, the overall channel describing the simultaneous effect of loss and dephasing results in a suitable composition of numerous concatenations between infinitesimal pure-loss and bosonic dephasing channels. However, given that:

- The pure-loss channel and the bosonic dephasing channel commute, i.e. $\mathcal{E}_\lambda \circ \mathcal{D}_\gamma = \mathcal{D}_\gamma \circ \mathcal{E}_\lambda$, as implied by Lemma 38 and Lemma 43;
- The composition of pure-loss channels is a pure-loss channel, $\mathcal{E}_{\lambda_1} \circ \mathcal{E}_{\lambda_2} = \mathcal{E}_{\lambda_1\lambda_2}$ (Lemma 41);
- The composition of bosonic dephasing channels is a bosonic dephasing channel, $\mathcal{D}_{\gamma_1} \circ \mathcal{D}_{\gamma_2} = \mathcal{D}_{\gamma_1+\gamma_2}$ (Lemma 48);

it follows that the combined effect of loss and dephasing can be modelled by the composition between pure-loss channel and bosonic dephasing channel,

$$\mathcal{N}_{\lambda,\gamma} := \mathcal{E}_\lambda \circ \mathcal{D}_\gamma\,, \tag{6.28}$$

dubbed the *bosonic loss-dephasing channel*.

Definition 49. *For all $\gamma \geq 0$ and $\lambda \in [0,1]$, the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is a quantum channel defined by the composition between pure-loss channel and bosonic dephasing channel: $\mathcal{N}_{\lambda,\gamma} := \mathcal{D}_\gamma \circ \mathcal{E}_\lambda$.*

Lemma 50. *For all $\lambda \in [0,1]$ and $\gamma \geq 0$, it holds that $\mathcal{N}_{\lambda,\gamma} := \mathcal{D}_\gamma \circ \mathcal{E}_\lambda = \mathcal{E}_\lambda \circ \mathcal{D}_\gamma$. Moreover, for all $n, m \in \mathbb{N}$ it holds that*

$$\begin{aligned}\mathcal{N}_{\lambda,\gamma}(|m\rangle\langle m|\,n) &= e^{-\frac{1}{2}\gamma(n-m)^2}\mathcal{E}_\lambda(|m\rangle\langle m|\,n) \\ &= e^{-\frac{1}{2}\gamma(n-m)^2}\sum_{\ell=0}^{\min(n,m)}\sqrt{\binom{n}{\ell}\binom{m}{\ell}}\,(1-\lambda)^\ell\lambda^{\frac{n+m}{2}-\ell}\,|m-\ell\rangle\langle m-\ell|\,n-\ell\,.\end{aligned}$$

Proof. It follows from Lemma 38 and Lemma 43. □

Lemma 51. *Let $\mathcal{H}_S, \mathcal{H}_{E_1}, \mathcal{H}_{E_2} := L^2(\mathbb{R})$. For all $\lambda \in [0,1]$ and all $\gamma \geq 0$, the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma} : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)$ admits the following Stinespring representation:*

$$\mathcal{N}_{\lambda,\gamma}(\rho) = \mathrm{Tr}_{E_1E_2}\left[U_\lambda^{SE_1}V_\gamma^{SE_2}\left(\rho_S \otimes |0\rangle\langle 0|_{E_1} \otimes |0\rangle\langle 0|_{E_2}\right)\left(U_\lambda^{SE_1}V_\gamma^{SE_2}\right)^\dagger\right], \qquad \forall\, \rho \in \mathcal{T}(\mathcal{H}_S),$$

The associated complementary channel $\mathcal{N}^c_{\lambda,\gamma} : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_{E_1} \otimes \mathcal{H}_{E_2})$ is defined as follows:

$$\mathcal{N}^c_{\lambda,\gamma}(\rho) := \mathrm{Tr}_S\left[U_\lambda^{SE_1}V_\gamma^{SE_2}\left(\rho_S \otimes |0\rangle\langle 0|_{E_1} \otimes |0\rangle\langle 0|_{E_2}\right)\left(U_\lambda^{SE_1}V_\gamma^{SE_2}\right)^\dagger\right], \qquad \forall\, \rho \in \mathcal{T}(\mathcal{H}_S)\,.$$

In particular,

$$\mathcal{N}^c_{\lambda,\gamma}(|m\rangle\langle m|\, n) = (-1)^{m-n}\mathcal{E}_{1-\lambda}\left(|m\rangle\langle m|\, n_{E_1}\right) \otimes |\sqrt{\gamma}m\rangle\langle\sqrt{\gamma}m|\, \sqrt{\gamma}n_{E_2}, \qquad \forall\, m,n \in \mathbb{N}\,, \tag{6.29}$$

where $|\sqrt{\gamma}n\rangle_{E_2}$ denotes the coherent state with parameter $\sqrt{\gamma}n$.

Proof. The Eq. (6.29) is derived by first applying (6.22) and subsequently utilising the dilation of the bosonic dephasing channel. Finally, the non-environment mode of the bosonic dephasing channel is traced out. □

Lemma 52. *Consider the Hilbert spaces $\mathcal{H}_S$, $\mathcal{H}_{E_1}$, and $\mathcal{H}_{E_2}$, all isomorphic to $L^2(\mathbb{R})$. Let $\lambda \in [0,1]$ and $\gamma \geq 0$. The bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma} : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)$ exhibits anti-degradability if and only if there exists a quantum channel $\mathcal{A}_{\lambda,\gamma} : \mathcal{T}(\mathcal{H}_{E_{out}}) \to \mathcal{T}(\mathcal{H}_S)$ satisfying the following condition:*

$$\mathcal{A}_{\lambda,\gamma} \circ \mathcal{N}^c_{\lambda,\gamma}(|m\rangle\langle m|\, n) = \mathcal{N}_{\lambda,\gamma}(|m\rangle\langle m|\, n), \quad \forall\, m,n \in \mathbb{N}\,, \tag{6.30}$$

where $\mathcal{N}^c_{\lambda,\gamma} : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_{E_{out}})$ denotes the complementary channel reported in (6.29), and $\mathcal{H}_{E_{out}} \subset \mathcal{H}_{E_1} \otimes \mathcal{H}_{E_2}$ is defined in (6.51). Moreover, the condition in (6.30) is equivalent to

$$\mathcal{A}_{\lambda,\gamma}\left(\mathcal{E}_{1-\lambda}\left(|m\rangle\langle m|\, n\right) \otimes |\sqrt{\gamma}m\rangle\langle\sqrt{\gamma}m|\, \sqrt{\gamma}n\right) = (-1)^{m-n}e^{-\frac{\gamma}{2}(m-n)^2}\mathcal{E}_\lambda(|m\rangle\langle m|\, n), \quad \forall\, m,n \in \mathbb{N}\,, \tag{6.31}$$

where $|\sqrt{\gamma}n\rangle$ denotes the coherent state with parameter $\sqrt{\gamma}n$.

Proof. $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable if and only if there exists a quantum channel $\mathcal{A}_{\lambda,\gamma} : \mathcal{T}(\mathcal{H}_{E_{\mathrm{out}}}) \to \mathcal{T}(\mathcal{H}_S)$ such that

$$\mathcal{A}_{\lambda,\gamma} \circ \mathcal{N}^{\mathrm{c}}_{\lambda,\gamma}(\rho) = \mathcal{N}_{\lambda,\gamma}(\rho), \qquad \forall\, \rho \in \mathcal{T}(\mathcal{H}_S)\,. \tag{6.32}$$

By linearity, it suffices to show the condition in (6.30), i.e. which corresponds to the condition in (6.32) restricted to rank-one Fock operators of the form $\rho = |m\rangle\langle m|\, n$ with $m,n \in \mathbb{N}$. Moreover, by exploiting Lemma 50 and 6.29, the condition in (6.30) is equivalent to (6.31). □

Lemma 53. *For all $\gamma_1, \gamma_2 \geq 0$ and all λ_1, λ_2 it holds that $\mathcal{N}_{\lambda_1,\gamma_1} \circ \mathcal{N}_{\lambda_2,\gamma_2} = \mathcal{N}_{\lambda_1\lambda_2\,,\,\gamma_1+\gamma_2}$.*

Proof. This can be shown by exploiting Lemma 41 and Lemma 48. □

Preliminaries on capacities of quantum channels

Quantum channels can be suitably exploited in order to transfer information from their input port to a possibly distant output port, a crucial task in quantum information theory [44, 59]. In particular, quantum Shannon theory [33, 34] primarily helps us understand the fundamental limits of quantum communication using a quantum channel $\mathcal{N}$. These limits are called *capacities* and tell us the ultimate amount of information we can send through the channel when we use it many times [33, 34]. Different capacities are defined based on the type of information that is being sent down the channel. For example, classical and quantum capacities of a quantum channel correspond to its ultimate capability of transmission of classical and quantum information, respectively. A channel can also be used to generate secret bits and the relevant capacity in this

context is the so-called secret-key capacity. Each of the above-mentioned capacities might be endowed with other resources such as initial shared entanglement between the sender and the receiver or (possibly interactive) classical communication over a noiseless but public channel. This latter scenario gives rise to the notion of two-way capacities.

Specifically, the *quantum capacity* $Q(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is the maximum rate at which qubits can be reliably transmitted through $\mathcal{N}$ [34]. We can further assume that both the sender, Alice, and the receiver, Bob, have free access to a public, noiseless two-way classical channel. In this two-way communication setting, the relevant notion of capacities are the *two-way quantum capacity* $Q_2(\mathcal{N})$ and the *secret-key capacity* $K(\mathcal{N})$ [33, 34], defined as the maximum achievable rate of qubits and secret-key bits, respectively, that can be reliably transmitted across $\mathcal{N}$ with the aid of two-way classical communication. Since an ebit (i.e. a maximally entangled state of Schmidt rank 2) can always be used to generate one bit of secret key [47], a trivial bound relates these capacities: $Q_2(\mathcal{N}) \leq K(\mathcal{N})$.

In practical scenarios, it is important to consider that the input state prepared by Alice can not have unlimited energy and it adheres to specific energy constraints. In bosonic systems, it is common to limit the average photon number of any input state ρ as $\mathrm{Tr}[\hat{a}^\dagger \hat{a} \rho] \leq N_s$, where $N_s > 0$ is a given energy constraint. For any $N_s > 0$, the energy-constrained (EC) two-way capacities for transmitting qubits and secret-key bits, denoted as $Q_2(\mathcal{N}, N_s)$ and $K(\mathcal{N}, N_s)$ respectively, are defined similarly to the unconstrained capacities with the difference that now the optimisation is performed over those strategies that exploit input states that adhere to the specified energy constraint. As in the unconstrained scenario, the relation between EC two-way capacities continue to hold, i.e. $Q_2(\mathcal{N}, N_s) \leq K(\mathcal{N}, N_s)$. Moreover, the unconstrained capacities are upper bounds for the corresponding energy constrained capacities and they become equal in the limit $N_s \to \infty$.

Two-way capacities of a quantum channel are closely related to another important information-processing task, namely, entanglement distillation over a quantum channel. Suppose Alice generates n copies of a state $\rho_{A'A}$ and sends the A' subsystems to Bob using the channel $\mathcal{N}$ for n times. Now, Alice and Bob share n copies of the state $\rho'_{AB} \coloneqq \mathrm{Id}_A \otimes \mathcal{N}_{A' \to B}(\rho_{AA'})$. The task of an entanglement distillation protocol concerns identifying the largest number m of ebits that can be extracted using n copies of ρ'_{AB} via LOCC (Local Operations and Classical Communication) operations. The rate of an entanglement distillation protocol is defined by the ratio m/n. The distillable entanglement $E_d(\rho'_{AB})$ of ρ'_{AB} is defined as the maximum rate over all the possible entanglement distillation protocols [66] [34, Chapter 8]. Note that without extra classical communication, entanglement distillation is not possible [67]. The following lemma establishes a link between the two-way quantum capacity, secret-key capacity, and distillable entanglement.

Lemma 54. *Let $\mathcal{H}_A, \mathcal{H}_{A'}, \mathcal{H}_B \coloneqq L^2(\mathbb{R})$. Let $\mathcal{N} : \mathcal{H}_{A'} \to \mathcal{H}_B$ be a quantum channel. Let $N_s > 0$ be the energy constaint, and let $\rho_{A'A} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$ be a two-mode state satisfying the energy constraint $\mathrm{Tr}[(\hat{a}^\dagger \hat{a} \otimes \mathrm{Id}_A)\rho_{A'A}] \leq N_s$, where $\hat{a}$ denotes the annihilation operator on $\mathcal{H}_{A'}$. Then, it holds that*

$$K(\mathcal{N}, N_s) \geq Q_2(\mathcal{N}, N_s) \geq E_d(\mathrm{Id}_A \otimes \mathcal{N}(\rho_{AA'})) \ , \tag{6.33}$$

where $K(\mathcal{N}, N_s)$ denotes the energy-constrained secret-key capacity of $\mathcal{N}$, $Q_2(\mathcal{N}, N_s)$ denotes the energy-constrained two-way quantum capacity of $\mathcal{N}$, and $E_d(\mathrm{Id}_A \otimes \mathcal{N}(\rho_{AA'}))$ denotes the distillable entanglement of the state $\mathrm{Id}_A \otimes \mathcal{N}(\rho_{AA'})$.

The proof idea of the above lemma is the following. Suppose that Alice produces n copies of a state $\rho_{AA'}$ such that the energy constraint is satisfied. Then, she can use the channel n times to send all subsystems A' to Bob. Then, Alice and Bob share n copies of $\mathrm{Id}_A \otimes \mathcal{N}(\rho_{AA'})$, which can now be used to generate $\approx n\, E_d(\mathrm{Id}_A \otimes \mathcal{N}(\rho_{AA'}))$ ebits by means of a suitable entanglement distillation protocol. The ebit rate of this protocol is thus $E_d(\mathrm{Id}_A \otimes \mathcal{N}(\rho_{AA'}))$, which provides a lower bound on $Q_2(\mathcal{N}, N_s)$ thanks to quantum teleportation [68]. In addition, it holds that $K(\mathcal{N}, N_s) \geq Q_2(\mathcal{N}, N_s)$ because an ebit can generate a secret-key bit [47]. Consequently, (6.33) holds.

6.7.2 Anti-degradability and Degradability of bosonic loss-dephasing channel

This section is split into two parts based on the observation that if the input state to the bosonic loss-dephasing channel is chosen from a finite-dimensional subspace, the bosonic loss-dephasing channel effectively becomes a finite-dimensional channel, a fact we show in Lemma 58. This property allows us to apply established insights about the finite-dimensional channels to the bosonic loss-dephasing channel. In subsection 6.7.2, we present our study of the bosonic loss-dephasing channel when the input resides in the entire infinite-dimensional space, while subsection 6.7.2 is dedicated to the findings resulting from analysis of finite-dimensional restrictions of the bosonic loss-dephasing channel.

Sufficient condition on anti-degradability

It is known that the bosonic dephasing channel $\mathcal{D}_\gamma$ is degradable across all dephasing parameter range $\gamma \geq 0$ and also it is never anti-degradable [40]. The pure-loss channel $\mathcal{E}_\lambda$ displays the peculiar characteristic of being anti-degradable for transmissivity values within the range $\lambda \in [0, \frac{1}{2}]$ and degradable for $\lambda \in [\frac{1}{2}, 1]$ [38, 39]. It turns out that when an anti-degradable channel is concatenated with another channel, the resulting channel inherits the property of being anti-degradable (see Lemma 71). This implies the following: if $\lambda \in [0, \frac{1}{2}]$ and $\gamma \geq 0$, then the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable [1]. The authors of [1] left as an open question to understand whether or not the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable in the region $\lambda \in (\frac{1}{2}, 1]$. In particular, they conjecture that $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable for all transmissivity values $\lambda \in (\frac{1}{2}, 1]$ and for all $\gamma \geq 0$. In the following theorem, we refute this conjecture by explicitly finding values of $\lambda \in (\frac{1}{2}, 1]$ and $\gamma \geq 0$ where the channel is anti-degradable. Our approach also yields an explicit expression for an anti-degrading map of the bosonic loss-dephasing channel.

Theorem 4. Each of the following is a sufficient condition for the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ to exhibit anti-degradability:

(i) $\lambda \in [0, \frac{1}{2}]$ and $\gamma \geq 0$.

(ii) $\lambda \in (\frac{1}{2}, 1)$ and $\theta\Big(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\Big) \leq \frac{3}{2}$, where θ is defined as $\theta(x, y) \coloneqq \sum_{n=0}^{\infty} x^{n^2} y^n$, $\forall\, x, y \in [0, 1)$.

In particular, $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable if $\lambda \leq \max\Big(\frac{1}{2}, \frac{1}{1+9e^{-\gamma}}\Big)$.

Proof. The proof of the sufficient condition (i) follows directly from the observation that the composition of a pure-loss channel with transmissivity $\lambda \in [0, \frac{1}{2}]$ with any other channel inherits the anti-degradability from the pure-loss channel (see Lemma 71).

The proof of the sufficient condition (ii) is more involved, and it is the main technical contribution of our work. We rely on the equivalence between anti-degradability of a quantum channel and the two-extendibility of its Choi state [41, 42]. To provide a comprehensive and intuitive understanding of this idea and to aid in the construction of an anti-degrading map of the bosonic loss-dephasing channel, we present this equivalence in Lemma 70 within the Appendix.

Assume $\lambda \in (\frac{1}{2}, 1)$ and $\theta\Big(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\Big) \leq \frac{3}{2}$. Let $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_{B_1}, \mathcal{H}_{B_2} = L^2(\mathbb{R})$ and suppose that the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is a quantum channel from the system A' to B. We want to show that the generalised Choi state of $\mathcal{N}_{\lambda,\gamma}$ is two-extendible on B. In other words, we want to show that there exists a tripartite state $\rho_{AB_1B_2}$ such that

$$\begin{aligned} \mathrm{Tr}_{B_2}\left[\rho_{AB_1B_2}\right] &= \tau_{AB_1}\,, \\ \mathrm{Tr}_{B_1}\left[\rho_{AB_1B_2}\right] &= \tau_{AB_2}\,, \end{aligned} \tag{6.34}$$

where $\tau_{AB} \coloneqq \mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}(|\psi(r)\rangle\langle\psi(r)|_{AA'})$ is the generalised Choi state of $\mathcal{N}_{\lambda,\gamma}$, and $|\psi(r)\rangle_{AA'}$ is the two-mode squeezed vacuum state with squeezing parameter $r > 0$ defined in (6.16). By

Lemma (50), the generalised Choi state of the bosonic loss-dephasing channel can be expressed as follows:

$$\tau_{AB} = \frac{1}{\cosh^2(r)} \sum_{m,n=0}^{\infty} \sum_{\ell=0}^{\min(m,n)} (\tanh(r))^{m+n} e^{-\frac{\gamma}{2}(m-n)^2} \sqrt{\mathcal{B}_\ell(m, 1-\lambda)\, \mathcal{B}_\ell(n, 1-\lambda)}\, |m\rangle\langle m|\, n_A \otimes |m-\ell\rangle\langle m-\ell|\, n-\ell_B\,, \tag{6.35}$$

where $\mathcal{B}_\ell(n,\lambda) \coloneqq \binom{n}{\ell}\lambda^\ell(1-\lambda)^{n-\ell}$. We observe that τ_{AB} is a linear combination of $|m\rangle\langle m|\, n \otimes |j_1\rangle\langle j_1|\, j_2$, where $m, n, j_1, j_2 \in \mathbb{N}$, $j_1 \leq n_1$, $j_2 \leq n$, and $m - n = j_1 - j_2$. This insight leads to the following educated guess about the structure of a potential two-extension:

$$\tilde{\rho}_{AB_1B_2} = \sum_{m,n=0}^{\infty} \sum_{\ell_1=0}^{m} \sum_{\ell_2=0}^{n} \sum_{k=0}^{\min(m-\ell_1, n-\ell_2)} c(m,n,\ell_1,\ell_2,k)\, |m\rangle\langle m|\, n_A \otimes |m-\ell_1\rangle\langle m-\ell_1|\, n-\ell_{2B_1} \otimes |\ell_1+k\rangle\langle \ell_1+k|\, \ell_2+k_{B_2}\,, \tag{6.36}$$

where $\{c(m,n,\ell_1,\ell_2,k)\}_{m,n,\ell_1,\ell_2,k}$ are some suitable coefficients. The soundness of this guess is confirmed by the fact that both $\mathrm{Tr}_{B_2}[\tilde{\rho}_{AB_1B_2}]$ and $\mathrm{Tr}_{B_1}[\tilde{\rho}_{AB_1B_2}]$ are linear combinations of $|m\rangle\langle m|\, n \otimes |j_1\rangle\langle j_1|\, j_2$ with $m, n, j_1, j_2 \in \mathbb{N}$, $j_1 \leq m$, $j_2 \leq n$, and $m - n = j_1 - j_2$, similar to the generalised Choi state τ_{AB} in (6.35).

At this point, one could try to define the coefficients $\{c(m,n,\ell_1,\ell_2,k)\}_{m,n,\ell_1,\ell_2,k}$ so as to satisfy the required conditions of the extendibility. However, the resulting tripartite operator may not qualify as a quantum state. In order to ensure that we obtain a quantum state, our approach consists in producing the operator $\tilde{\rho}_{AB_1B_2}$ via a physical process that consists in applying a sequence of quantum channels to a quantum state.

We begin by constructing a quantum state that has the same operator structure as the operator $\tilde{\rho}_{AB_1B_2}$ in (6.36). This means that at this initial stage, we only aim to build a tripartite state consisting of a linear combination of operators $|m\rangle\langle m|\, n_A \otimes |m-\ell_1\rangle\langle m-\ell_1|\, n-\ell_{2B_1} \otimes |\ell_1+k\rangle\langle \ell_1+k|\, \ell_2+k_{B_2}$ with the summation limits identical to those in (6.36). This ensures that the coefficients equal to zero coincide in the two operators.

We now illustrate on each step of this construction. For a fixed $n \in \mathbb{N}$, consider the state $|n\rangle_A \otimes |n\rangle_{B_1}$. We introduce two auxiliary single-mode systems B_2 and C initially in vacuum states: $|n\rangle_A \otimes |n\rangle_{B_1} \otimes |0\rangle_{B_2} \otimes |0\rangle_C$. We next send the systems B_2 and B_1 through the ports of a beam splitter, resulting in a superposition of $|n\rangle_A \otimes |n-\ell\rangle_{B_1} \otimes |\ell\rangle_{B_2} \otimes |0\rangle_C$ for $\ell = 0, 1, \ldots, n$, as implied by Lemma (35). We repeat this for systems B_1 and C, thus obtaining a superposition of $|n\rangle_A \otimes |n-\ell-k\rangle_{B_1} \otimes |\ell\rangle_{B_2} \otimes |k\rangle_C$, with $k = 0, 1, \ldots, n-\ell$ and $\ell = 0, 1, \ldots, n$. Consider now the isometry $W^{CB_1B_2}$ defined by

$$W^{CB_1B_2} |n\rangle_{B_1} \otimes |m\rangle_{B_2} \otimes |k\rangle_C = |n+k\rangle_{B_1} \otimes |m+k\rangle_{B_2} \otimes |k\rangle_C\,, \quad \forall\, n, m, k \in \mathbb{N}\,, \tag{6.37}$$

dubbed *controlled-add-add isometry* (mode C is the control mode). By applying this isometry to the superposition we created by using beam splitters, we obtain a superposition of $|n\rangle_A \otimes |n-\ell\rangle_{B_1} \otimes |\ell+k\rangle_{B_2} \otimes |k\rangle_C$ with $k = 0, 1, \ldots, n-\ell$ and $\ell = 0, 1, \ldots, n$. Finally, by tracing out system C, we obtain the same operator structure of the operator $\tilde{\rho}_{AB_1B_2}$ in (6.36). Having focused on the operator structure, we have not considered the transmissivities of the two beam splitters so far. We will see that these transmissivities can be chosen carefully such that the diagonal elements of the operator at hand becomes equal to those of the Choi state.

We now apply the outlined construction to the two-mode squeezed vacuum state with two vacuum states appended to it, i.e. $|\psi(r)\rangle_{AB_1} \otimes |0\rangle_{B_2} \otimes |0\rangle_C$. For reasons that will become clear in a moment, we choose the two beam splitter transmissivities to be λ (for the beam splitter acting on B_2B_1) and $\frac{1-\lambda}{\lambda}$ (for the one acting on CB_1). By doing so we obtain the state

$$\begin{aligned} |\phi\rangle_{AB_1B_2C} &\coloneqq W^{C,B_1B_2} U^{CB_1}_{\frac{1-\lambda}{\lambda}} U^{B_2B_1}_{\lambda} |\psi(r)\rangle_{AB_1} \otimes |0\rangle_{B_2} \otimes |0\rangle_C \\ &= \frac{1}{\cosh(r)} \sum_{n=0}^{\infty} \sum_{\ell=0}^{n} \sum_{k=0}^{n-\ell} \tanh^n(r) \sqrt{\mathcal{B}_\ell(n, 1-\lambda)\, \mathcal{B}_k\left(n-\ell, \frac{2\lambda-1}{\lambda}\right)}\, |n\rangle_A \otimes |n-\ell\rangle_{B_1} \otimes |\ell+k\rangle_{B_2} \otimes |k\rangle_C\,, \end{aligned} \tag{6.38}$$

where we used (6.16) and Lemma 35. Note that the transmissivities of the beam splitters $U^{CB_1}_{\frac{1-\lambda}{\lambda}}$ and $U^{B_2B_1}_{\lambda}$ are chosen such that the diagonal elements of $\mathrm{Tr}_{B_2C}\big[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\big]$ and $\mathrm{Tr}_{B_1C}\big[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\big]$ both coincide with those of the Choi state τ_{AB} in (6.35). To verify this, let us calculate the state $\mathrm{Tr}_C\big[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\big]$:

$$\mathrm{Tr}_C\left[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\right] = \frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell_1=0}^{m}\sum_{\ell_2=0}^{n}\sum_{k=0}^{\min(m-\ell_1,n-\ell_2)}(\tanh(r))^{m+n}\sqrt{\mathcal{B}_{\ell_1}(m,1-\lambda)\,\mathcal{B}_{\ell_2}(n,1-\lambda)}$$
$$\left[\sqrt{\mathcal{B}_k\Big(m-\ell_1,\frac{2\lambda-1}{\lambda}\Big)\,\mathcal{B}_k\Big(n-\ell_2,\frac{2\lambda-1}{\lambda}\Big)}\right]|m\rangle\langle m|\,n_A\otimes|m-\ell_1\rangle\langle m-\ell_1|\,n-\ell_{2B_1}\otimes|\ell_1+k\rangle\langle\ell_1+k|\,\ell_2+k_{B_2}.$$

Notably, the structure of the state $\mathrm{Tr}_C\big[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\big]$ mirrors that of (6.36) with specific coefficients $c(m,n,\ell_1,\ell_2,k)$. Moreover, it holds that

$$\mathrm{Tr}_{B_2C}[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,] = \frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell=0}^{\min(m,n)}(\tanh(r))^{m+n}\sqrt{\mathcal{B}_\ell(m,1-\lambda)\,\mathcal{B}_\ell(n,1-\lambda)}$$
$$\left[\sum_{k=0}^{\min(m-\ell,n-\ell)}\sqrt{\mathcal{B}_k\Big(m-\ell,\frac{2\lambda-1}{\lambda}\Big)\,\mathcal{B}_k\Big(n-\ell,\frac{2\lambda-1}{\lambda}\Big)}\right]|m\rangle\langle m|\,n_A\otimes|m-\ell\rangle\langle m-\ell|\,n-\ell_{B_1} \tag{6.39}$$

and that

$$\mathrm{Tr}_{B_1C}\big[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\big] = \mathrm{Tr}_{B_2C}\big[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\big]. \tag{6.40}$$

In order to prove (6.40), observe that

$$\begin{aligned}
\mathrm{Tr}_{B_1C}[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,] &= \frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell_1=\max(m-n,0)}^{m}\sum_{k=0}^{m-\ell_1}(\tanh(r))^{m+n}\sqrt{\mathcal{B}_{\ell_1}(m,1-\lambda)\,\mathcal{B}_{n-m+\ell_1}(n,1-\lambda)}\\
&\quad \mathcal{B}_k\Big(m-\ell_1,\frac{2\lambda-1}{\lambda}\Big)\,|m\rangle\langle m|\,n_A\otimes|\ell_1+k\rangle\langle\ell_1+k|\,n-m+\ell_1+k_{B_2}\\
&= \frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell_1=\max(m-n,0)}^{m}\sum_{\ell=0}^{m-\ell_1}(\tanh(r))^{m+n}\sqrt{\mathcal{B}_{\ell_1}(m,1-\lambda)\,\mathcal{B}_{n-m+\ell_1}(n,1-\lambda)}\\
&\quad \mathcal{B}_{m-\ell_1-\ell}\Big(m-\ell_1,\frac{2\lambda-1}{\lambda}\Big)\,|m\rangle\langle m|\,n_A\otimes|m-\ell\rangle\langle m-\ell|\,n-\ell_{B_2}\\
&= \frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell=0}^{\min(m,n)}(\tanh(r))^{m+n}\sum_{\ell_1=\max(m-n,0)}^{m-\ell}\sqrt{\mathcal{B}_{\ell_1}(m,1-\lambda)\,\mathcal{B}_{n-m+\ell_1}(n,1-\lambda)}\\
&\quad \mathcal{B}_{m-\ell_1-\ell}\Big(m-\ell_1,\frac{2\lambda-1}{\lambda}\Big)\,|m\rangle\langle m|\,n_A\otimes|m-\ell\rangle\langle m-\ell|\,n-\ell\\
&= \frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell=0}^{\min(m,n)}(\tanh(r))^{m+n}\sum_{k=0}^{\min(n-\ell,m-\ell)}\sqrt{\mathcal{B}_{m-\ell-k}(m,1-\lambda)\,\mathcal{B}_{n-\ell-k}(n,1-\lambda)}\\
&\quad \mathcal{B}_k\Big(k+\ell,\frac{2\lambda-1}{\lambda}\Big)\,|m\rangle\langle m|\,n_A\otimes|m-\ell\rangle\langle m-\ell|\,n-\ell\\
&\overset{\text{(i)}}{=} \frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell=0}^{\min(m,n)}(\tanh(r))^{m+n}\sqrt{\mathcal{B}_\ell(m,1-\lambda)\,\mathcal{B}_\ell(n,1-\lambda)}\\
&\quad \left[\sum_{k=0}^{\min(m-\ell,n-\ell)}\sqrt{\mathcal{B}_k\Big(m-\ell,\frac{2\lambda-1}{\lambda}\Big)\,\mathcal{B}_k\Big(n-\ell,\frac{2\lambda-1}{\lambda}\Big)}\right]|m\rangle\langle m|\,n_A\otimes|m-\ell\rangle\langle m-\ell|\,n-\ell_{B_2}\\
&\overset{\text{(ii)}}{=} \mathrm{Tr}_{B_2C}[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,]
\end{aligned} \tag{6.41}$$

Here, in (i), we used the identity

$$\begin{aligned}&\sqrt{\mathcal{B}_{m-\ell-k}(m,1-\lambda)\,\mathcal{B}_{n-\ell-k}(n,1-\lambda)}\mathcal{B}_k\left(k+\ell,\frac{2\lambda-1}{\lambda}\right)\\&=\sqrt{\mathcal{B}_\ell(m,1-\lambda)\,\mathcal{B}_\ell(n,1-\lambda)}\sqrt{\mathcal{B}_k\left(m-\ell,\frac{2\lambda-1}{\lambda}\right)\,\mathcal{B}_k\left(n-\ell,\frac{2\lambda-1}{\lambda}\right)}\,,\end{aligned}\tag{6.42}$$

which can be easily proved by substituting the definition $\mathcal{B}_\ell(n,\lambda) \coloneqq \binom{n}{\ell}\lambda^\ell(1-\lambda)^{n-\ell}$ and by leveraging the binomial identity

$$\binom{n}{l+k}\binom{l+k}{k}=\binom{n}{l}\binom{n-l}{k}\,.\tag{6.43}$$

Moreover, in (ii), we exploited (6.39).

Note that the off-diagonal terms of the state in (6.39) are not equal to those of the Choi state τ_{AB} in (6.35). Specifically, the points of difference with the Choi state τ_{AB} are the presence of the term inside the square brackets and the absence of the dephasing exponent. To address these additional terms, let us use the toolbox of *Hadamard maps*. Let H be the Hadamard map, introduced in Sec. 6.7.1, associated with the infinite matrix $A \coloneqq (a_{mn})_{m,n\in\mathbb{N}}$ defined as follows:

$$a_{mn} \coloneqq \frac{e^{-\frac{\gamma}{2}(n-m)^2}}{\sum_{j=0}^{\min(n,m)}\sqrt{\mathcal{B}_j\left(n,\frac{2\lambda-1}{\lambda}\right)\,\mathcal{B}_j\left(m,\frac{2\lambda-1}{\lambda}\right)}}\,,\quad \forall\, n,m\in\mathbb{N}\,.\tag{6.44}$$

By construction, we have that

$$\begin{aligned}&\mathrm{Id}_A\otimes H_{B_1}\left(\mathrm{Tr}_{B_2C}\left[\,|\phi\rangle\langle\phi|_{AB_1B_2C}\,\right]\right)\\&=\frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell=0}^{\min(m,n)}(\tanh(r))^{m+n}\sqrt{\mathcal{B}_\ell(m,1-\lambda)\,\mathcal{B}_\ell(n,1-\lambda)}\\&\quad\left[\sum_{k=0}^{\min(m-\ell,n-\ell)}\sqrt{\mathcal{B}_k\left(m-\ell,\frac{2\lambda-1}{\lambda}\right)\,\mathcal{B}_k\left(n-\ell,\frac{2\lambda-1}{\lambda}\right)}\right]a_{m-\ell,\,n-\ell}\;|m\rangle\langle m|\,n_A\otimes|m-\ell\rangle\langle m-\ell|\,n-\ell_{B_1}\\&=\frac{1}{\cosh^2(r)}\sum_{m,n=0}^{\infty}\sum_{\ell=0}^{\min(m,n)}(\tanh(r))^{m+n}e^{-\frac{\gamma}{2}(n-m)^2}\sqrt{\mathcal{B}_\ell(m,1-\lambda)\,\mathcal{B}_\ell(n,1-\lambda)}\,|m\rangle\langle m|\,n_A\otimes|m-\ell\rangle\langle m-\ell|\,n-\ell_{B_1}\\&=\tau_{AB_1}\,.\end{aligned}\tag{6.45}$$

This means that the operator

$$\rho_{AB_1B_2}\coloneqq\mathrm{Id}_A\otimes H_{B_1}\otimes H_{B_2}\left(\mathrm{Tr}_C\left[|\phi\rangle\langle\phi|_{AB_1B_2C}\right]\right)\tag{6.46}$$

satisfies the extendibility conditions in (6.34). All that remains to prove is that $\rho_{AB_1B_2}$ is in fact a quantum state. We will do this by showing that the superoperator H is a quantum channel. In Sec. 6.7.1, we establish that a Hadamard map is a quantum channel if its defining infinite matrix is Hermitian, has diagonal elements equal to one, and is diagonally dominant. The first two properties are trivially satisfied by the infinite matrix A defined in (6.44). We only need to demonstrate that for the parameter region $\lambda > \frac{1}{2}$ and $\theta\left(e^{-\gamma/2},\sqrt{\frac{\lambda}{1-\lambda}}\right)\leq\frac{3}{2}$, the infinite matrix A is diagonally dominant. We recall that, by definition, A is diagonally dominant if it holds that

$\sum_{\substack{m=0\\m\neq n}}^{\infty} |a_{mn}| \leq 1$, $\forall\, n \in \mathbb{N}$. Note that for any $n, m \in \mathbb{N}$ we have that

$$\begin{aligned} |a_{nm}| &= \frac{e^{-\frac{\gamma}{2}(n-m)^2}}{\sum_{j=0}^{\min(n,m)} \sqrt{\binom{n}{j}\binom{m}{j}\left(\frac{1-\lambda}{\lambda}\right)^{n+m-2j}\left(\frac{2\lambda-1}{\lambda}\right)^{2j}}} \\ &\leq \frac{e^{-\frac{\gamma}{2}(n-m)^2}}{\sum_{j=0}^{\min(n,m)} \sqrt{\binom{\min(n,m)}{j}^2\left(\frac{1-\lambda}{\lambda}\right)^{n+m-2j}\left(\frac{2\lambda-1}{\lambda}\right)^{2j}}} \\ &= e^{-\frac{\gamma}{2}(n-m)^2}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{|n-m|}{2}} . \end{aligned} \tag{6.47}$$

Consequently, if λ and γ are such that $\lambda > \frac{1}{2}$ and $\theta\Big(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\Big) \leq \frac{3}{2}$, for any $n \in \mathbb{N}$ we have that

$$\begin{aligned} \sum_{\substack{m=0\\m\neq n}}^{\infty} |a_{mn}| &\leq \sum_{\substack{m=0\\m\neq n}}^{\infty} e^{-\frac{\gamma}{2}(m-n)^2}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{|m-n|}{2}} \\ &= \sum_{k=1}^{\infty} e^{-\frac{\gamma}{2}k^2}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k}{2}} + \sum_{k=1}^{n} e^{-\frac{\gamma}{2}k^2}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k}{2}} \\ &\leq 2\sum_{k=1}^{\infty} e^{-\frac{\gamma}{2}k^2}\left(\frac{\lambda}{1-\lambda}\right)^{\frac{k}{2}} \\ &= 2\,\theta\left(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\right) - 2 \\ &\leq 1\,. \end{aligned} \tag{6.48}$$

Therefore, the infinite matrix A is diagonally dominant if $\lambda > \frac{1}{2}$ and $\theta\Big(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\Big) \leq \frac{3}{2}$. This establishes that H_{B_1} and H_{B_2} are valid quantum channels in this parameter range, implying that $\rho_{AB_1B_2}$ is a valid two-extension of the Choi state of $\mathcal{N}_{\lambda,\gamma}$, and in turn entailing that $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable. Finally, note that if $\lambda > \frac{1}{2}$ the condition

$$\theta\left(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\right) \leq \frac{3}{2} \tag{6.49}$$

is implied by $\lambda \leq \frac{1}{1+9e^{-\gamma}}$. Indeed,

$$\theta\left(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\right) := \sum_{k=0}^{n} e^{-\frac{\gamma}{2}k^2}\left(\sqrt{\frac{\lambda}{1-\lambda}}\right)^k \leq \sum_{k=0}^{\infty}\left(\frac{e^{-\gamma}\lambda}{1-\lambda}\right)^{k/2} = \frac{1}{1-\sqrt{\frac{e^{-\gamma}\lambda}{1-\lambda}}} \leq \frac{3}{2}\,,$$

where the last inequality follows from $\frac{e^{-\gamma}\lambda}{1-\lambda} \leq \frac{1}{9}$, which is implied by $\lambda \leq \frac{1}{1+9e^{-\gamma}}$. □

Expanding the anti-degradability region numerically

Note that Theorem 4 does not identify the entire anti-degradability region of the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$. In fact, from the above argument it becomes clear that a way to obtain a better inner approximation of this region is to check for which values of the parameters the infinite matrix A defined by (6.44) is positive semi-definite. This is established in the following theorem.

Theorem 5. Let $\ell^2(\mathbb{N})$ be the space of square-summable complex-valued sequences (defined by (6.97) below). For any $\lambda \in (\frac{1}{2}, 1)$ and $\gamma > 0$, let $A = (a_{mn})_{m,n\in\mathbb{N}}$ be the infinite matrix whose components are defined by (6.44). If $A \geq 0$ is positive semi-definite as an operator on $\ell^2(\mathbb{N})$, then the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable.

Proof. In the proof of Theorem 4 we have seen that the bosonic loss-dephasing channel is anti-degradable if the Hadamard map associated with the infinite matrix A (given in (6.44)) is a quantum channel. Since the diagonal elements of A are equal to one, from Lemma 66 we deduce that the Hadamard map associated to A is a quantum channel if and only if A is positive semi-definite. This concludes the proof. □

In Theorem 4, we showed that the above-mentioned infinite matrix A is positive semi-definite if $\theta\Big(e^{-\gamma/2}, \sqrt{\lambda/(1-\lambda)}\Big) \leq \frac{3}{2}$, where $\theta(x,y) \coloneqq \sum_{n=0}^{\infty} x^{n^2} y^n$. This identifies just a portion of the full region of parameters of λ and γ where the infinite matrix A is positive semi-definite.

To analyse the positive semi-definiteness of the infinite matrix A further, let $A^{(d)}$ denote its $d \times d$ top-left corner. Note that it is well-known that an infinite matrix is positive semi-definite if and only if its $d \times d$ top-left corner is positive semi-definite for all $d \in \mathbb{N}$. For modest values of d, we can numerically determine the parameter region where $A^{(d)}$ is positive semi-definite. To achieve this, we plot in Fig. 6.3 the quantity

$$\eta_d(\gamma) \coloneqq \max\left(\lambda \in \left(\frac{1}{2}, 1\right] : \ A^{(d)} \text{ is positive semi-definite}\right), \tag{6.50}$$

with respect to $e^{-\gamma}$ for various values of d. This quantity is relevant because $A^{(d)}$ is positive semi-definite if and only if $\lambda \leq \eta_d(\gamma)$. Moreover, the quantity $\eta_d(\gamma)$ monotonically decreases in d and converges to some $\bar{\eta}(\gamma)$ as $d \to \infty$. Notably, the condition $\lambda \leq \bar{\eta}(\gamma)$ is necessary and sufficient for positive semi-definiteness of the infinite matrix A, and also a sufficient condition for the anti-degradability of the the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$. Our numerical investigation seems to suggest that when d is approximately 20, the quantity $\eta_d(\gamma)$ has already reached its limiting value $\bar{\eta}(\gamma)$, which can be approximated, for instance, by considering, the curve $\eta_{30}(\gamma)$.

Anti-degrading maps

Theorem 4 discovers parameter regions of transmissivity λ and dephasing γ in which the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable. Although the proof of this theorem ensures the existence of anti-degrading maps for $\mathcal{N}_{\lambda,\gamma}$ within these parameter regions, it does not offer explicit constructions of such anti-degrading maps. In the forthcoming Theorem 6, we present such explicit constructions. Note that, thanks to Lemma 51, the output operators of the complementary channel $\mathcal{N}^{\mathrm{c}}_{\lambda,\gamma}$ reside within the space $\mathcal{T}(\mathcal{H}_{E_{\mathrm{out}}})$, where $\mathcal{H}_{E_{\mathrm{out}}}$ is the following subspace of the two-mode Hilbert space $\mathcal{H}_{E_1} \otimes \mathcal{H}_{E_2} = L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$:

$$\mathcal{H}_{E_{\mathrm{out}}} \coloneqq \mathrm{Span}\left\{|\ell\rangle_{E_1} \otimes |\sqrt{\gamma} n\rangle_{E_2} : \ \ell \leq n \text{ with } \ell, n \in \mathbb{N}\right\}, \tag{6.51}$$

where $|n\rangle$ represents the nth Fock state, and $|\sqrt{\gamma} n\rangle$ denotes the coherent state with a parameter of $\sqrt{\gamma} n$. These states correspond to the environmental modes of the pure-loss and dephasing channels, respectively (we shall maintain this notation throughout).

Theorem 6. Anti-degrading maps corresponding to each parameter region in Theorem 4 can be defined as follows. In the region (i), i.e. $\lambda \in \left[0, \frac{1}{2}\right]$ and $\gamma \geq 0$, an anti-degrading map is given by

$$\mathcal{A}_{\lambda,\gamma} = \left(\mathcal{E}_{\frac{\lambda}{1-\lambda}} \circ \mathcal{R}_{E_1}\right) \otimes \mathrm{Tr}_{E_2}, \tag{6.52}$$

where $\mathcal{R}_{E_1}(\cdot) \coloneqq (-1)^{\hat{e}_1^\dagger \hat{e}_1} \cdot (-1)^{\hat{e}_1^\dagger \hat{e}_1}$, with $\hat{e}_1$ as the annihilation operator of the output mode of the pure-loss channel E_1.

In the region (ii), i.e. $\lambda \in (\frac{1}{2}, 1)$ and γ such that $\theta\Big(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\Big) \leq \frac{3}{2}$, an anti-degrading map $\mathcal{A}_{\lambda,\gamma} : \mathcal{T}(\mathcal{H}_{E_{\mathrm{out}}}) \to \mathcal{T}(\mathcal{H}_{\mathrm{B}})$, with $\mathcal{H}_{E_{\mathrm{out}}}$ given by (6.51), is defined as follows. For all $\ell_1, \ell_2, n_1, n_2 \in \mathbb{N}$ with $\ell_1 \leq n_1$ and $\ell_2 \leq n_2$, it holds that

$$\mathcal{A}_{\lambda,\gamma}(|\ell_1\rangle\langle\ell_1|\, \ell_2 \otimes |\sqrt{\gamma} n_1\rangle\langle\sqrt{\gamma} n_1|\, \sqrt{\gamma} n_2) \coloneqq \sum_{k=0}^{\min(n_1-\ell_1, n_2-\ell_2)} c_k^{(\ell_1,\ell_2,n_1,n_2)}\, |k+\ell_1\rangle\langle k+\ell_1|\, k+\ell_2\,, \tag{6.53}$$

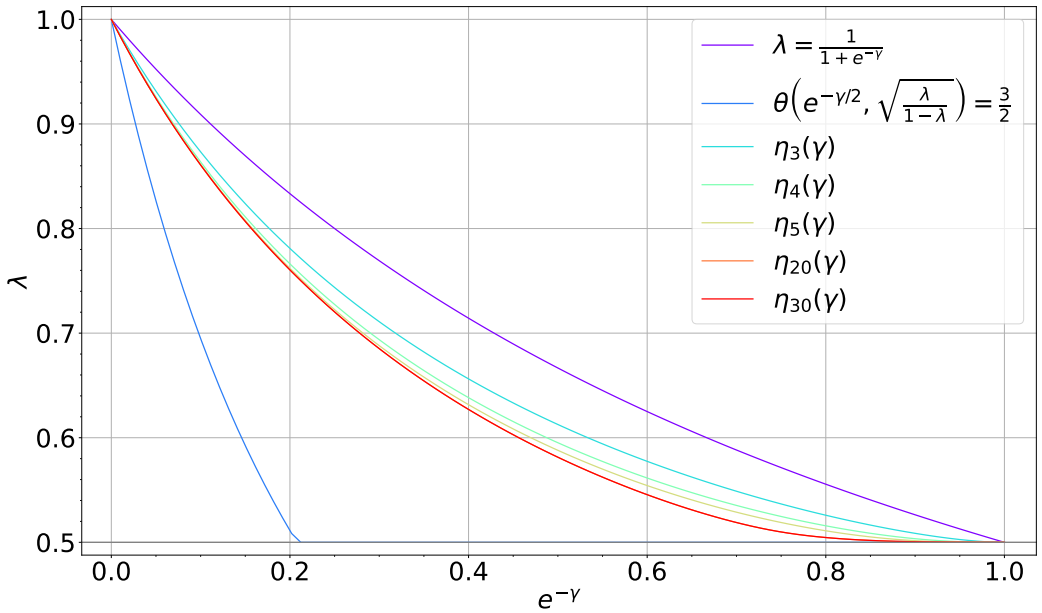


Figure 6.3: Numerical estimation of the anti-degradability of the loss-dephasing channel. The horizontal axis shows the quantity $e^{-\gamma}$, varying from 0 to 1 as dephasing parameter γ decreases from ∞ to 0, while the vertical axis corresponds to the transmissivity λ. Theorem 4 establishes that below the region defined by the blue curve, corresponding to the condition $\theta(e^{-\gamma/2}, \sqrt{\lambda/(1-\lambda)}) = 3/2$, the bosonic loss-dephasing channel is anti-degradable. Moreover, Theorem 8 establishes that above the region defined by the purple curve, corresponding to $\lambda = \frac{1}{1+e^{-\gamma}}$, the bosonic loss-dephasing channel is not anti-degradable. The other curves depict the quantity $\eta_d(\gamma)$, which is defined in (6.50) as the maximum value of the transmissivity where $A^{(d)}$ is positive semi-definite, for various values of d. Our numerical analysis seems to indicate that in the region below the red curve, corresponding to $\lambda \leq \eta_{30}(\gamma)$, the bosonic loss-dephasing channel is anti-degradable. Here, we employ $d = 30$, as increasing d beyond $d \geq 20$ yields no discernible change in the plot.

where for all $k \in \{0, 1, \ldots, \min(n_1 - \ell_1, n_2 - \ell_2)\}$ the coefficients $c_k^{(\ell_1,\ell_2,n_1,n_2)}$ are defined as

$$c_k^{(\ell_1,\ell_2,n_1,n_2)} := (-1)^{\ell_1-\ell_2} \sqrt{\mathcal{B}_k\left(n_1 - \ell_1, \frac{2\lambda-1}{\lambda}\right) \mathcal{B}_k\left(n_2 - \ell_2, \frac{2\lambda-1}{\lambda}\right)}\, a_{n_1-\ell_1 n_2-\ell_2}\, a_{k+\ell_1 k+\ell_2}\,, \tag{6.54}$$

where $\mathcal{B}_l(n, \lambda) := \binom{n}{l}\lambda^l(1-\lambda)^{n-l}$, and a_{mn} is defined in (6.44).

Proof of Theorem 6. Let us suppose that λ and γ fall within the parameter region (i). A complementary channel of the pure-loss channel $\mathcal{E}_\lambda$ is given by $\mathcal{E}_\lambda^{\mathrm{c}} = \mathcal{E}_{1-\lambda} \circ \mathcal{R}$. Consequently, Lemma 41 implies that $\left(\mathcal{E}_{\frac{\lambda}{1-\lambda}} \circ \mathcal{R}\right) \circ \mathcal{E}_\lambda^{\mathrm{c}} = \mathcal{E}_\lambda$, i.e. the channel $\mathcal{E}_{\frac{\lambda}{1-\lambda}} \circ \mathcal{R}$ is an anti-degrading map of the pure-loss channel. The general construction detailed in the proof of Lemma 71 for the anti-degrading map of the composition between an anti-degradable channel and another channel demonstrates that the map given in (6.52) is an anti-degrading map of $\mathcal{N}_{\lambda,\gamma}$.

Let us now suppose that λ and γ fall within the parameter region (ii). To come up with the anti-degrading map defined in (6.53), we drew intuition from the proof of Lemma 70, which demonstrates the equivalence between two-extendibility of the Choi state and the existence of an anti-degrading map, while also considering the two-extension of the Choi state of $\mathcal{N}_{\lambda,\gamma}$ explicitly found in (6.46). In order to show that the map $\mathcal{A}_{\lambda,\gamma}$ in (6.53) is an anti-degrading map of $\mathcal{N}_{\lambda,\gamma}$, we need to show that it is a quantum channel satisfying $\mathcal{A}_{\lambda,\gamma} \circ \mathcal{N}_{\lambda,\gamma}^{\mathrm{c}} = \mathcal{N}_{\lambda,\gamma}$. We begin by proving that $\mathcal{A}_{\lambda,\gamma}$ is trace preserving. By linearity, it suffices to show that for any $\ell_1, \ell_2, n_1, n_2 \in \mathbb{N}$ with

$\ell_1 \leq n_1$ and $\ell_2 \leq n_2$ it holds that

$$\mathrm{Tr}\Big[\mathcal{A}_{\lambda,\gamma}\Big(|\ell_1\rangle\langle\ell_1|\,\ell_2{}_{E_1} \otimes |\sqrt{\gamma}n_1\rangle\langle\sqrt{\gamma}n_1|\,\sqrt{\gamma}n_2{}_{E_2}\Big)\Big] = \mathrm{Tr}\Big[|\ell_1\rangle\langle\ell_1|\,\ell_2{}_{E_1} \otimes |\sqrt{\gamma}n_1\rangle\langle\sqrt{\gamma}n_1|\,\sqrt{\gamma}n_2{}_{E_2}\Big]\,. \tag{6.55}$$

Indeed, we obtain

$$\begin{aligned}
\mathrm{Tr}\Big[\mathcal{A}_{\lambda,\gamma}\Big(|\ell_1\rangle\langle\ell_1|\,\ell_2{}_{E_1} \otimes |\sqrt{\gamma}m\rangle\langle\sqrt{\gamma}m|\,\sqrt{\gamma}n_{E_2}\Big)\Big] &= \delta_{\ell_1,\ell_2} \sum_{k=0}^{\min(m-\ell_1,n-\ell_1)} c_k^{(\ell_1,\ell_1,m,n)} \\
&= \delta_{\ell_1,\ell_2} \sum_{k=0}^{\min(m-\ell_1,n-\ell_1)} \sqrt{\mathcal{B}_k\Big(m-\ell_1,\frac{2\lambda-1}{\lambda}\Big)\,\mathcal{B}_k\Big(n-\ell_1,\frac{2\lambda-1}{\lambda}\Big)}\, a_{m-\ell_1,n-\ell_1}\, a_{k+\ell_1,k+\ell_1} \\
&= \delta_{\ell_1,\ell_2} e^{-\frac{\gamma}{2}(m-n)^2} \\
&= \mathrm{Tr}\Big[|\ell_1\rangle\langle\ell_1|\,\ell_2{}_{E_1} \otimes |\sqrt{\gamma}n_1\rangle\langle\sqrt{\gamma}n_1|\,\sqrt{\gamma}n_2{}_{E_2}\Big]\,,
\end{aligned}$$

where δ_{ℓ_1,ℓ_2} denotes the kronecker delta and where we have exploited the formula for the overlap between coherent states provided in (6.15). Now, let us show that $\mathcal{A}_{\lambda,\gamma}$ is completely positive. To achieve this, we need to find a pure state Ψ on $\mathcal{H}_{\mathrm{anc}} \otimes \mathcal{H}_{E_{\mathrm{out}}}$, where $\mathcal{H}_{\mathrm{anc}}$ in an auxiliary reference, such that $\mathrm{Tr}_{\mathrm{anc}}[|\Psi\rangle\langle\Psi|] > 0$ and $(\mathrm{Id}_{\mathrm{anc}} \otimes \mathcal{A}_{\lambda,\gamma})(|\Psi\rangle\langle\Psi|) \geq 0$ [61, 62, 69]. Let $\mathcal{H}_{\mathrm{anc}} := \mathcal{H}_A \otimes \mathcal{H}_{B_1} = L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$ and let us construct the pure state $|\Psi\rangle_{AB_1E_1E_2} \in \mathcal{H}_{\mathrm{anc}} \otimes \mathcal{H}_{E_{\mathrm{out}}}$ as follows:

$$\begin{aligned}
|\Psi\rangle_{AB_1E_1E_2} &:= U_\lambda^{B_1E_1} V_\gamma^{B_1E_2} |\psi(r)\rangle_{AB_1} |0\rangle_{E_1} |0\rangle_{E_2} \\
&= \frac{1}{\cosh(r)} \sum_{n=0}^{\infty} \sum_{\ell=0}^{n} (-1)^\ell \tanh^n(r) \sqrt{\mathcal{B}_l(n,1-\lambda)}\, |n\rangle_A |n-\ell\rangle_{B_1} |\ell\rangle_{E_1} |\sqrt{\gamma}n\rangle_{E_2}\,,
\end{aligned} \tag{6.56}$$

where $U_\lambda^{B_1E_1}$ is the beam splitter unitary, $V_\gamma^{B_1E_2}$ is the conditional displacement unitary, and $|\psi(r)\rangle_{AB_1}$ is the two-mode squeezed vacuum state with squeezing $r > 0$. Let us now show that $\mathrm{Tr}_{AB_1}\left[|\Psi\rangle\langle\Psi|_{AB_1E_1E_2}\right]$ is positive definite on $\mathcal{H}_{E_{\mathrm{out}}}$. Let $|\phi\rangle_{E_1E_2} \in \mathcal{H}_{E_{\mathrm{out}}}$. Since there exists $\bar{\ell}, \bar{n} \in \mathbb{N}$ with $\bar{l} \leq \bar{n}$ such that $\langle\phi|_{E_1E_2} |\bar{\ell}\rangle_{E_1} \otimes |\sqrt{\gamma}\bar{n}\rangle_{E_2} \neq 0$, (6.56) implies that

$$\begin{aligned}
&\langle\phi|_{E_1E_2} \mathrm{Tr}_{AB_1}\left[|\Psi\rangle\langle\Psi|_{AB_1E_1E_2}\right] |\phi\rangle_{E_1E_2} = \frac{1}{\cosh^2(r)} \sum_{n=0}^{\infty} \sum_{l=0}^{n} \tanh^{2n}(r)\, \mathcal{B}_\ell(n,1-\lambda) \left|\langle\phi|_{E_1E_2} |\ell\rangle_{E_1} \otimes |\sqrt{\gamma}n\rangle_{E_2}\right|^2 \\
&\geq \frac{1}{\cosh^2(r)} \tanh^{2\bar{n}}(r)\, \mathcal{B}_{\bar{l}}(\bar{n},1-\lambda) \left|\langle\phi|_{E_1E_2} |\bar{\ell}\rangle_{E_1} \otimes |\sqrt{\gamma}\bar{n}\rangle_{E_2}\right|^2 \\
&> 0\,.
\end{aligned}$$

We next show that $\mathrm{Id}_{AB_1} \otimes \mathcal{A}_{\lambda,\gamma}(|\Psi\rangle\langle\Psi|)$ is positive semi-definite. Let B_2 denote the output system of $\mathcal{A}_{\lambda,\gamma}$. Note that

$$\begin{aligned}
\mathrm{Id}_{AB_1} \otimes \mathcal{A}_{\lambda,\gamma}(|\Psi\rangle\langle\Psi|_{AB_1E_1E_2}) &\overset{\text{(i)}}{=} \frac{1}{\cosh^2(r)} \sum_{m,n=0}^{\infty} \sum_{\ell_1=0}^{m} \sum_{\ell_2=0}^{n} (-1)^{\ell_1+\ell_2} [\tanh(r)]^{m+n} \sqrt{\mathcal{B}_{\ell_1}(m,1-\lambda)\mathcal{B}_{\ell_2}(n,1-\lambda)} \\
&\qquad |m\rangle\langle m|\,n_A \otimes |m-\ell_1\rangle\langle m-\ell_1|\,n-\ell_2{}_{B_1} \otimes \mathcal{A}_{\lambda,\gamma}\Big(|\ell_1\rangle\langle\ell_1|\,\ell_2{}_{E_1} \otimes |\sqrt{\gamma}m\rangle\langle\sqrt{\gamma}m|\,\sqrt{\gamma}n_{E_2}\Big) \\
&\overset{\text{(ii)}}{=} \frac{1}{\cosh^2(r)} \sum_{m,n=0}^{\infty} \sum_{\ell_1=0}^{m} \sum_{\ell_2=0}^{n} \sum_{k=0}^{\min(m-\ell_1,n-\ell_2)} [\tanh(r)]^{m+n} \\
&\qquad \sqrt{\mathcal{B}_{\ell_1}(m,1-\lambda)\,\mathcal{B}_{\ell_2}(n,1-\lambda)\,\mathcal{B}_k\left(m-\ell_1,\frac{2\lambda-1}{\lambda}\right)\mathcal{B}_k\left(n-\ell_2,\frac{2\lambda-1}{\lambda}\right)} \\
&\qquad a_{m-\ell_1,n-\ell_2}\, a_{k+\ell_1,k+\ell_2}\, |n_1\rangle\langle n_1|\,n_A \otimes |m-\ell_1\rangle\langle m-\ell_1|\,n-\ell_2{}_{B_1} \otimes |k+\ell_1\rangle\langle k+\ell_1|\,k+\ell_2{}_{B_2} \\
&\overset{\text{(iii)}}{=} \rho_{AB_1B_2}\,.
\end{aligned} \tag{6.57}$$

Here, in (i) we used the definition of $|\Psi\rangle_{AB_1E_1E_2}$ given in (6.56); in (ii) we utilised the definition of the map $\mathcal{A}_{\lambda,\gamma}$ from (6.53); and in (iii) we recognised the tripartite operator $\rho_{AB_1B_2}$ defined

in (6.46), which is a quantum state provided that λ and γ satisfy $\theta\left(e^{-\gamma/2}, \sqrt{\frac{\lambda}{1-\lambda}}\right) \leq \frac{3}{2}$. Therefore, in such a parameter region, $\mathrm{Id}_{AB_1} \otimes \mathcal{A}_{\lambda,\gamma}(|\Psi\rangle\langle\Psi|_{AB_1E_1E_2})$ is positive semi-definite, and thus $\mathcal{A}_{\lambda,\gamma}$ is a quantum channel. Let us now verify that $\mathcal{A}_{\lambda,\gamma} \circ \mathcal{N}^{\mathrm{c}}_{\lambda,\gamma} = \mathcal{N}_{\lambda,\gamma}$. To show this, note that

$$\begin{aligned} \mathrm{Id}_A \otimes \left(\mathcal{A}_{\lambda,\gamma} \circ \mathcal{N}^{\mathrm{c}}_{\lambda,\gamma}\right)\left(|\psi(r)\rangle\langle\psi(r)|_{AB_1}\right) &\overset{\text{(iv)}}{=} \mathrm{Tr}_{B_1}\left[\mathrm{Id}_{AB_1} \otimes \mathcal{A}_{\lambda,\gamma}(|\Psi\rangle\langle\Psi|_{AB_1E_1E_2})\right] \\ &\overset{\text{(v)}}{=} \mathrm{Tr}_{B_1}\left[\rho_{AB_1B_2}\right] \\ &\overset{\text{(vi)}}{=} \mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}\left(|\psi(r)\rangle\langle\psi(r)|_{AB_2}\right), \end{aligned} \tag{6.58}$$

Here, in (iv) we employed (6.56); in (v) we exploited (6.57); and in (vi) we used that $\rho_{AB_1B_2}$ is a two-extension of $\mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}\left(|\psi(r)\rangle\langle\psi(r)|\right)$, as established in the proof of Theorem 4. Finally, since the two-mode squeezed vacuum state $|\psi(r)\rangle_{AB}$ satisfies $\mathrm{Tr}_B[|\psi(r)\rangle\langle\psi(r)|_{AB}] > 0$, we conclude $\mathcal{A}_{\lambda,\gamma} \circ \mathcal{N}^{\mathrm{c}}_{\lambda,\gamma} = \mathcal{N}_{\lambda,\gamma}$. □

Analysis of the bosonic loss-dephasing channel via its finite-dimensional restrictions

Definition 55. *Let $d \in \mathbb{N}$ and let $\mathcal{H}_d := \mathrm{Span}(\{|n\rangle\}_{n=0,\ldots,d-1})$ be the subspace spanned by the first d Fock states. The qudit restriction of the bosonic loss-dephasing channel $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is a quantum channel defined by*

$$\mathcal{N}^{(d)}_{\lambda,\gamma}(\Theta) := \mathcal{N}_{\lambda,\gamma}(\Theta) \qquad \forall\, \Theta \in \mathcal{T}(\mathcal{H}_d) \tag{6.59}$$

Lemma 56. *Let $\mathcal{H}_A, \mathcal{H}_B := L^2(\mathbb{R})$. Let $\mathcal{N} : \mathcal{T}(\mathcal{H}_A) \to \mathcal{T}(\mathcal{H}_B)$ be a quantum channel and let $\mathcal{N}^{(d)}$ be its qudit restriction, defined by*

$$\mathcal{N}^{(d)}(\Theta) := \mathcal{N}(\Theta) \qquad \forall\, \Theta \in \mathcal{T}(\mathcal{H}_d)\,. \tag{6.60}$$

If $\mathcal{N}$ is (anti-)degradable, then its qudit restriction $\mathcal{N}^{(d)}$ is (anti-)degradable.

Proof. Let $\mathcal{N}(\cdot) = \mathrm{Tr}_E[V_{A\to BE}(\cdot)V^\dagger_{A\to BE}]$ be a Stinespring representation of $\mathcal{N}$, and let $\mathcal{N}^{\mathrm{c}}(\cdot) = \mathrm{Tr}_B[V_{A\to BE}(\cdot)V^\dagger_{A\to BE}]$ be the associated complementary channel. Note that the isometry $V_{A\to BE}$ provides a Stinespring representation also for the qudit restriction $\mathcal{N}^{(d)}$. Hence, the qudit restriction of the complementary channel $\mathcal{N}^{\mathrm{c}}$ is a complementary channel of the qudit restriction $\mathcal{N}^{(d)}$, i.e.

$$(\mathcal{N}^{(d)})^{\mathrm{c}}(\Theta) = \mathcal{N}^{\mathrm{c}}(\Theta) \qquad \forall\, \Theta \in \mathcal{T}(\mathcal{H}_d)\,. \tag{6.61}$$

Moreover, note that any degrading or anti-degrading map of $\mathcal{N}$ is effective for all input states, including those restricted to $\mathcal{H}_d$. Consequently, an (anti-)degrading map of $\mathcal{N}$ is also an (anti-)degrading map of its qudit restriction $\mathcal{N}^{(d)}$. □

Corollary 57. *If the qudit restriction $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is not (anti)-degradable, then the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is also not (anti)-degradable.*

The following lemma shows that qudit restriction $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is a qu*d*it-to-qu*d*it channel, mapping the space spanned by $\{|n\rangle\}_{n=0,\ldots,d-1}$ into itself.

Lemma 58. *If the input state to the bosonic loss-dephasing channel is confined into the finite-dimensional subspace spanned by $\{|n\rangle\}_{n=0,\ldots,d-1}$, the resulting output state will similarly be confined to this subspace.*

Proof. Examining Lemma 50 reveals that the operator $|m\rangle\langle m|\, n$, when acted on by the bosonic loss-dephasing channel, is transformed into linear combinations of operators $\{|\ell\rangle\langle\ell|\, k\}_{\ell\leq m, k\leq n}$. This means that if the input state to the bosonic loss-dephasing channel is restricted to the d-dimensional subspace $\{|n\rangle\}_{n=0,\ldots,d-1}$, the output of the channel will reside within the same subspace. □

The qubit restriction $\mathcal{N}^{(2)}_{\lambda,\gamma}$ of the bosonic loss-dephasing channel coincides with the composition between the amplitude damping channel and the qubit dephasing channel [33, 34], which we dub *amplitude-phase damping channel*. Theorems 7 and 8 utilise the amplitude-phase damping channel $\mathcal{N}^{(2)}_{\lambda,\gamma}$ to find that the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is never degradable and, additionally, it is not anti-degradabile for $\lambda > \frac{1}{1+e^{-\gamma}}$, respectively.

The bosonic loss-dephasing channel is never degradable

The bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is never degradable, except when it coincides with either the bosonic dephasing channel ($\gamma > 0$ and $\lambda = 1$) or the degradable pure-loss channel ($\gamma = 0$ and $\lambda \geq \frac{1}{2}$), thereby complicating the derivation of its quantum capacity [1]. This result has been previously demonstrated in [1] through pages-long proof; however, here we provide a significantly simpler proof of this result.

Theorem 7. Let $\lambda \in [0,1]$ and $\gamma \geq 0$. The bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is degradable if and only if one of the following conditions is satisfied:

- $\gamma = 0$ and $\lambda \in [\frac{1}{2}, 1]$
- $\gamma \geq 0$ and $\lambda = 1$

Proof. Thanks to Corollary 57, a necessary condition for $\mathcal{N}_{\lambda,\gamma}$ to be degradable is the degradability of the amplitude-phase damping channel $\mathcal{N}^{(2)}_{\lambda,\gamma}$. We now apply [51, Theorem 4], which establishes a necessary condition on the degradability of any qubit channel. Specifically, the rank of the Choi state of a degradable qubit channel is necessarily less or equal to 2. By using the notation used in (6.12), the matrix associated with the Choi state of the amplitude-phase damping channel $C\left(\mathcal{N}^{(2)}_{\lambda,\gamma}\right)$ in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ is given by:

$$\frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & \sqrt{e^{-\gamma}\lambda} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\lambda & 0 \\ \sqrt{e^{-\gamma}\lambda} & 0 & 0 & \lambda \end{pmatrix}. \tag{6.62}$$

One can easily see that its rank is equal to 3 for $\gamma > 0$ and $\lambda \in (0,1)$. In addition, for $\lambda = 1$ the bosonic loss-dephasing channel coincides with the bosonic dephasing channel, $\mathcal{N}_{1,\gamma} = \mathcal{D}_{\gamma}$, which is degradable for any value of $\gamma \geq 0$ [40]. Finally, for $\gamma = 0$ the bosonic loss-dephasing channel coincides with the pure-loss channel, $\mathcal{N}_{\lambda,0} = \mathcal{E}_{\lambda}$, which is degradable if and only if $\lambda \in [\frac{1}{2}, 1]$ [38, 39]. □

Necessary condition on anti-degradability via qubit restriction

The next theorem establishes the parameter range where the bosonic loss-dephasing channel is not anti-degradable. We provide three different proofs for this theorem.

Theorem 8. Let $\gamma \geq 0$. If $\lambda > \frac{1}{1+e^{-\gamma}}$, then the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable.

Proof 1. Assume that $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable; then substituting $m = 0$ and $n = 1$ in (6.31) yields

$$\mathcal{A}_{\lambda,\gamma}\left(\mathcal{E}_{1-\lambda}(|0\rangle\langle 0|\,1) \otimes |0\rangle\langle 0|\,\sqrt{\gamma}\right) = -e^{-\frac{1}{2}\gamma}\mathcal{E}_{\lambda}(|0\rangle\langle 0|\,1)\,. \tag{6.63}$$

By exploiting $\mathcal{E}_{\lambda}(|0\rangle\langle 0|\,1) = \sqrt{\lambda}\,|0\rangle\langle 0|\,1$, we have

$$\mathcal{A}_{\lambda,\gamma}\left(|0\rangle\langle 0|\,1 \otimes |0\rangle\langle 0|\,\sqrt{\gamma}\right) = -\sqrt{\frac{e^{-\gamma}\lambda}{1-\lambda}}\,|0\rangle\langle 0|\,1\,.$$

Using Lemma 68 in the Appendix, we find $\sqrt{\frac{e^{-\gamma}\lambda}{1-\lambda}} \leq 1$, or $\lambda \leq \frac{1}{1+e^{-\gamma}}$. □

Proof 2. Assume that $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable. As a consequence of (6.31) and of the data-processing inequality for the fidelity [59], we find

$$F\Big(\mathcal{E}_\lambda(|0\rangle\langle 0|)\,,\,\mathcal{E}_\lambda(|1\rangle\langle 1|)\Big) \geq F\Big(\mathcal{E}_{1-\lambda}(|0\rangle\langle 0|)\otimes|0\rangle\langle 0|\;,\;\mathcal{E}_{1-\lambda}(|1\rangle\langle 1|)\otimes|\sqrt{\gamma}\rangle\langle\sqrt{\gamma}|\Big)\,. \tag{6.64}$$

Furthermore, we obtain

$$\begin{aligned}
\sqrt{1-\lambda} &= F\Big(|0\rangle\langle 0|\,,\lambda\,|1\rangle\langle 1|+(1-\lambda)\,|0\rangle\langle 0|\Big)\\
&\overset{\text{(i)}}{=} F\Big(\mathcal{E}_\lambda(|0\rangle\langle 0|),\mathcal{E}_\lambda(|1\rangle\langle 1|)\Big)\\
&\overset{(ii)}{\geq} F\Big(|0\rangle\langle 0|\,,|\sqrt{\gamma}\rangle\langle\sqrt{\gamma}|\Big)\,F\Big(\mathcal{E}_{1-\lambda}(|0\rangle\langle 0|),\mathcal{E}_{1-\lambda}(|1\rangle\langle 1|)\Big)\\
&\overset{(iii)}{\geq} F\Big(|0\rangle\langle 0|\,,|\sqrt{\gamma}\rangle\langle\sqrt{\gamma}|\Big)\,F\Big(|0\rangle\langle 0|\,,\lambda\,|0\rangle\langle 0|+(1-\lambda)\,|1\rangle\langle 1|\Big)\\
&\overset{\text{(iv)}}{=} \sqrt{e^{-\gamma}\lambda}\,.
\end{aligned} \tag{6.65}$$

Here, (i) follows from Lemma 38, (ii) follows from (6.64) and from the fact that the fidelity is multiplicative under tensor product [59], (iii) uses Lemma 38 again, and in (iv) we exploited that $|\langle 0|\sqrt{\gamma}\rangle| = \sqrt{e^{-\gamma}}$. This yields $\sqrt{1-\lambda} \geq \sqrt{e^{-\gamma}\lambda}$, or $\lambda \leq \frac{1}{1+e^{-\gamma}}$. □

Proof 3. By exploiting Lemma 32 and the Choi matrix of the qubit channel $\mathcal{N}^{(2)}_{\lambda,\gamma}$ reported in (6.62), one can easily obtain that $\mathcal{N}^{(2)}_{\lambda,\gamma}$ is anti-degradable if and only if $\lambda \leq \frac{1}{1+e^{-\gamma}}$. Consequently, thanks to Corollary 57, the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable for $\lambda > \frac{1}{1+e^{-\gamma}}$. □

Necessary condition on anti-degradability via qudit restrictions

Let us introduce the following quantity for any $d \in \mathbb{N}$ and $\gamma \geq 0$:

$$\lambda_d(\gamma) := \max\Big(\lambda \in [0,1] :\ \mathcal{N}^{(d)}_{\lambda,\gamma} \text{ is anti-degradable}\Big)\,. \tag{6.66}$$

This quantity is relevant since it allows us to find parameter region where the bosonic loss-dephasing channel is not anti-degradable. Specifically, for $\lambda > \lambda_d(\gamma)$ the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is not anti degradable, as established by Corollary 57. Thanks to the Proof 3 of Theorem 8, it follows that for $d = 2$ we have that $\lambda_2(\gamma) = \frac{1}{1+e^{-\gamma}}$, thereby establishing that $\mathcal{N}_{\lambda,\gamma}$ is not anti degradable for $\lambda > \frac{1}{1+e^{-\gamma}}$. Through an examination of larger values of d, we aim to identify an extended parameter region where the channel is not anti-degradable (see Fig. 6.4). We begin by proving some useful properties of the quantity $\lambda_d(\gamma)$.

Lemma 59. *For any $\gamma \geq 0$ and $d \in \mathbb{N}, d \geq 2$, the following facts hold:*

- *Fact 1: The qudit restriction of the bosonic loss-dephasing channel $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is anti-degradable if and only if $\lambda \leq \lambda_d(\gamma)$*
- *Fact 2: The quantity $\lambda_d(\gamma)$ is monotonically increasing in γ*
- *Fact 3: For $d = 2$, it precisely holds that $\lambda_2(\gamma) = \frac{1}{1+e^{-\gamma}}$*
- *Fact 4: The quantity $\lambda_d(\gamma)$ is monotonically non-increasing in d*
- *Fact 5: It holds that $\frac{1}{2} \leq \lambda_d(\gamma) \leq \frac{1}{1+e^{-\gamma}}$*
- *Fact 6: When $d = 3$ and $e^{-\gamma} \leq \sqrt{2}-1$ (or $\gamma \geq 0.881$), it exactly holds that $\lambda_3(\gamma) = \frac{1}{1+e^{-\gamma}}$*

Proof. Fact 1. It suffices to show that for any $\gamma \geq 0$ and $\lambda, \lambda' \in [0,1]$ with $\lambda' < \lambda$, if $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is anti-degradable, then so is $\mathcal{N}^{(d)}_{\lambda',\gamma}$. To show this, we exploit the composition rule

$$\mathcal{E}_{\lambda_1} \circ \mathcal{N}_{\lambda_2,\gamma} = \mathcal{N}_{\lambda_1\lambda_2,\gamma} \qquad \forall\, \lambda_1, \lambda_2 \in [0,1]\,, \tag{6.67}$$

as established by Lemma 53, implying that the channel $\mathcal{N}^{(d)}_{\lambda',\gamma}$ can be written as the composition between $\mathcal{N}^{(d)}_{\lambda,\gamma}$ and another channel. Consequently, Lemma 71 concludes the proof.

Fact 2. Analogously to Fact 1, it suffices to show that for any $\lambda \in [0,1]$ and for any $\gamma' \geq \gamma \geq 0$, if $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is anti-degradable, then so is $\mathcal{N}^{(d)}_{\lambda,\gamma'}$. This follows from the composition rule

$$\mathcal{D}_{\gamma_1} \circ \mathcal{N}_{\lambda,\gamma_2} = \mathcal{N}_{\lambda,\gamma_1+\gamma_2} \qquad \forall\, \gamma_1, \gamma_2 \geq 0\,, \tag{6.68}$$

as proved in Lemma 53. Furthermore, Lemma 71 concludes the proof.

Fact 3. This has already been proved in the Proof 3 of Theorem 8.

Fact 4. This follows from the observation that for all $d' \geq d$, if $\mathcal{N}^{(d')}_{\lambda,\gamma}$ is anti-degradable, then so is $\mathcal{N}^{(d)}_{\lambda,\gamma}$.

Fact 5. The upper bound $\lambda_d(\gamma) \leq \frac{1}{1+e^{-\gamma}}$ follows from Fact 3 and Fact 4. Moreover, since the pure-loss channel $\mathcal{E}_\lambda$ is anti-degradable if and only if $\lambda \leq \frac{1}{2}$, Fact 4 implies that $\lambda_d(0) \geq \frac{1}{2}$ (more specifically, one can also show that $\lambda_d(0) = \frac{1}{2}$). Consequently, Fact 2 concludes the proof.

Fact 6. This proof relies on the equivalence between anti-degradability of a channel and two-extendibility of its Choi state, as established in Lemma 70. Let λ and γ be such that $\lambda = \frac{1}{1+e^{-\gamma}}$ and $e^{-\gamma} \leq \sqrt{2}-1$, implying that $\lambda \geq \frac{1}{\sqrt{2}}$. By using Lemma 50, we obtain the Choi state of $\mathcal{N}^{(3)}_{\lambda,\gamma}$ as follows:

$$\mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}(|\Phi_3\rangle\langle\Phi_3|) = \frac{1}{3}\sum_{m=0}^{2}\sum_{n=0}^{2}\sum_{\ell=0}^{\min(m,n)} e^{-\frac{\gamma}{2}(m-n)^2}\sqrt{\binom{m}{\ell}\binom{n}{\ell}}\lambda^{\frac{m+n}{2}-\ell}(1-\lambda)^l\, |m\rangle\langle m|\, n_A \otimes |m-\ell\rangle\langle m-\ell|\, n-\ell_B,$$

where Φ_3 is the maximally entangled state of Schmidt rank 3. We define a two-extension $\tilde{\rho}_{AB_1B_2}$ of the Choi state by the following conditions. First, $\tilde{\rho}_{AB_1B_2}$ satisfies the following $B_1 \leftrightarrow B_2$ symmetry for all $i_1, i_2, i_3, j_1, j_2, j_3 \in \{0,1,2\}$:

$$\begin{aligned} \langle j_1|_A \langle j_2|_{B_1} \langle j_3|_{B_2} \tilde{\rho}_{AB_1B_2} |i_1\rangle_A |i_2\rangle_{B_1} |i_3\rangle_{B_2} &= \langle j_1|_A \langle j_3|_{B_1} \langle j_2|_{B_2} \tilde{\rho}_{AB_1B_2} |i_1\rangle_A |i_2\rangle_{B_1} |i_3\rangle_{B_2} \\ \langle j_1|_A \langle j_2|_{B_1} \langle j_3|_{B_2} \tilde{\rho}_{AB_1B_2} |i_1\rangle_A |i_2\rangle_{B_1} |i_3\rangle_{B_2} &= \langle j_1|_A \langle j_2|_{B_1} \langle j_3|_{B_2} \tilde{\rho}_{AB_1B_2} |i_1\rangle_A |i_3\rangle_{B_1} |i_2\rangle_{B_2}\,. \end{aligned} \tag{6.69}$$

Furthermore, if $i_1 < \max(i_2, i_3)$ or $j_1 < \max(j_2, j_3)$, or if $i_1 > i_2 + i_3$ or $j_1 > j_2 + j_3$, then $\langle j_1|_A \langle j_2|_{B_1} \langle j_3|_{B_2} \tilde{\rho}_{AB_1B_2} |i_1\rangle_A |i_2\rangle_{B_1} |i_3\rangle_{B_2} = 0$. We can thus define $\tilde{\rho}_{AB_1B_2}$ by writing only the matrix elements with respect the set $\{|i_1\rangle_A |i_2\rangle_{B_1} |i_3\rangle_{B_2}\}$ with $2 \geq i_1 \geq i_2 \geq i_3 \geq 0$ such that $i_2 + i_3 \geq i_1$. Hence, in order to fully define $\tilde{\rho}_{AB_1B_2}$, it suffices to write the matrix elements of $\tilde{\rho}_{AB_1B_2}$ with respect to the set $\{\, |0\rangle_A |0\rangle_{B_1} |0\rangle_{B_2}, |1\rangle_A |0\rangle_{B_1} |0\rangle_{B_2}, |1\rangle_A |1\rangle_{B_1} |1\rangle_{B_2}, |2\rangle_A |1\rangle_{B_1} |1\rangle_{B_2}\,,$ $|2\rangle_A |2\rangle_{B_1} |0\rangle_{B_2}, |2\rangle_A |2\rangle_{B_1} |1\rangle_{B_2}\,, |2\rangle_A |2\rangle_{B_1} |2\rangle_{B_2}\,\}$. This gives rise to the following 7×7 matrix:

	$0_A0_{B_1}0_{B_2}$	$1_A1_{B_1}0_{B_2}$	$1_A1_{B_1}1_{B_2}$	$2_A1_{B_1}1_{B_2}$	$2_A2_{B_1}0_{B_2}$	$2_A2_{B_1}1_{B_2}$	$2_A2_{B_1}2_{B_2}$
$0_A0_{B_1}0_{B_2}$	1	$\sqrt{1-\lambda}$	0	$\sqrt{2}(1-\lambda)$	$\frac{(1-\lambda)^2}{\lambda}$	0	0
$1_A1_{B_1}0_{B_2}$	$\sqrt{1-\lambda}$	$1-\lambda$	0	$\sqrt{2(1-\lambda)^3}$	$\sqrt{\frac{(1-\lambda)^5}{\lambda^2}}$	0	0
$1_A1_{B_1}1_{B_2}$	0	0	$2\lambda-1$	0	0	$\frac{2\lambda-1}{\lambda}\sqrt{1-\lambda}$	0
$2_A1_{B_1}1_{B_2}$	$\sqrt{2}(1-\lambda)$	$\sqrt{2(1-\lambda)^3}$	0	$2(1-\lambda)^2$	$\sqrt{2}\frac{(1-\lambda)^3}{\lambda}$	0	0
$2_A2_{B_1}0_{B_2}$	$\frac{(1-\lambda)^2}{\lambda}$	$\sqrt{\frac{(1-\lambda)^5}{\lambda^2}}$	0	$\sqrt{2}\frac{(1-\lambda)^3}{\lambda}$	$(1-\lambda)^2$	0	0
$2_A2_{B_1}1_{B_2}$	0	0	$\frac{2\lambda-1}{\lambda}\sqrt{1-\lambda}$	0	0	$2(1-\lambda)(2\lambda-1)$	0
$2_A2_{B_1}2_{B_2}$	0	0	0	0	0	0	$(2\lambda-1)^2$

One can show by direct calculation that this matrix is positive semi-definite if and only $\lambda \geq \frac{1}{\sqrt{2}}$. Note that the matrix is positive semi-definite if and only if $\tilde{\rho}_{AB_1B_2}$ is positive semi-definite. The latter follows from the following two simple facts: (i) A $n \times n$ symmetric matrix with a duplicate column is positive semi-definite if and only if the $(n-1) \times (n-1)$ matrix obtained by deleting one of the two equal column and the corresponding row is positive semi-definite, and (ii) A $n \times n$ symmetric matrix with a zero column is positive semi-definite if and only if the $(n-1) \times (n-1)$ matrix obtained by deleting such a column and the corresponding row is positive semi-definite. One can also verify $\mathrm{Tr}_{B_1} \tilde{\rho}_{AB_1B_2} = \mathrm{Tr}_{B_2} \tilde{\rho}_{AB_1B_2}$, and they are equal to the Choi state of the qutrit restriction with $\lambda = \frac{1}{1+e^{-\gamma}}$. We therefore conclude that the curve $\lambda = \frac{1}{1+e^{-\gamma}}$ provides a necessary and sufficient condition for the anti-degradability of the qutrit channel $\mathcal{N}^{(3)}_{\lambda,\gamma}$ when $\lambda \geq \frac{1}{\sqrt{2}}$, or equivalently when $e^{-\gamma} \leq \sqrt{2}-1$. □

To numerically compute the quantity $\lambda_d(\gamma)$ given in (6.66), we utilise the equivalence between anti-degradability of a channel and two-extendibility of its Choi state [41]. Specifically, for small values of d, we can determine necessary and sufficient conditions for the anti-degradability of $\mathcal{N}^{(d)}_{\lambda,\gamma}$ by numerically solving the following *semi-definite program*:

$$\begin{aligned} &\min_{\rho_{AB_1B_2}} \ 1 \\ &\text{s.t. } \rho_{AB_1B_2} \geq 0\,, \\ &\quad \mathrm{Tr}[\rho_{AB_1B_2}] = 1, \\ &\quad \mathrm{Tr}_{B_2}\left[\rho_{AB_1B_2}\right] = \mathrm{Id}_A \otimes \mathcal{N}^{(d)}_{\lambda,\gamma}(|\Phi_d\rangle\langle\Phi_d|_{AA'})\,, \\ &\quad \mathrm{Tr}_{B_1}\left[\rho_{AB_1B_2}\right] = \mathrm{Id}_A \otimes \mathcal{N}^{(d)}_{\lambda,\gamma}(|\Phi_d\rangle\langle\Phi_d|_{AA'})\,. \end{aligned} \tag{6.70}$$

where $|\Phi_d\rangle$ is the maximally entangled state of schmidt rank d. The channel $\mathcal{N}^{(d)}_{\lambda,\gamma}$ is anti-degradable if and only if the semi-definite program admits a feasible solution. We compute the quantity $\lambda_d(\gamma)$ defined in (6.66) by numerically solving the semi-definite program. The results are plotted with respect to $e^{-\gamma}$ for various values of d in Fig. 6.4, showcasing the dependence of $\lambda_d(\gamma)$ on γ for small values of d. Our numerical analysis reveals that when γ is sufficiently large, i.e. $e^{-\gamma} \lesssim 0.41$ or $\gamma \gtrsim 0.89$, the value of $\lambda_d(\gamma)$ consistently equals $\frac{1}{1+e^{-\gamma}}$ for all examined values of d. In particular, this seems to suggest that within this range of the dephasing parameter, $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable if and only if $\lambda \leq \frac{1}{1+e^{-\gamma}}$. Based on this numerical exploration, we propose the following conjecture:

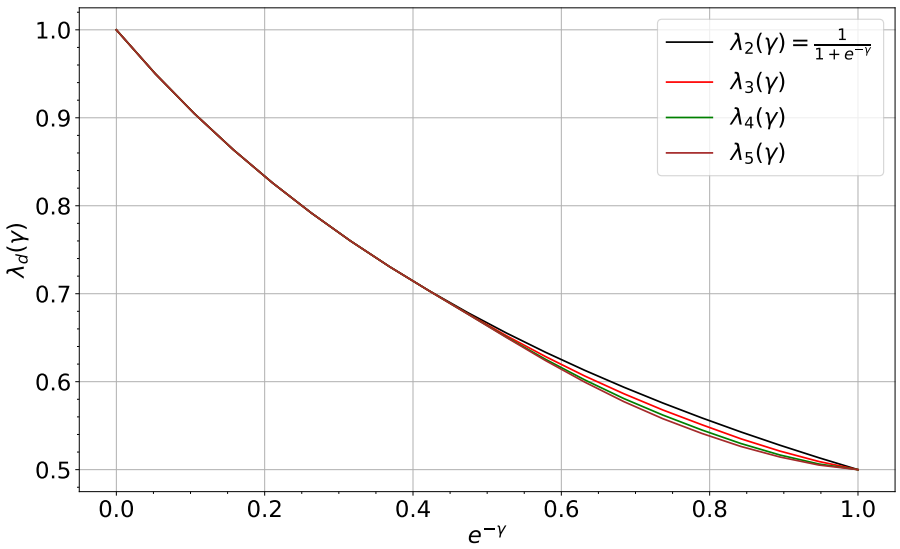


Figure 6.4: Each curve indicates necessary and sufficient conditions where the qu*d*it restriction $\mathcal{N}^{(d)}_{\lambda,\gamma}$ of the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable. In the region above the curve $\lambda_5(\gamma)$, the bosonic loss-dephasing channel is never anti-degradable.

Conjecture 60. *If γ is sufficiently large ($\gamma \gtrsim 0.89$), then the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is anti-degradable if and only if $\lambda \leq \frac{1}{1+e^{-\gamma}}$.*

Notably, from Fig. 6.4 we observe that if λ and γ satisfy $\lambda = \frac{1}{1+e^{-\gamma}}$ with $e^{-\gamma} \gtrsim 0.41$ (or $\gamma \lesssim 0.89$), then $\mathcal{N}_{\lambda,\gamma}$ is not anti-degradable.

6.7.3 Generalisation of our methods to general bosonic dephasing channels

In Theorem 4 we introduced a method to analyse anti-degradability of the bosonic loss-dephasing channel. In this section, we show that this method can be applied also to analyse the anti-degradability of the composition between a *general* bosonic dephasing channel and the pure-loss channel channel.

Given a probability distribution $p(\cdot)$ over $\mathbb{R}$, the associated *general bosonic dephasing channel* is given by

$$\mathcal{D}^{(p)}(X) := \int_{-\infty}^{\infty} \mathrm{d}\phi\, p(\phi)\, e^{i\phi\hat{a}^\dagger\hat{a}}\, X\, e^{-i\phi\hat{a}^\dagger\hat{a}}\,. \tag{6.71}$$

If $p(\phi)$ is the Gaussian distribution $p(\phi) := \frac{1}{\sqrt{2\pi\gamma}} e^{-\frac{\phi^2}{2\gamma}}$, the general bosonic dephasing channel $\mathcal{D}^{(p)}$ exactly coincides with the bosonic dephasing channel $\mathcal{D}_\gamma$ analysed in this work. The action of $\mathcal{D}^{(p)}$ on operators of the form $|n\rangle\langle n|\, m$ is given by

$$\mathcal{D}^{(p)}(|n\rangle\langle n|\, m) = \tilde{p}(n-m)\, |n\rangle\langle n|\, m\,, \tag{6.72}$$

where $\tilde{p}$ is the Fourier transform of the probability distribution p, i.e.

$$\tilde{p}(k) := \int_{-\infty}^{\infty} \mathrm{d}\phi\, p(\phi)\, e^{ik\phi}\,. \tag{6.73}$$

Let $\mathcal{N}^{(p)}_\lambda$ be the composition between such a general bosonic dephasing channel $\mathcal{D}^{(p)}$ and the pure-loss channel, i.e.

$$\mathcal{N}^{(p)}_\lambda := \mathcal{D}^{(p)} \circ \mathcal{E}_\lambda = \mathcal{E}_\lambda \circ \mathcal{D}^{(p)}\,. \tag{6.74}$$

We can apply the exact same method that we have introduced in the proof of Thereom 4 in order to analyse the anti-degradability of $\mathcal{N}_{\lambda}^{(p)}$. The key observation is that in the proof of Theorem 4 we did not use the explicit expression of the channel $\mathcal{D}^{(p)}$ before stating (6.46). This simple observation allows us to generalise our results to arbitrary bosonic dephasing channels, as stated in the following theorem.

Theorem 9 (Sufficient condition on the anti-degradability of the composition between a general bosonic dephasing channel and pure-loss channel)**.** Let $\lambda \in [0,1)$ and let $p(\cdot)$ be a probability distribution over $\mathbb{R}$. Let $A = (a_{mn})_{m,n\in\mathbb{N}}$ be the infinite matrix whose components are defined by

$$a_{mn} := \frac{\tilde{\phi}(n-m)}{\sum_{j=0}^{\min(n,m)} \sqrt{\mathcal{B}_j\left(n, \frac{2\lambda-1}{\lambda}\right)\, \mathcal{B}_j\left(m, \frac{2\lambda-1}{\lambda}\right)}}, \quad \forall\, n,m \in \mathbb{N}\,. \tag{6.75}$$

The channel $\mathcal{N}_{\lambda}^{(p)}$ is anti-degradable as long as either $\lambda \in [0, \frac{1}{2}]$ or the infinite matrix A is positive semi-definite.

6.7.4 Coherence preservation of the bosonic loss-dephasing channel

In this section we use the notation introduced in Section 6.7.1.

Theorem 10. Let $\gamma \geq 0$ and $\lambda \in (0,1]$. For any energy constraint $N_s > 0$, the energy-constrained two-way quantum and secret-key capacities of the bosonic loss-dephasing channel are strictly positive, $K(\mathcal{N}_{\lambda,\gamma}, N_s) \geq Q_2(\mathcal{N}_{\lambda,\gamma}, N_s) > 0$.

Proof. We begin by assuming $N_s \in (0,1)$ and defining the two-mode state

$$|\Psi_{N_s}\rangle_{AA'} := \sqrt{1-N_s}\,|00\rangle_{AA'} + \sqrt{N_s}\,|11\rangle_{AA'}\,, \tag{6.76}$$

where the mean photon number of A' system is equal to N_s. By exploiting Lemma 50, one can observe that the state

$$\rho_{AB} := \mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}) \tag{6.77}$$

is effectively a two-qubit state and its matrix with respect to the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ is given by:

$$\rho_{AB} = \begin{pmatrix} 1-N_s & 0 & 0 & \sqrt{(1-N_s)N_s e^{-\gamma\lambda}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & (1-\lambda)N_s & 0 \\ \sqrt{(1-N_s)N_s e^{-\gamma\lambda}} & 0 & 0 & \lambda N_s \end{pmatrix}.$$

If we perform partial transpose with respect to the system B, we find the matrix

$$(\rho_{AB})^{\intercal_B} = \begin{pmatrix} 1-N_s & 0 & 0 & 0 \\ 0 & 0 & \sqrt{(1-N_s)N_s e^{-\gamma\lambda}} & 0 \\ 0 & \sqrt{(1-N_s)N_s e^{-\gamma\lambda}} & (1-\lambda)N_s & 0 \\ 0 & 0 & 0 & \lambda N_s \end{pmatrix},$$

whose eigenvalues are not all positive, i.e. the state ρ_{AB} is not PPT [70]. By exploiting the fact that any two-qubit state is distillable if and only if it is not PPT [70], it follows that $E_d\left(\mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'})\right) > 0$, where E_d is the distillable entanglement. On the other hand, from Lemma 54 we have that

$$K(\mathcal{N}_{\lambda,\gamma}, N_s) \geq Q_2(\mathcal{N}_{\lambda,\gamma}, N_s) \geq E_d(\mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'}))\,. \tag{6.78}$$

This concludes the proof for $N_s \in (0,1)$. Since the energy-constrained capacities are monotonically non-decreasing in the energy constraint N_s, the proof follows for any $N_s > 0$. □

Since the state $\mathrm{Id}_A \otimes \mathcal{N}_{\lambda,\gamma}(|\Psi_{N_s}\rangle\langle\Psi_{N_s}|_{AA'})$ in (6.77) is always entangled, it follows that the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is never entanglement breaking [33]. We state this formally in the following theorem.

Theorem 11. For all $\gamma \geq 0$ and all $\lambda \in (0,1]$, the bosonic loss-dephasing channel $\mathcal{N}_{\lambda,\gamma}$ is not entanglement breaking.

In the following subsection we will find an explicit strictly positive lower bound on the two-way capacities of the bosonic loss-dephasing channel.

Multi-rail multi-photon encoding

Let $\Pi(N,k)$ be the set of partitions of N objects into k (possibly empty) parts. It is well known that $|\Pi(N,k)| = \binom{N+k-1}{k-1} = \binom{N+k-1}{N}$. Clearly, we can think of each $p \in \Pi(N,k)$ as a vector in $\mathbb{N}_+^k$, also denoted by p, with the constraint that $\sum_{\ell=1}^k p_\ell = N$. For each $p \in \Pi(N,k)$, define the associate k-mode Fock state as

$$|\psi_p\rangle \coloneqq |p_1\rangle \dots |p_k\rangle \ . \tag{6.79}$$

Note that for $p, q \in \Pi(N,k)$, we have that $\langle\psi_p|\psi_q\rangle = \delta_{p,q}$. Let us call

$$P_{N,k} \coloneqq \sum_{p\in\Pi(N,k)} \psi_p \tag{6.80}$$

the projector onto the k-mode subspace of total photon number N. Note that the bosonic dephasing channel satisfies that

$$\mathcal{D}_\gamma^{\otimes k}\left(|\psi_p\rangle\langle\psi_p|\,\psi_q\right) = e^{-\frac{\gamma}{2}\sum_\ell (p_\ell - q_\ell)^2}\,|\psi_p\rangle\langle\psi_p|\,\psi_q = e^{-\frac{\gamma}{2}\|p-q\|^2}\,|\psi_p\rangle\langle\psi_p|\,\psi_q = (K_{N,k,\gamma})_{pq}\,|\psi_p\rangle\langle\psi_p|\,\psi_q\,, \tag{6.81}$$

where $K_{N,k,\gamma}$ is the $\binom{N+k-1}{N} \times \binom{N+k-1}{N}$ matrix with entries

$$(K_{N,k,\gamma})_{pq} \coloneqq e^{-\frac{\gamma}{2}\|p-q\|^2}. \tag{6.82}$$

For any $\binom{N+k-1}{N}$-dimensional state σ, let us denote as $\overline{\sigma}$ the following isometrically equivalent state

$$\overline{\sigma} \coloneqq \sum_{p,q\in\Pi(N,k)} \sigma_{pq}\,|\psi_p\rangle\langle\psi_p|\,\psi_q\,. \tag{6.83}$$

The state $\overline{\sigma}$, which is termed as the *rail encoding* of σ, is supported on the subspace of k modes with total photon number equal to N. Now, let ρ be a $\binom{N+k-1}{N}$-dimensional state and let us calculate the output of $\mathcal{D}_\gamma^{\otimes k}$ when the input is $\overline{\rho}$:

$$\mathcal{D}_\gamma^{\otimes k}\left(\overline{\rho}\right) = \sum_{p,q\in\Pi(N,k)} \rho_{pq}(K_{N,k,\gamma})_{pq}\,|\psi_p\rangle\langle\psi_p|\,\psi_q = \overline{K_{N,k,\gamma}\circ\rho} = \overline{\Theta_{N,k,\gamma}(\rho)}\,, \tag{6.84}$$

where the operation $\circ$ denotes the element-wise product between matrices and where we have introduced the following Hadamard channel:

$$\Theta_{N,k,\gamma}(X) \coloneqq K_{N,k,\gamma}\circ X\,, \tag{6.85}$$

Since $\mathcal{D}_\gamma^{\otimes k}$ is a (completely) positive map, this in particular shows that $K_{N,k,\gamma} \geq 0$. (This latter statement can also be proved directly with techniques similar to that in the proof of [71, Lemma 15].) In practice, the k-fold application of the bosonic dephasing channel on the k-mode N-photon code space behaves as a new Hadamard channel $\Theta_{N,k,\gamma}$ with associated matrix $K_{N,k,\gamma}$.

Lower bound on the two-way capacity of the bosonic loss-dephasing channel

Since under the action of $\mathcal{N}_{\lambda,\gamma}$ photons can only be lost and never added, and each photon has a probability λ of being transmitted, the probability that an N-photon state will retain N photons at the output of the channel is exactly λ^N. *If that happens*, then the state in the code space is effectively left untouched by the loss and only dephased under the action of the Hadamard channel $\Theta_{N,k,\gamma}$.

More formally, from the Kraus representation

$$\mathcal{E}_\lambda(X) = \sum_{n=0}^{\infty} \frac{1}{n!} (1-\lambda)^n \lambda^{\frac{a^\dagger a}{2}} a^n X (a^\dagger)^n \lambda^{\frac{a^\dagger a}{2}} \tag{6.86}$$

it is easy to deduce the handy identity

$$\mathcal{E}_\lambda^{\otimes k}\left(\overline{\rho}\right) = \lambda^N \overline{\rho} + \left(1-\lambda^N\right) \delta_{N,k,\lambda}\,, \tag{6.87}$$

valid for all $\binom{N+k-1}{N}$-dimensional states ρ, with the notation of (6.83). Here, $\delta_{N,k,\lambda}$ is a suitable k-mode state supported on the subspace of total photon number at most $N-1$, and thus $\overline{\rho}\delta_{N,k,\lambda} = \delta_{N,k,\lambda}\overline{\rho} = 0$. In turn, the above identity implies that

$$\mathcal{N}_{\lambda,\gamma}^{\otimes k}\left(\overline{\rho}\right) = \lambda^N \overline{K_{N,k,\gamma} \circ \rho} + \left(1-\lambda^N\right) \delta'_{N,k,\lambda,\gamma} = \lambda^N \overline{\Theta_{N,k,\gamma}(\rho)} + \left(1-\lambda^N\right) \delta'_{N,k,\lambda,\gamma}\,, \tag{6.88}$$

where once again $\delta'_{N,k,\lambda,\gamma}$ is a suitable k-mode state supported on the subspace of total photon number at most $N-1$.

Therefore, we can use the channel $\mathcal{N}_{\lambda,\gamma}^{\otimes k}$ to simulate $\Theta_{N,k,\gamma}$ *probabilistically*, with probability λ^N. The simulation works as follows:

(i) The input state ρ is encoded in the k-mode N-photon subspace according to the mapping $\rho \mapsto \overline{\rho}$.

(ii) The k-mode state $\overline{\rho}$ is sent across $\mathcal{N}_{\lambda,\gamma}^{\otimes k}$, via k uses of the bosonic loss-dephasing channel.

(iii) The total photon number is measured at the output. If N photons are found then the simulation is successful, otherwise the protocol is aborted.

A wealth of operational resource inequalities can be deduced from the above considerations. Here we limit ourselves to the observation that the two-way quantum capacity must satisfy

$$Q_2(\mathcal{N}_{\lambda,\gamma}) \overset{(i)}{\geq} \frac{\lambda^N}{k} Q_2(\Theta_{N,k,\gamma})\,. \tag{6.89}$$

Consequently, it holds that

$$\begin{aligned} Q_2(\mathcal{N}_{\lambda,\gamma}) &\geq \frac{\lambda^N}{k} Q_2(\Theta_{N,k,\gamma}) \\ &\overset{(i)}{\geq} \frac{\lambda^N}{k} I_{\mathrm{coh}}\left(\mathrm{Id} \otimes \Theta_{N,k,\gamma}(|\Psi\rangle\langle\Psi|\,\Psi)\right) \\ &\overset{(iii)}{=} \frac{\lambda^N}{k} \left[\log_2 \binom{N+k-1}{N} - S\left(\binom{N+k-1}{N}^{-1} K_{N,k,\gamma}\right)\right]. \end{aligned} \tag{6.90}$$

Here, in (ii), we used the fact that the two-way quantum capacity of a channel can be lower bounded in terms of the coherent information [33, 34] and we introduced the two-qu*d*it maximally entangled

state $|\Psi\rangle$ of dimension $d = \binom{N+k-1}{N}$. In (iii), we used the definition of coherent information $I_{\text{coh}}(\rho_{AB}) \coloneqq S(\rho_B) - S(\rho_{AB})$, with $S(\cdot)$ being the von Neumann entropy, and the fact that

$$\begin{aligned} \text{Id} \otimes \Theta_{N,k,\gamma}(|\Psi\rangle\langle\Psi| \, \Psi) &= \frac{1}{\binom{N+k-1}{N}} \sum_{p,q \in \Pi(N,k)} |p\rangle\langle p| \, q \otimes \Theta_{N,k,\gamma}(|p\rangle\langle p| \, q) \\ &= \frac{1}{\binom{N+k-1}{N}} \sum_{p,q \in \Pi(N,k)} (K_{N,k,\gamma})_{pq} \, |p\rangle\langle p| \, q \otimes |p\rangle\langle p| \, q \, , \end{aligned} \tag{6.91}$$

which implies that the spectrum of $\text{Id} \otimes \Theta_{N,k,\gamma}(|\Psi\rangle\langle\Psi| \, \Psi)$ coincides with the spectrum of the matrix $\binom{N+k-1}{N}^{-1} K_{N,k,\gamma}$. Consequently, we have that

$$Q_2(\mathcal{N}_{\lambda,\gamma}) \geq \max_{N,k \in \mathbb{N}_+} \frac{\lambda^N}{k} \left[\log_2 \binom{N+k-1}{N} - S\left(\binom{N+k-1}{N}^{-1} K_{N,k,\gamma} \right) \right] . \tag{6.92}$$

One can obtain a lower bound on the energy-constrained two-way quantum capacity $Q_2(\mathcal{N}_{\lambda,\gamma}, N_s)$ by restricting the optimisation to the values of N and k such that $\frac{N}{k} \leq N_s$. Indeed, note that the rail-encoded state $\bar{\rho}$ satisfies the energy constraint as its mean photon number per mode is $\frac{N}{k}$. In formula, we have that

$$Q_2(\mathcal{N}_{\lambda,\gamma}, N_s) \geq \max_{N,k \in \mathbb{N}_+ : \, \frac{N}{k} \leq N_s} \frac{\lambda^N}{k} \left[\log_2 \binom{N+k-1}{N} - S\left(\binom{N+k-1}{N}^{-1} K_{N,k,\gamma} \right) \right] . \tag{6.93}$$

Note that $\log_2 \binom{N+k-1}{N} - S\left(\binom{N+k-1}{N}^{-1} K_{N,k,\gamma} \right)$ is always positive because $\binom{N+k-1}{N}^{-1} K_{N,k,\gamma}$ is a $\binom{N+k-1}{N}$-dimensional, non-maximally mixed, state and thus its von Neumann entropy is strictly smaller than by $\log_2 \binom{N+k-1}{N}$. Consequently, we have the following theorem.

Theorem 12. Let $\gamma \geq 0$ and $\lambda \in (0, 1]$. For any energy constraint $N_s > 0$, the energy-constrained two-way quantum and secret-key capacities of the bosonic loss-dephasing channel are lower bounded by

$$K(\mathcal{N}_{\lambda,\gamma}, N_s) \geq Q_2(\mathcal{N}_{\lambda,\gamma}, N_s) \geq \max_{\substack{N,k \in \mathbb{N}_+ \\ \frac{N}{k} \leq N_s}} \frac{\lambda^N}{k} \left[\log_2 \binom{N+k-1}{N} - S\left(\rho_{N,k,\gamma} \right) \right] > 0 \, . \tag{6.94}$$

Here, $S(\cdot)$ is the von Neumann entropy, $\rho_{N,k,\gamma}$ is a $\binom{N+k-1}{N}$-dimensional state defined by

$$\rho_{N,k,\gamma} \coloneqq \binom{N+k-1}{N}^{-1} \sum_{p,q \in \Pi(N,k)} e^{-\frac{\gamma}{2}\|p-q\|_2^2} \, |p\rangle\langle p| \, q \, , \tag{6.95}$$

where $\Pi(N,k) \coloneqq \left\{ p \in \mathbb{N}^k : \sum_{i=1}^k p_i = N \right\}$ represents the set of partitions of a set of N elements into k parts, and the vectors $\{|p\rangle\}_{p \in \Pi(N,k)}$ are orthonormal. In particular,

$$K(\mathcal{N}_{\lambda,\gamma}, N_s) \geq Q_2(\mathcal{N}_{\lambda,\gamma}, N_s) > \max_{N,k \in \mathbb{N}_+} \frac{\lambda^N}{k} \left[\log_2 \binom{N+k-1}{N} - S\left(\rho_{N,k,\gamma} \right) \right] > 0 \, . \tag{6.96}$$

6.7.5 Technical lemmas

6.7.6 Hadamard maps

Given an infinite matrix $A = (a_{mn})_{m,n \in \mathbb{N}}$, $a_{mn} \in \mathbb{C}$, we can introduce a superoperator H, recognised as the *Hadamard map*, whose action is defined as $H(|m\rangle\langle m| \, n) = a_{n,m} \, |m\rangle\langle m| \, n$ for all $n, m \in \mathbb{N}$.

We are interested in establishing requirements for an infinite matrix A to ensure that the associated Hadamard map H is a quantum channel. We begin with some preliminaries. Let $\ell^2(\mathbb{N})$ be the space of square-summable complex-valued sequences defined as

$$\ell^2(\mathbb{N}) := \left\{ x := \{x_n\}_{n\in\mathbb{N}}, x_n \in \mathbb{C} : \|x\| := \sqrt{\sum_{n=0}^{\infty} |x_n|^2} < \infty \right\} . \tag{6.97}$$

An infinite matrix $A := (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$ defines a linear operator on $\ell^2(\mathbb{N})$. The operator norm of A is defined as follows:

$$\|A\|_\infty := \sup_{\substack{x\in\ell^2(\mathbb{N}) \\ \|x\|=1}} \|Ax\| = \sup_{\substack{\{x_n\}_{n\in\mathbb{N}}, x_n\in\mathbb{C} \\ \sum_{n=0}^{\infty}|x_n|^2=1}} \sqrt{\sum_{m=0}^{\infty} \left| \sum_{n=0}^{\infty} a_{mn} x_n \right|^2} .$$

A is said to be bounded if $\|A\|_\infty < \infty$. The following lemma, referred to as *Schur test*, gives a sufficent condition for an infinite matrix to be bounded (e.g. [72, Page 24, Problem 45]).

Lemma 61. *Let $A := (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$, be an infinite matrix. Suppose that there exist $\{p_n\}_{n\in\mathbb{N}}, p_n \in \mathbb{R}_{>0}$ and $\{q_m\}_{m\in\mathbb{N}}, q_m \in \mathbb{R}_{>0}$, and $\beta > 0$, and $\gamma > 0$ such that*

$$\sum_{m=0}^{\infty} |a_{mn}| p_m \le \beta q_n \quad \textit{and} \quad \sum_{n=0}^{\infty} |a_{mn}| q_n \le \gamma p_m , \quad \forall\, m, n \in \mathbb{N} .$$

Then the matrix A satisfies $\|A\|_\infty \le \beta\gamma$. In particular, A is bounded.

By choosing $p_n = q_n = 1$ and $\gamma = \beta = \sup_{n\in\mathbb{N}} \sum_{m=0}^{\infty} |a_{mn}|$, we obtain the following corollary:

Corollary 62. *Let $A = (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$, be an infinite Hermitian matrix. If $\sup_{n\in\mathbb{N}} \sum_{m=0}^{\infty} |a_{mn}|$ is finite, then A is bounded.*

Lemma 63. *Let $A = (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$, be a bounded Hermitian infinite matrix. Then A is positive semi-definite as an operator on $\ell^2(\mathbb{N})$ if and only if $A^{(d)} := (a_{mn})_{m,n=0,1,\ldots,d-1}$ is positive semi-definite for all $d \in \mathbb{N}$, where $A^{(d)}$ is the $d \times d$ top left corner of A.*

Proof. Assume that $A^{(d)}$ is positive semi-definite for all $d \in \mathbb{N}$. Let us pick an arbitrary $x \in \ell^2(\mathbb{N})$. It is known that for any $\varepsilon > 0$, there exists $d \in \mathbb{N}$ and $y^{(d)} := (y_n^{(d)})_{n\in\mathbb{N}}, y_n^{(d)} \in \mathbb{C}$, with $y_n^{(d)} = 0$ for all $n > d$, such that $\|x - y^{(d)}\| < \varepsilon$. Note that

$$\begin{aligned} x^\dagger A x &= \left(x - y^{(d)}\right)^\dagger A\, x + \left(y^{(d)}\right)^\dagger A\, (x - y^{(d)}) + \left(y^{(d)}\right)^\dagger A y^{(d)} \\ &\overset{(1)}{\ge} -\|x - y^{(d)}\|\, \|A\|_\infty \left(\|x\| + \|y^{(d)}\|\right) + \left(y^{(d)}\right)^\dagger A^{(d)}\, y^{(d)} \\ &\overset{(ii)}{\ge} -\varepsilon\, \|A\|_\infty\, (2\|x\| + \varepsilon) , \end{aligned}$$

where in (i) we applied Cauchy-Schwarz inequality twice and the definition of infinity norm as follows

$$\begin{aligned} |(x - y^{(d)})^\dagger A\, x| &\le \|x - y^{(d)}\| \|A\, x\| \le \|x - y^{(d)}\| \|A\|_\infty \|x\| , \\ |(y^{(d)})^\dagger A\, (x - y^{(d)})| &\le \|y^{(d)}\| \|A\, (x - y^{(d)})\| \le \|y^{(d)}\| \|A\|_\infty \|x - y^{(d)}\| , \end{aligned}$$

and, in (ii), we exploited triangular inequality to derive

$$\|y^{(d)}\| \le \|y^{(d)} - x\| + \|x\| \le \varepsilon + \|x\| , \tag{6.98}$$

together with the fact that $A^{(d)}$ is positive semi-definite. Hence, since $\varepsilon > 0$ is arbitrary, we conclude that $x^\dagger A x \ge 0$, meaning that A is positive semi-definite as an operator on $\ell^2(\mathbb{N})$. □

A square matrix is said to be diagonally dominant if

$$\sum_{m\neq n} |a_{mn}| \leq |a_{nn}|, \quad \forall n.$$

In words, a square matrix is said to be diagonally dominant if for every row of the matrix, the absolute value of the diagonal entry in a row is larger than or equal to the sum of the absolute values of all the other (non-diagonal) entries in that row. Note that for Hermitian matrices one can exchange row with column in this definition. We proceed to state a sufficient condition for a matrix A to be positive semi-definite as an operator on $\ell^2(\mathbb{N})$. While the following sufficient condition is usually stated for finite matrices, it can be generalised to infinite matrices as per Lemma 63.

Lemma 64. *[43, Chapter 6] For any $d \in \mathbb{N}$, let $A = (a_{mn})_{m,n=0,1,\ldots,d-1}, a_{mn} \in \mathbb{C}$, be a $d \times d$ Hermitian matrix with $a_{nn} \in \mathbb{R}_{\geq 0}$ for all $n \in \{0, \ldots, d-1\}$. A is positive semi-definite if it is diagonally dominant.*

Lemma 65. *Let $A = (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$ be an infinite Hermitian matrix with $a_{nn} \in \mathbb{R}_{\geq 0}, \forall n \in \mathbb{N}$. Assume that $\sup_{n\in\mathbb{N}} a_{nn}$ is finite and that A is diagonally dominant. Then A is bounded and positive semi-definite when seen as an operator on $\ell^2(\mathbb{N})$.*

Proof. Since

$$\sup_{n\in\mathbb{N}} \sum_{m=0}^{\infty} |a_{mn}| \leq 2 \sup_{n\in\mathbb{N}} a_{n,n} < \infty\,,$$

Corollary 62 implies that A is bounded. The fact that A is positive semi-definite as an operator on $\ell^2(\mathbb{N})$ follows from Lemma 64 together with Lemma 63. □

We now present a lemma from the literature that establishes necessary and sufficient conditions for an infinite matrix A to ensure that the associated Hadamard map H is a quantum channel. We then use it to derive an explicit sufficient condition for an infinite matrix A to give rise to a cptp Hadamard map.

Lemma 66. *[29, Lemma S4] Let $A := (a_{mn})_{m,n\in\mathbb{N}}, a_{mn} \in \mathbb{C}$ be a bounded infinite matrix. The following requirements establish the necessary and sufficient conditions for the associated Hadamard map H to qualify as a quantum channel:*

(i) $a_{nn} = 1$, $\forall n \in \mathbb{N}$;

(ii) A is positive semi-definite as on operator on $\ell^2(\mathbb{N})$.

As a consequence of Lemma 66 and Lemma 65, we obtain:

Lemma 67. *Let $A := (a_{mn})_{m,n\in\mathbb{N}}$ be an infinite Hermitian matrix that is diagonally dominant with $a_{n,n} = 1$ for all $n \in \mathbb{N}$. In this case, its associated Hadamard map H is a quantum channel.*

Miscellaneous Lemmas

Lemma 68. *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert spaces and let $\mathcal{N} : \mathcal{T}(\mathcal{H}_A) \to \mathcal{T}(\mathcal{H}_B)$ be a quantum channel. For all normalised states $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}_A$ and $|\phi_1\rangle, |\phi_2\rangle \in \mathcal{H}_B$ it holds that*

$$|\langle\phi_1| \mathcal{N}(|\psi_1\rangle\langle\psi_1| \psi_2) |\phi_2\rangle| \leq 1\,. \tag{6.99}$$

Proof. It holds that

$$\begin{aligned} |\langle\phi_1|\,\mathcal{N}(|\psi_1\rangle\langle\psi_1|\,\psi_2)\,|\phi_2\rangle| &\overset{(i)}{\leq} \|\mathcal{N}(|\psi_1\rangle\langle\psi_1|\,\psi_2)\|_\infty \\ &\overset{(ii)}{\leq} \|\mathcal{N}(|\psi_1\rangle\langle\psi_1|\,\psi_2)\|_1 \\ &\overset{(iii)}{\leq} \|\,|\psi_1\rangle\langle\psi_1|\,\psi_2\|_1 \\ &= 1\,. \end{aligned} \tag{6.100}$$

Here, in (i), we exploited one of the definition of the operator norm in (6.11). In (ii), we exploited that the trace norm is always an upper bound on the operator norm. Finally, in (iii), we leveraged the monotonicity of the trace norm under quantum channels [59]. □

Lemma 69 [61, 62]**.** *Let $\mathcal{H}_A, \mathcal{H}_{A'}$ be isomorphic Hilbert spaces, possibly infinite dimensional. Let $|\psi\rangle_{A'A}$ be a pure state that satisfies $\mathrm{Tr}_{A'}[|\psi\rangle\langle\psi|_{AA'}] > 0$. The generalised Choi–Jamiołkowski matrix defines an isomorphism between the set of quantum channels from $\mathcal{H}_{A'}$ to $\mathcal{H}_B$ and the set of bipartite states $\sigma_{AB} \in \mathcal{P}(\mathcal{H}_{AB})$ such that $\mathrm{Tr}_B\,\sigma_{AB} = \mathrm{Tr}_{A'}[|\psi\rangle\langle\psi|_{AA'}]$. Specifically, for any quantum channel $\mathcal{N}_{A'\to B} : \mathcal{T}(\mathcal{H}_{A'}) \to \mathcal{T}(\mathcal{H}_B)$, it holds that*

$$\mathcal{N}_{A'\to B}(|e_i\rangle\langle e_i|\,e_j) = \frac{1}{\sqrt{\lambda_i\lambda_j}}\,\mathrm{Tr}_A[(|e_j\rangle\langle e_j|\,e_{iA}\otimes\mathbb{1}_B)\;\sigma_{AB}]\,, \quad \forall\, i,j\in\mathbb{N}\,, \tag{6.101}$$

where $(|e_i\rangle)_{i\in\mathbb{N}}$ and $(\lambda_i)_{i\in\mathbb{N}}$ form a spectral decomposition of $\mathrm{Tr}_{A'}[|\psi\rangle\langle\psi|_{AA'}]$, i.e. $\mathrm{Tr}_{A'}[|\psi\rangle\langle\psi|_{AA'}] = \sum_i \lambda_i\,|e_i\rangle\langle e_i|_A$, and where the state $\sigma_{AB} \coloneqq \mathrm{Id}_A\otimes\mathcal{N}_{A'\to B}(|\psi\rangle\langle\psi|_{AA'})$ is called the generalised Choi state of $\mathcal{N}$. Eq. 6.101 is enough to specify the channel $\mathcal{N}_{A'\to B}$ completely, as the linear span of the operators $(|e_i\rangle\langle e_i|\,e_j)_{i,j\in\mathbb{N}}$ (i.e. the set of finite-rank operators) is dense in $\mathcal{T}(\mathcal{H}_{A'})$.

Lemma 70 [41]**.** *Let $\mathcal{H}_A, \mathcal{H}_{A'}, \mathcal{H}_B$ be isomorphic Hilbert spaces, possibly infinite dimensional. Let $\mathcal{N}_{A'\to B} : \mathcal{T}(\mathcal{H}_{A'}) \to \mathcal{T}(\mathcal{H}_B)$ be a quantum channel. Let $|\psi\rangle_{A'A} \in \mathcal{H}_A\otimes\mathcal{H}_{A'}$ be a pure state such that the reduced state $\mathrm{Tr}_{A'}[|\psi\rangle\langle\psi|_{AA'}]$ is positive definite. Then, $\mathcal{N}_{A'\to B}$ is anti-degradable if and only if the state $\mathrm{Id}_A\otimes\mathcal{N}_{A'\to B}(|\psi\rangle\langle\psi|_{AA'})$ is two-extendible on B, meaning that there exists a state $\rho_{AB_1B_2} \in \mathcal{P}(\mathcal{H}_A\otimes\mathcal{H}_{B_1}\otimes\mathcal{H}_{B_2})$ such that*

$$\begin{aligned} \mathrm{Tr}_{B_2}[\rho_{AB_1B_2}] &= \mathrm{Id}_A\otimes\mathcal{N}_{A'\to B_1}(|\psi\rangle\langle\psi|_{AA'})\,, \\ \mathrm{Tr}_{B_1}[\rho_{AB_1B_2}] &= \mathrm{Id}_A\otimes\mathcal{N}_{A'\to B_2}(|\psi\rangle\langle\psi|_{AA'}), \end{aligned} \tag{6.102}$$

where $\mathcal{H}_{B_1}$ and $\mathcal{H}_{B_2}$ are Hilbert spaces that are isomorphic to $\mathcal{H}_B$.

Proof. Let $U_{A'E\to BE}$ be a Stinespring dilation of the channel $\mathcal{N}_{A'\to B}$. Further assume that $\mathcal{N}_{A'\to B}$ is anti-degradable. By definition, there exists a quantum channel $\mathcal{A}_{E\to B}$ such that $\mathcal{A}_{E\to B}\circ\mathcal{N}^{\mathrm{c}}_{A'\to B} = \mathcal{N}_{A'\to B}$. Let us consider the tripartite state $\rho_{AB_1B_2}$, with B_1, B_2 being copies of B, defined as

$$\rho_{AB_1B_2} = \mathrm{Id}_A\otimes\mathrm{Id}_{B_1}\otimes\mathcal{A}_{E\to B_2}\left(U_{A'E\to B_1E}\left(|\psi\rangle\langle\psi|_{AA'}\otimes|0\rangle\langle 0|_E\right)U^\dagger_{A'E\to B_1E}\right). \tag{6.103}$$

It holds that

$$\begin{aligned} \mathrm{Tr}_{B_2}[\rho_{AB_1B_2}] &= \mathrm{Tr}_{B_2}\left[\mathrm{Id}_A\otimes\mathrm{Id}_{B_1}\otimes\mathcal{A}_{E\to B_2}\left(U_{A'E\to B_1E}\left(|\psi\rangle\langle\psi|_{AA'}\otimes|0\rangle\langle 0|_E\right)U^\dagger_{A'E\to B_1E}\right)\right] \\ &= \mathrm{Id}_A\otimes\mathrm{Tr}_E\left[U_{A'E\to B_1E}\left(|\psi\rangle\langle\psi|_{AA'}\otimes|0\rangle\langle 0|_E\right)U^\dagger_{A'E\to B_1E}\right] \\ &= \mathrm{Id}_A\otimes\mathcal{N}_{A'\to B_1}(|\psi\rangle\langle\psi|_{AA'})\,, \end{aligned}$$

and

$$\begin{aligned} \mathrm{Tr}_{B_1}[\rho_{AB_1B_2}] &= \mathrm{Tr}_{B_1}\left[\mathrm{Id}_A\otimes\mathrm{Id}_{B_1}\otimes\mathcal{A}_{E\to B_2}\left(U_{A'E\to B_1E}\left(|\psi\rangle\langle\psi|_{AA'}\otimes|0\rangle\langle 0|_E\right)U^\dagger_{A'E\to B_1E}\right)\right] \\ &= \mathrm{Id}_A\otimes\mathcal{A}_{E\to B_2}\left(\mathrm{Tr}_{B_1}\left[U_{A'E\to B_1E}\left(|\psi\rangle\langle\psi|_{AA'}\otimes|0\rangle\langle 0|_E\right)U^\dagger_{A'E\to B_1E}\right]\right) \\ &= \mathrm{Id}_A\otimes\mathcal{A}_{E\to B_2}\circ\mathcal{N}^{\mathrm{c}}_{A'\to B_2}(|\psi\rangle\langle\psi|_{AA'}) \\ &= \mathrm{Id}_A\otimes\mathcal{N}_{A'\to B_2}(|\psi\rangle\langle\psi|_{AA'})\,. \end{aligned}$$

Now, let us establish the converse. Assume that there exists $\rho_{AB_1B_2}$ which satisfies (6.102). Let $|\Psi\rangle_{AB_1B_2P} \in \mathcal{H}_A \otimes \mathcal{H}_{B_1} \otimes \mathcal{H}_{B_2} \otimes \mathcal{H}_P$ be a purification of $\rho_{AB_1B_2}$, with $\mathcal{H}_P$ being the purifying Hilbert space. Note that both $|\Psi\rangle_{AB_1B_2P}$ and $U_{A'E\to B_1E}\left(|\psi\rangle_{AA'} \otimes |0\rangle_E\right)$ are purifications of $\mathrm{Id}_A \otimes \mathcal{N}_{A'\to B_1}(|\psi\rangle\langle\psi|_{AA'})$, with $\mathcal{H}_{B_2} \otimes \mathcal{H}_P$ and $\mathcal{H}_E$ being their purifying Hilbert spaces, respectively. It follows that [59] there exists an isometry $V_{E\to B_2P} : \mathcal{H}_E \to \mathcal{H}_{B_2} \otimes \mathcal{H}_P$ such that $V_{E\to B_2P}\, U_{A'E\to B_1E}\left(|\psi\rangle_{AA'} \otimes |0\rangle_E\right) = |\Psi\rangle_{AB_1B_2P}$. Hence, the quantum channel $\mathcal{A}_{E\to B_2} : \mathcal{T}(\mathcal{H}_E) \to \mathcal{T}(\mathcal{H}_{B_2})$, defined by $\mathcal{A}_{E\to B_2}(\cdot) = \mathrm{Tr}_P\left[V_{E\to B_2P}(\cdot)V^\dagger_{E\to B_2P}\right]$, satisfies that

$$\begin{aligned}
\mathrm{Id}_A \otimes \mathcal{A}_{E\to B_2} \circ \mathcal{N}^{\mathrm{c}}_{A'\to B_2}(|\psi\rangle\langle\psi|_{AA'}) &= \mathrm{Id}_A \otimes \mathcal{A}_{E\to B_2}\left(\mathrm{Tr}_{B_1}\left[U_{A'E\to B_1E}(|\psi\rangle\langle\psi|_{AA'} \otimes |0\rangle\langle 0|_E)U^\dagger_{A'E\to B_1E}\right]\right) \\
&= \mathrm{Tr}_{B_1P}\left[|\Psi\rangle\langle\Psi|_{AB_1B_2P}\right] \\
&= \mathrm{Tr}_{B_1}\left[\rho_{AB_1B_2}\right] \\
&= \mathrm{Id}_A \otimes \mathcal{N}_{A'\to B_2}\left(|\psi\rangle\langle\psi|_{AA'}\right) .
\end{aligned}$$

Consequently, since the pure state $|\psi\rangle_{A'A}$ satisfies $\mathrm{Tr}_{A'}[|\psi\rangle\langle\psi|_{AA'}] > 0$, Lemma 69 implies that $\mathcal{A}_{E\to B_2} \circ \mathcal{N}^{\mathrm{c}}_{A'\to B_2} = \mathcal{N}_{A'\to B_2}$, meaning that $\mathcal{N}_{A'\to B_2}$ is anti-degradable. □

Lemma 71. *Let $\mathcal{N}, \mathcal{M} : \mathcal{T}(\mathcal{H}_S) \to \mathcal{T}(\mathcal{H}_S)$ be quantum channels. If either $\mathcal{M}$ or $\mathcal{N}$ is anti-degradable, then the composition $\mathcal{M} \circ \mathcal{N}$ is anti-degradable. Specifically, let E_1 and E_2 be the Stinespring environments of $\mathcal{N}$ and $\mathcal{M}$, respectively. If $\mathcal{N}$ is anti-degradable with anti-degrading map $\mathcal{A}_{E_1\to S}$, then $(\mathcal{M} \circ \mathcal{A}_{E_1\to S}) \otimes \mathrm{Tr}_{E_2}$ is an anti-degrading map of $\mathcal{M} \circ \mathcal{N}$. Analogously, if $\mathcal{M}$ is anti-degradable with anti-degrading map $\mathcal{A}_{E_2\to S}$, then $\mathrm{Tr}_{E_1} \otimes \mathcal{A}_{E_2\to S}$ is an anti-degrading map of $\mathcal{M} \circ \mathcal{N}$.*

Proof. Let $V^{S\to SE_1}$ and $W^{S\to SE_2}$ be Stinespring isometries associated with $\mathcal{N}$ and $\mathcal{M}$, respectively. By considering the following complementary channel of $\mathcal{M} \circ \mathcal{N}$,

$$(\mathcal{M} \circ \mathcal{N})^{\mathrm{c}}(\rho) = \mathrm{Tr}_S\left[W^{S\to SE_2}V^{S\to SE_1}\,\rho\,\left(V^{S\to SE_1}\right)^\dagger\left(W^{S\to SE_2}\right)^\dagger\right], \qquad \forall\,\rho \in \mathcal{T}(\mathcal{H}_S)\,,$$

one can easily check that if $\mathcal{N}$ is anti-degradable with anti-degrading map $\mathcal{A}_{E_1\to S}$, then

$$[(\mathcal{M} \circ \mathcal{A}_{E_1\to S}) \otimes \mathrm{Tr}_{E_2}] \circ (\mathcal{M} \circ \mathcal{N})^{\mathrm{c}} = \mathcal{M} \circ \mathcal{N}\,. \tag{6.104}$$

Analogously, one can easily verify that if $\mathcal{M}$ is anti-degradable with anti-degrading map $\mathcal{A}_{E_2\to S}$, then

$$[\mathrm{Tr}_{E_1} \otimes \mathcal{A}_{E_2\to S}] \circ (\mathcal{M} \circ \mathcal{N})^{\mathrm{c}} = \mathcal{M} \circ \mathcal{N}\,. \tag{6.105}$$

□

Bibliography

[1] P. Leviant, Q. Xu, L. Jiang, and S. Rosenblum. Quantum capacity and codes for the bosonic loss-dephasing channel. *Quantum*, 6:821, sep 2022. doi: 10.22331/q-2022-09-29-821. URL https://doi.org/10.22331/q-2022-09-29-821.

[2] M. Hayashi. *Bosonic System and Quantum Optics*, pages 231–262. Springer International Publishing, Cham, 2017. ISBN 978-3-319-44906-7. doi: 10.1007/978-3-319-44906-7_7. URL https://doi.org/10.1007/978-3-319-44906-7_7.

[3] A. Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, Taylor & Francis Group, Boca Raton, USA, 2017.

[4] C. M. Caves and P. D. Drummond. Quantum limits on bosonic communication rates. *Reviews of Modern Physics*, 66:481–537, 1994. doi: 10.1103/RevModPhys.66.481. URL https://link.aps.org/doi/10.1103/RevModPhys.66.481.

[5] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84:621–669, 2012.

[6] S. Yokoyama, R. Ukai, S. Armstrong, and et al. Ultra-large-scale continuous-variable cluster states multiplexed in the time domain. *Nat. Photo*, 7:982–986, 2013. doi: https://doi.org/10.1038/nphoton.2013.287.

[7] M. H. Devoret and R. J. Schoelkopf. Superconducting circuits for quantum information: An outlook. *Science*, 339(6124):1169–1174, 2013. doi: 10.1126/science.1231930. URL https://www.science.org/doi/abs/10.1126/science.1231930.

[8] Z. Leghtas, G. Kirchmair, B. Vlastakis, R. J. Schoelkopf, M. H. Devoret, and M. Mirrahimi. Hardware-efficient autonomous quantum memory protection. *Phys. Rev. Lett.*, 111:120501, Sep 2013. doi: 10.1103/PhysRevLett.111.120501. URL https://link.aps.org/doi/10.1103/PhysRevLett.111.120501.

[9] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature*, 536(7617):441–445, 2016.

[10] X. B. Wang, T. Hiroshima, A. Tomita, and M. Hayashi. Quantum information with gaussian states. *Physics Reports*, 448(1):1–111, 2007. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2007.04.005. URL https://www.sciencedirect.com/science/article/pii/S0370157307001822.

[11] M. Schlosshauer. Decoherence, the measurement problem, and interpretations of quantum mechanics. *Rev. Mod. Phys.*, 76:1267–1305, Feb 2005. doi: 10.1103/RevModPhys.76.1267. URL https://link.aps.org/doi/10.1103/RevModPhys.76.1267.

[12] C. Vuillot, H. Asasi, Y. Wang, L. P. Pryadko, and B. M. Terhal. Quantum error correction with the toric gottesman-kitaev-preskill code. *Phys. Rev. A*, 99:032344, Mar 2019. doi: 10.1103/PhysRevA.99.032344. URL https://link.aps.org/doi/10.1103/PhysRevA.99.032344.

[13] B. M. Terhal, J. Conrad, and C. Vuillot. Towards scalable bosonic quantum error correction. *Quantum Science and Technology*, 5(4):043001, jul 2020. doi: 10.1088/2058-9565/ab98a5. URL https://dx.doi.org/10.1088/2058-9565/ab98a5.

[14] V. V. Albert, K. Noh, K. Duivenvoorden, D. J. Young, R. T. Brierley, P. Reinhold, C. Vuillot, L. Li, C. Shen, S. M. Girvin, B. M. Terhal, and L. Jiang. Performance and structure of single-mode bosonic codes. *Phys. Rev. A*, 97:032346, Mar 2018. doi: 10.1103/PhysRevA.97.032346. URL https://link.aps.org/doi/10.1103/PhysRevA.97.032346.

[15] A. Denys and A. Leverrier. The 2T-qutrit, a two-mode bosonic qutrit. *Quantum*, 7:1032, June 2023. ISSN 2521-327X. doi: 10.22331/q-2023-06-05-1032. URL https://doi.org/10.22331/q-2023-06-05-1032.

[16] Arne L. Grimsmo, Joshua Combes, and Ben Q. Baragiola. Quantum computing with rotation-symmetric bosonic codes. *Phys. Rev. X*, 10:011058, Mar 2020. doi: 10.1103/PhysRevX.10.011058. URL https://link.aps.org/doi/10.1103/PhysRevX.10.011058.

[17] K. Noh and C. Chamberland. Fault-tolerant bosonic quantum error correction with the surface–gottesman-kitaev-preskill code. *Phys. Rev. A*, 101:012316, Jan 2020. doi: 10.1103/PhysRevA.101.012316. URL https://link.aps.org/doi/10.1103/PhysRevA.101.012316.

[18] D. S. Schlegel, F. Minganti, and V. Savona. Quantum error correction using squeezed schrödinger cat states. *Phys. Rev. A*, 106:022431, Aug 2022. doi: 10.1103/PhysRevA.106.022431. URL https://link.aps.org/doi/10.1103/PhysRevA.106.022431.

[19] T. Hillmann and F. Quijandría. Quantum error correction with dissipatively stabilized squeezed-cat qubits. *Phys. Rev. A*, 107:032423, Mar 2023. doi: 10.1103/PhysRevA.107.032423. URL https://link.aps.org/doi/10.1103/PhysRevA.107.032423.

[20] F. A. Mele, L. Lami, and V. Giovannetti. Maximum tolerable excess noise in CV-QKD and improved lower bound on two-way capacities. *Preprint arXiv:2303.12867*, 2023. doi: 10.48550/arXiv.2303.12867.

[21] F. A. Mele, L. Lami, and V. Giovannetti. Restoring quantum communication efficiency over high loss optical fibers. *Physical Review Letters*, 129:180501, 2022. doi: 10.1103/PhysRevLett.129.180501. URL https://link.aps.org/doi/10.1103/PhysRevLett.129.180501.

[22] F. A. Mele, L. Lami, and V. Giovannetti. Quantum optical communication in the presence of strong attenuation noise. *Physical Review A*, 106:042437, 2022. doi: 10.1103/PhysRevA.106.042437. URL https://link.aps.org/doi/10.1103/PhysRevA.106.042437.

[23] F. A. Mele, G. De Palma, M. Fanizza, V. Giovannetti, and L. Lami. Optical fibres with memory effects and their quantum communication capacities, 2023.

[24] M. H. Michael, M. Silveri, R. T. Brierley, V. V. Albert, J. Salmilehto, L. Jiang, and S. M. Girvin. New class of quantum error-correcting codes for a bosonic mode. *Phys. Rev. X*, 6:031006, Jul 2016. doi: 10.1103/PhysRevX.6.031006. URL https://link.aps.org/doi/10.1103/PhysRevX.6.031006.

[25] F. Brito, D. P. DiVincenzo, R. H. Koch, and M. Steffen. Efficient one- and two-qubit pulsed gates for an oscillator-stabilized josephson qubit. *New Journal of Physics*, 10(3): 033027, mar 2008. doi: 10.1088/1367-2630/10/3/033027. URL https://dx.doi.org/10.1088/1367-2630/10/3/033027.

[26] K. H. Wanser. Fundamental phase noise limit in optical fibres due to temperature fluctuations. *Electronics Letters*, 28:53, January 1992. doi: 10.1049/el:19920033.

[27] K. Noh, V. V. Albert, and L. Jiang. Quantum capacity bounds of gaussian thermal loss channels and achievable rates with gottesman-kitaev-preskill codes. *IEEE Transactions on Information Theory*, 65(4):2563–2582, 2019. doi: 10.1109/TIT.2018.2873764.

[28] M. M. Wilde and A. Winter. Strong converse for the classical capacity of the pure-loss bosonic channel. *Problems of Information Transmission*, 50(2):117–132, April 2014.

[29] L. Lami and M. M. Wilde. Exact solution for the quantum and private capacities of bosonic dephasing channels. *Nat. Photonics*, 17(6):525–530, 2023. doi: 10.1038/s41566-023-01190-4.

[30] P. Campagne-Ibarcq, A. Eickbusch, S. Touzard, E. Zalys-Geller, N. E. Frattini, V. V. Sivak, P. Reinhold, S. Puri, S. Shankar, R. J. Schoelkopf, L. Frunzio, M. Mirrahimi, and M. H. Devoret. Quantum error correction of a qubit encoded in grid states of an oscillator. *Nature*, 584(7821):368–372, August 2020.

[31] M. Reagor, W. Pfaff, C. Axline, R. W. Heeres, N. Ofek, K. Sliwa, E. Holland, C. Wang, J. Blumoff, K. Chou, M. J. Hatridge, L. Frunzio, M. H. Devoret, L. Jiang, and R. J. Schoelkopf. Quantum memory with millisecond coherence in circuit qed. *Phys. Rev. B*, 94:014506, Jul 2016. doi: 10.1103/PhysRevB.94.014506. URL https://link.aps.org/doi/10.1103/PhysRevB.94.014506.

[32] S. Rosenblum, P. Reinhold, M. Mirrahimi, L. Jiang, L. Frunzio, and R. J. Schoelkopf. Fault-tolerant detection of a quantum error. *Science*, 361(6399):266–270, July 2018. ISSN 1095-9203. doi: 10.1126/science.aat3996. URL http://dx.doi.org/10.1126/science.aat3996.

[33] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2nd edition, 2017. ISBN 9781316813300.

[34] S. Khatri and M. M. Wilde. Principles of quantum communication theory: A modern approach, 2020. URL https://arxiv.org/abs/2011.04672.

[35] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Commun. Math. Phys.*, 256(2):287–303, 2005. doi: 10.1007/s00220-005-1317-6.

[36] F. Caruso and V. Giovannetti. Degradability of bosonic gaussian channels. *Physical Review A*, 74:062307, 2006. doi: 10.1103/PhysRevA.74.062307. URL https://link.aps.org/doi/10.1103/PhysRevA.74.062307.

[37] V. Giovannetti, S. Lloyd, L. Maccone, and P. W. Shor. Entanglement assisted capacity of the broadband lossy channel. *Physical Review Letters*, 91:047901, 2003. doi: 10.1103/PhysRevLett.91.047901. URL https://link.aps.org/doi/10.1103/PhysRevLett.91.047901.

[38] F. Caruso, V. Giovannetti, and A. S. Holevo. One-mode bosonic Gaussian channels: a full weak-degradability classification. *New Journal of Physics*, 8(12):310–310, 2006. doi: 10.1088/1367-2630/8/12/310.

[39] M. M. Wolf, D. Pérez-García, and G. Giedke. Quantum capacities of bosonic channels. *Physical Review Letters*, 98:130501, 2007. doi: 10.1103/PhysRevLett.98.130501.

[40] A. Arqand, L. Memarzadeh, and S. Mancini. Quantum capacity of a bosonic dephasing channel. *Phys. Rev. A*, 102:042413, Oct 2020. doi: 10.1103/PhysRevA.102.042413. URL https://link.aps.org/doi/10.1103/PhysRevA.102.042413.

[41] G. O. Myhr and N. Lütkenhaus. Spectrum conditions for symmetric extendible states. *Physical Review A*, 79:062307, 2009. doi: 10.1103/PhysRevA.79.062307.

[42] L. Lami, S. Khatri, G. Adesso, and M. M. Wilde. Extendibility of bosonic Gaussian states. *Physical Review Letters*, 123:050501, 2019. doi: 10.1103/PhysRevLett.123.050501.

[43] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[44] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. doi: 10.1017/9781316848142.

[45] C. Paddock and J. Chen. A characterization of antidegradable qubit channels, 2017.

[46] J. Chen, Z. Ji, D. Kribs, N. Lutkenhaus, and B. Zeng. Symmetric extension of two-qubit states. *Physical Review A*, 90(3), sep 2014. doi: 10.1103/physreva.90.032318. URL https://doi.org/10.11032Fphysreva.90.032318.

[47] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67: 661–663, 1991.

[48] N. Davis, M. E. Shirokov, and M. M. Wilde. Energy-constrained two-way assisted private and quantum capacities of quantum channels. *Physical Review A*, 97:062310, 2018. doi: 10.1103/PhysRevA.97.062310.

[49] M. M. Wilde and H. Qi. Energy-constrained private and quantum capacities of quantum channels. *IEEE Transactions on Information Theory*, 64(12):7802–7827, 2018. doi: 10.1109/TIT.2018.2854766.

[50] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. Royal Soc. A*, 461(2053):207–235, 2005.

[51] T. S. Cubitt, M. B. Ruskai, and G. Smith. The structure of degradable quantum channels. *J. Math. Phys.*, 49(10):102104, 2008. doi: 10.1063/1.2953685.

[52] Graeme Smith and John A. Smolin. Detecting incapacity of a quantum channel. *Phys. Rev. Lett.*, 108:230507, Jun 2012. doi: 10.1103/PhysRevLett.108.230507. URL https://link.aps.org/doi/10.1103/PhysRevLett.108.230507.

[53] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi. Fundamental limits of repeaterless quantum communications. *Nature Communications*, 8(1):15043, 2017. doi: 10.1038/ncomms15043.

[54] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54:3824–3851, 1996. doi: 10.1103/PhysRevA.54.3824.

[55] A. S. Holevo and R. F. Werner. Evaluating capacities of bosonic Gaussian channels. *Physical Review A*, 63:032312, 2001. doi: 10.1103/PhysRevA.63.032312.

[56] M. M. Wilde, P. Hayden, and S. Guha. Quantum trade-off coding for bosonic communication. *Physical Review A*, 86:062306, 2012. doi: 10.1103/PhysRevA.86.062306.

[57] M. M. Wilde and H. Qi. Energy-constrained private and quantum capacities of quantum channels. *IEEE Transactions on Information Theory*, 64(12):7802–7827, 2018. ISSN 1557-9654. doi: 10.1109/TIT.2018.2854766.

[58] K. Noh, V. V. Albert, and L. Jiang. Quantum capacity bounds of Gaussian thermal loss channels and achievable rates with Gottesman-Kitaev-Preskill codes. *IEEE Transactions on Information Theory*, 65(4):2563–2582, 2019. ISSN 1557-9654. doi: 10.1109/TIT.2018.2873764.

[59] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010.

[60] M. L. Nowakowski. The symmetric extendibility of quantum states. *Journal of Physics A: Mathematical and Theoretical*, 49(38):385301, aug 2016. doi: 10.1088/1751-8113/49/38/385301. URL https://dx.doi.org/10.1088/1751-8113/49/38/385301.

[61] A. S. Holevo. The Choi–Jamiolkowski forms of quantum Gaussian channels. *J. Math. Phys.*, 52(4):042202, 2011. doi: 10.1063/1.3581879.

[62] A. S. Holevo. On the Choi–Jamiolkowski correspondence in infinite dimensions. *Preprint arXiv:1004.0196*, 2010.

[63] Z. Baghali Khanian and A. Winter. Distributed compression of correlated classical-quantum sources or: The price of ignorance. *IEEE Transactions on Information Theory*, 66(9): 5620–5633, 2020. doi: 10.1109/TIT.2020.2981322.

[64] M. Rexiti, L. Memarzadeh, and S. Mancini. Discrimination of dephasing channels. *Journal of Physics A: Mathematical and Theoretical*, 55(24):245301, may 2022. doi: 10.1088/1751-8121/ac6d2c. URL https://dx.doi.org/10.1088/1751-8121/ac6d2c.

[65] S. Dehdashti, J. Nötzel, and P. V. Loock. Quantum capacity of a deformed bosonic dephasing channel, 2022.

[66] H. Dür and H. J. Briegel. Entanglement purification and quantum error correction. *Reports on Progress in Physics*, 70(8):1381–1424, 2007. doi: 10.1088/0034-4885/70/8/r03. URL https://doi.org/10.1088/0034-4885/70/8/r03.

[67] F. Salek and A. Winter. New protocols for conference key and multipartite entanglement distillation. *Preprint arXiv:2308.01134*, 2023. doi: 10.48550/arXiv.2308.0113.

[68] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, 1993.

[69] A. S. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction*. Texts and Monographs in Theoretical Physics. De Gruyter, Berlin, Germany, 2nd edition, 2019. ISBN 9783110642407.

[70] M. Horodecki, P. Horodecki, and R. Horodecki. Inseparable two spin-$\frac{1}{2}$ density matrices can be distilled to a singlet form. *Physical Review Letters*, 78:574–577, 1997. doi: 10.1103/PhysRevLett.78.574.

[71] L. Lami, K. K. Sabapathy, and A. Winter. All phase-space linear bosonic channels are approximately Gaussian dilatable. *New Journal of Physics*, 20(11):113012, 2018.

[72] P. R. Halmos. *A Hilbert Space Problem Book*. Graduate Texts in Mathematics. Springer New York, NY, 1982.

Chapter 7

Optical fibres with memory effects and their quantum communication capacities

The development of quantum repeaters remains challenging, both technologically and in terms of cost and long-term maintenance. This motivates a careful study of what can be achieved *without* repeaters. In the standard memoryless (i.i.d.) model of optical fibres, it is known that fundamental quantum communication tasks become impossible once the fibre transmissivity drops below a critical value, imposing a stringent limit on the achievable distance. However, the memoryless approximation is not always justified: when successive signals are sent sufficiently close in time, the noise can acquire temporal correlations, and the implications of the memoryless bound must be re-examined.

In this chapter, we introduce a model of optical fibre that captures such memory effects over long transmission lines. We then characterise its communication capabilities by solving exactly its quantum capacity, two-way quantum capacity, and secret-key capacity. The resulting picture is qualitatively different from the memoryless one: owing to memory-induced cross-talk between consecutive signals, reliable quantum communication can become possible even in highly noisy regimes where it would be ruled out under the i.i.d. assumption. As part of our analysis, we identify a critical time separation between subsequent signals below which quantum communication, two-way entanglement distribution, and quantum key distribution become achievable.

[This chapter is based on F. A. Mele, G. De Palma, M. Fanizza, V. Giovannetti, L. Lami, Optical fibres with memory effects and their quantum communication capacities, IEEE Transactions on Information Theory 70(12), 8844–8869 (2024)]

7.1 Introduction

Quantum information [1], and in particular quantum communication, will likely play a pivotal role in our future technology. The potential applications of a global quantum internet [2, 3] include secure communication [4], efficient entanglement and qubits distribution, enhanced quantum sensing capabilities [5], distributed and blind quantum computing [6, 7], as well as groundbreaking experiments in fundamental physics [5]. These applications heavily rely on establishing long-distance quantum communication across optical fibres or free-space links. However, the vulnerability of optical signals to noise poses a significant obstacle to achieve this goal. To overcome this challenge, one possible known solution is to exploit quantum repeaters [8, 9] along the communication line. Nonetheless, current implementations of quantum repeaters remain in the realm of proof-of-principle experiments. In addition, quantum repeaters will likely impose substantial demands on technology resources, making them potentially expensive to implement. Consequently, a pressing problem is to develop quantum communication protocols that can operate without — or with a modest number of — quantum repeaters. Recently, in [10, 11] it has been proposed a theoretical proof-of-principle solution to the problem of establishing quantum communication

over long distance without relying on quantum repeaters. The crux of such a solution is to take advantage of *memory effects* [12, 13] in optical fibres.

Memory effects arise when signals are fed into the optical fibre separated by a sufficiently short time interval [12–15]. In this scenario, the noise within the fibre is influenced by prior input signals, exhibiting a form of "memory." Consequently, the commonly assumed memoryless (iid) paradigm, which posits that noise acts uniformly and independently on each signal, becomes invalid. The exploration of memory effects in optical fibres has been extensively addressed in [10, 11, 16–18]. These studies primarily focus on modelling the interaction between two signals through a single localised interaction. While they capture fundamental aspects of the problem, especially in the context of "short" communication lines or setups with signal cross-talk confined in a specific spatial region, a more comprehensive analysis reveals limitations when extending these findings to practical configurations. Specifically, in scenarios involving "long" communication lines where consecutive signals continually interact throughout the entire length of the fibre, the existing analyses become inadequate, hindering the incorporation of two essential requirements vital for the theory's conceptual self-consistency:

- *Property 1*: It is imperative that no information can be transmitted across the fibre if the transmissivity associated with a single signal is precisely zero;
- *Property 2*: The model must remain consistent when optical fibres are composed. In other words, the combination of the model associated with a fibre of length L_1 and the model linked to a fibre of length L_2 should yield the model associated with a fibre of length $L_1 + L_2$.

Aim of the present paper is to overcome the limitations of [10, 11, 16–18] by developing a new model of optical fibres with memory effects that accurately encapsulates the essential attributes of extended optical fibres, such as those outlined in Properties 1 and 2 above. We shall call such model "Delocalised Interaction Model", or DIM in brief, and employ it to determine the ultimate quantum communication capabilities attainable through the strategic utilisation of memory effects in these systems. In particular, we will find the precise range of parameters that allow for the successful transmission of qubits, entanglement, or secret keys. Our main result is the calculation of the exact value of the quantum capacity Q, the two-way quantum capacity Q_2, and the secret-key capacity K [19, 20] of the optical fibre in the absence of thermal noise. Additionally, our investigation will encompass an examination of the existence of the phenomenon known as the "die-hard quantum communication" effect [10, 11, 21]. This effect, potentially enabling communication across optical fibres with arbitrarily low transmissivity, will be scrutinised to confirm its presence in our model, ensuring it is not merely an artifact of the localised signal cross-talking assumption made in [10, 11, 21]. Specifically we shall see that *for any arbitrarily low non-zero value of the single-signal transmissivity* $\lambda \in (0, 1]$ *of the fibre and for any arbitrarily large value of its associated thermal noise* $\nu \geq 0$*, there exists a non-zero time interval separating successive signals below which* Q*,* Q_2*, and* K *all become strictly positive.* In particular, we show that memory effects enable qubit distribution ($Q > 0$) even when $\lambda \in (0, \frac{1}{2}]$, which is not achievable with the corresponding memoryless optical fibre. Additionally, using the sufficient condition for entanglement distribution reported in [22], we show that memory effects enable two-way entanglement distribution ($Q_2 > 0$) and quantum key distribution ($K > 0$) even when $\lambda \in (0, \frac{\nu}{\nu+1}]$, thus surpassing the performance of memoryless optical fibres.

Preliminaries

In the framework of quantum Shannon theory [19, 20], the fundamental limitations of point-to-point quantum communication are determined by the *capacities* of quantum channels. The capacities quantify the maximum amount of information that can be reliably transmitted per channel use in the asymptotic limit of many uses. Different notions of capacities have been defined, based on

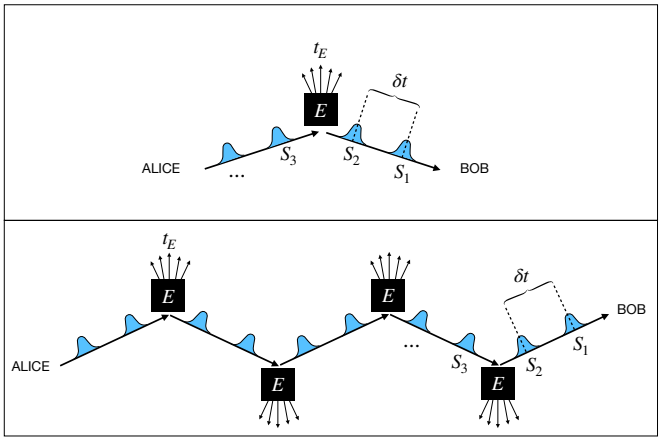


Figure 7.1: Pictorial representation of the mechanism which is responsible for the intra-signal interactions between a sequence of pulses S_1, S_2, $\cdots$ that travel along an optical fibre. In the LIM [16–18] such couplings occurs in a single location, through the mediation of a single common environment (upper panel). A more realistic description of the effect would instead allow for multiple cross-talks events distributed over the entire length of the fibre (bottom panel).

the type of information to be transmitted, such as qubits or secret-key bits, and the additional resources permitted in the protocol design, such as classical feedback. In this paper, we investigate three distinct capacities: on the one hand the quantum capacity Q, which measures the efficiency in the transmission of qubits with no additional resources; on the other, the two-way quantum capacity Q_2 and the secret-key capacity K, which instead gauge the efficiency in the transmission of qubits and secret key, respectively, with the additional free resource of a public two-way classical communication channel between the sender (Alice) and the receiver (Bob) [19, 20].

The signals transmitted along an optical fibre can be described in terms of an ordered array of localised e.m. pulses S_1, S_2, $\cdots$, S_n, of assigned mean frequency ω_0 and bandwidth $\Delta\omega$ that, in the absence of dispersion, propagate rigidly through the fibre separated by a fixed delay time δt. In the quantum setting such modes are conventionally identified with a corresponding collection of independent annihilation operators a_1, a_2, $\cdots$, a_n that fulfil canonical commutation rules [23]. The memoryless regime is reached when δt is sufficiently large to prevent cross-talking among the transmitted signals: accordingly they will experience the same type of noise which, under very general conditions, is typically identified with a *thermal attenuator* channel $\mathscr{E}_{\lambda,\nu}$. This is a continuous-variable [24] single-mode quantum channel characterised by two parameters: $\lambda \in [0,1]$, which represents the transmissivity of the fibre (i.e. the ratio between the output energy and the input energy of the transmitted signal), and $\nu \in [0,\infty)$, which quantifies the thermal noise added by the environment (in the limit of zero temperature $\nu = 0$, the transformation is conventionally referred to as the *pure-loss channel*). Mathematically, the action of $\mathscr{E}_{\lambda,\nu}$ on a generic input state ρ of the ith signal can be expressed as a beam splitter mixing the latter with a dedicated local Bosonic bath E_i, initialised in the thermal state τ_ν with mean photon number ν. In formula, this can be written as

$$\mathscr{E}_{\lambda,\nu}(\rho) \coloneqq \mathrm{Tr}_{E_i}\left[U_\lambda\Big(\rho\otimes\tau_\nu\Big)U_\lambda{}^\dagger\right], \tag{7.1}$$

where a_i and b_i are the annihilation operators of S_i and E_i, $U_\lambda \coloneqq e^{\arccos(\sqrt{\lambda})(a_i^\dagger b_i - a_i\, b_i^\dagger)}$ is the unitary describing the beam splitter interaction, and Tr_{E_i} represents the partial trace w.r.t. E_i. The capacities Q [25–27], Q_2 [28], and K [28] of the pure-loss channel $\mathscr{E}_{\lambda,0}$ have been determined exactly. In contrast, only bounds are known for the capacities Q [25, 28–33], Q_2 [22, 28, 34, 35], and K [22, 28, 34, 35] of the thermal attenuator $\mathscr{E}_{\lambda,\nu}$. In particular, it is known that the quantum capacity of the thermal attenuator $Q(\mathscr{E}_{\lambda,\nu})$ vanishes if the transmissivity of the fibre falls below

the critical value of $\lambda \leq \frac{1}{2}$:

$$\lambda \leq \frac{1}{2} \Longrightarrow Q(\mathscr{E}_{\lambda,\nu}) = 0\,, \tag{7.2}$$

and, additionally, the equivalence "$\Longleftrightarrow$" holds for $\nu = 0$. It is also known that the two-way quantum capacity $Q_2(\mathscr{E}_{\lambda,\nu})$ and the secret-key capacity $K(\mathscr{E}_{\lambda,\nu})$ of the thermal attenuator vanish if and only if the transmissivity falls below the critical value of $\lambda \leq \frac{\nu}{\nu+1}$ [22]:

$$\lambda \leq \frac{\nu}{\nu+1} \Longleftrightarrow Q_2(\mathscr{E}_{\lambda,\nu}) = K(\mathscr{E}_{\lambda,\nu}) = 0\,. \tag{7.3}$$

Since typically the transmissivity of an optical fibre decreases exponentially with its length, under the memoryless assumption there are strong limitations on the distance at which it is possible to perform qubit distribution ($Q > 0$), two-way entanglement distribution ($Q_2 > 0$), and quantum key distribution ($K > 0$) without relying on quantum repeaters. For instance, modern optical fibres typically exhibit signal attenuation rates of around 0.2 dB/km, with the best recorded value being 0.14 dB/km [36, 37], meaning that the quantum capacity vanishes if the fibre is longer than 15 km or at most 21 km.

Early attempts to incorporate memory effects into optical fibres were made in Refs. [16–18] with a model which from now on we shall refer to as "Localised Interaction Model" (LIM). In these works, following the approach outlined in [13], intra-signal couplings are induced by an ordered sequence of collisional events in which each transmitted pulse interacts with a common reservoir E (see upper panel of Fig. 7.1). The latter is characterised by a resetting mechanism that endeavors to restore it to its initial configuration over a thermalization time-scale t_E. Consequently, when the time delay δt separating two successive input signals exceeds the thermalization time t_E, each pulse encounters the same environmental state, rendering the communication effectively memoryless. Conversely, when δt is smaller than or comparable to t_E, after colliding with one of the signals, the reservoir E does not have sufficient time to revert to τ_ν and functions as a mediator for pulse interactions. As illustrated in Fig. 7.2(b), LIM emulates this intricate dynamics of the n input signals via the n-mode quantum channel $\Phi^{(1,n)}_{\lambda,\mu,\nu}$, obtained by connecting the parallel optical lines which describe the noisy propagation of the modes associated with the annihilation operators a_1, a_2, $\cdots$, a_n in the memoryless regime (panel (a) of Fig. 7.2), with a series of additional beam splitters of transmissivity μ. This parameter serves as the "memory parameter" of the model, quantifying what is the fraction of the energy lost by the ith input signal that can potentially be absorbed by the subsequent ones by mixing it with the thermal contributions of the local environments E_{i+1}, E_{i+2}, $\cdots$, E_n. Ranging from 0 (where $\Phi^{(1,n)}_{\lambda,0,\nu}$ reduces to $n$fold memoryless channels $\mathscr{E}^{\otimes n}_{\lambda,\nu}$) to 1 (full memory), μ effectively encapsulates the interplay between the time interval δt separating two consecutive input signals and the characteristic thermalization time t_E (for instance, one plausible expression for μ might be defined as $\mu := \exp(-\delta t/t_E)$). Note also that similarly to the memoryless case, the LIM transformation $\Phi^{(1,n)}_{\lambda,\mu,\nu}$ is still characterised by an effective transmissivity parameter λ which in this case represents the attenuation experienced when a single signal traverses the line in isolation, and by the thermal noise parameter ν which is responsible for defining the temperature of the local baths.

A crucial insight which emerges from Refs. [16–18] is that memory couplings can improve the communication efficiency of optical fibres, thus opening the door to the realisation of the "die-hard quantum communication" effect [10, 11, 21]. For instance, in the zero-temperature ($\nu = 0$) limit the value of Q computed for LIM turns out to be an increasing function of μ for each assigned transmissivity value λ (see also Appendix 7.7). Unfortunately, as outlined in the introduction, utilising LIM for investigating real-world scenarios is problematic, mainly because it fails to consider the possibility that the signals could experience multiple cross-talks over different locations of the optical fibre as depicted in the bottom panel of Fig. 7.1. The limitations of LIM become particularly evident when one observes that the associated mapping $\Phi^{(1,n)}_{\lambda,\mu,\nu}$ does not fulfil neither Property 1 nor Property 2, which should instead hold for long transmission lines. For

instance, a close inspection of the interferometric representation of Fig. 7.2(b) reveals that, as long as $\mu > 0$, even for $\lambda = 0$ the channel $\Phi^{(1,n)}_{0,\mu,\nu}$ is still capable of transmitting signals from Alice to Bob (e.g. the photons of the first input mode a_1 will be received by Bob at the output of the mode a_2). The possibility of having multiple cross-talks will arguably prevent this possibility via destructive interferences that could in principle spoil the advantages pointed out in [10, 11, 16–18, 21]. The primary objective of this paper is to shed light on this problem, introducing a new model for memory effects in optical fibres that is immune to the shortcomings of LIM.

Results

In this section we present an improved version of LIM which can be used to describe memory effects in optical fibres when the transmitted signals have a chance of experiencing multiple interactions along the entire length of the communication line. As we shall see, such construction, which we dub "Delocalised Interaction Model" (DIM), is particularly apt to represent the input-output relations occurring in long optical fibres as, unlike LIM, it fulfils Properties 1 and 2 detailed in the introductory section.

7.1.1 Delocalised Interaction Model

The idea behind the DIM approach is relatively simple: a spatially homogeneous optical fibre of finite length L, characterised by a thermal noise $\nu \geq 0$ and by a single-pulse transmissivity $\lambda \in [0, 1]$, is seen as the composition of M identical optical fibres of length L/M. In the limit of large M, such pieces are sufficiently small that they can be effectively described via LIM mappings. The resulting input-output relation of the global fibre is hence computed by first properly concatenating such individual terms and then taking the continuum limit $M \to \infty$.

More specifically, considering that for a spatially homogeneous fibre the single-signal transmissivity λ is exponentially decreasing in the length L, to each of the M infinitesimal components of the fibre we can assign an effective single-signal transmissivity $\lambda^{1/M}$. Their individual LIM channel representations are given by maps $\Phi^{(1,n)}_{\lambda^{1/M},\mu,\nu}$, each of which is characterised by the same local temperature parameter ν and by the same memory parameter μ of the global fibre. Similarly to the cascade construction of [38] for fixed M, the resulting input-output DIM map is hence provided by the M-fold concatenation

$$\Phi^{(M,n)}_{\lambda,\mu,\nu} = \left(\Phi^{(1,n)}_{\lambda^{1/M},\mu,\nu}\right)^{M} := \underbrace{\Phi^{(1,n)}_{\lambda^{1/M},\mu,\nu} \circ \cdots \circ \Phi^{(1,n)}_{\lambda^{1/M},\mu,\nu}}_{M}\,, \tag{7.4}$$

whose interferometric representation in terms of beam splitter couplings is given in Fig. 7.2(c). Note that as for LIM, the present scheme reduces to the most common model of memoryless optical fibre when the memory parameter μ vanishes: indeed, if $\mu = 0$ each of the M infinitesimal optical fibres can be modelled as a thermal attenuator $\mathscr{E}_{\lambda^{1/M},\nu}$ of transmissivity $\lambda^{1/M}$, and (7.4) gives $\Phi^{(M,n)}_{\lambda,0,\nu} = \mathscr{E}_{\lambda,\nu}$. If $\mu > 0$, instead, then the environments $E^{(j)}_i$ are not always in the thermal state τ_ν, and their state depends on all previous input signals — the model exhibits memory effects.

As shown in Theorem 80 in the Supplementary, for $M \to \infty$ the mapping $\Phi^{(M,n)}_{\lambda,\mu,\nu}$ converges to an n-mode quantum channel denoted by $\Phi^{(n)}_{\lambda,\mu,\nu}$. The mathematically precise sense in which this convergence happens involves the notion of *strong convergence*, discussed in the Methods. The family of quantum channels $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$, which forms a *quantum memory channel* [12], characterises our model of optical fibre with memory effects. Most importantly, contrary to the LIM of Ref. [10, 11, 16–18], the map $\Phi^{(n)}_{\lambda,\mu,\nu}$ satisfies the above Properties 1–2. The proof of this fact is rather technical; the interested reader can find it in Sec. 7.4 of the Supplementary.

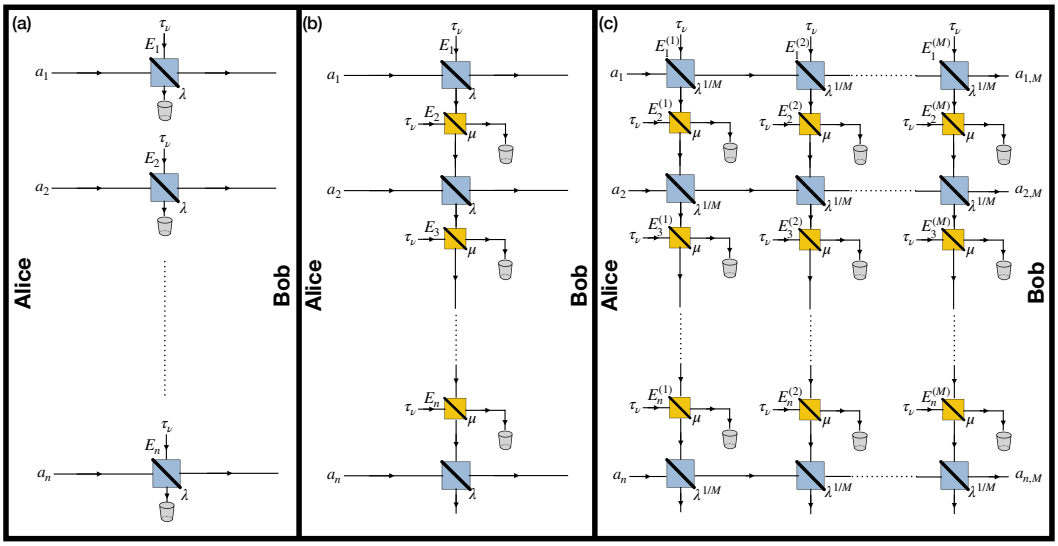


Figure 7.2: Interferometric representation of the n-mode quantum channels which describe the transmission of e.m. pulses along an optical fibre of single-signal transmissivity λ and local temperature ν for different memory configurations. Panel (a) memoryless regime: here each one of the input modes a_1, a_2, $\cdots$, a_n evolves independently from the others, undergoing to the same thermal attenuation mapping $\mathscr{E}_{\lambda,\nu}$ induced by beam splitter couplings (7.1) with the local thermal state τ_ν of the local environments E_1, E_2, $\cdots$, E_n. Panel (b) LIM channel $\Phi^{(1,n)}_{\lambda,\mu,\nu}$: in this case signal cross-talks are mediated by the yellow beam splitters of transmissivity μ. These allow the photons lost by the ith input mode to emerge in the output of the subsequent ones by letting them to interfere with the local bath E_{i+1}, E_{i+2}, $\cdots$, E_n. Setting $\mu = 0$ the LIM reduces to the memoryless case of panel (a), i.e. $\Phi^{(1,n)}_{\lambda,0,\nu} = \mathscr{E}^{\otimes n}_{\lambda,\nu}$. Panel (c) DIM channel $\Phi^{(M,n)}_{\lambda,\mu,\nu}$: the fibre is described by the concatenation (7.4) of M LIM channels $\Phi^{(1,n)}_{\lambda^{1/M},\mu,\nu}$ of transmissivity $\lambda^{1/M}$. In the Heisenberg representation, $\Phi^{(M,n)}_{\lambda,\mu,\nu}$ maps the annihilation operator a_i (depicted on the left) of the ith input signal into the annihilation operators $a_{i,M}$ (depicted on the right) of the ith output signal for all $i = 1, 2, \ldots, n$. For $j = 1, \cdots, M$, the symbols $E^{(j)}_1$, $E^{(j)}_2$, $\cdots$, $E^{(j)}_n$ represent the single-mode environments associated with the jth infinitesimal optical fibre element: all of them are initialised in the same thermal state τ_ν. Supplementary 7.8 presents a generalisation of DIM where memoryless attenuation along the fibre occurs, concurrently with the memory effects.

7.1.2 Capacities of the model

Let $Q(\lambda,\mu,\nu)$, $Q_2(\lambda,\mu,\nu)$, and $K(\lambda,\mu,\nu)$ be the quantum capacity, two-way quantum capacity, and secret key capacity, respectively, of the DIM quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$. The forthcoming Theorem 72, which is proved in Theorem 91 in the Supplementary, provides conditions on the parameter region where these capacities are strictly positive. In particular, it states that one can make the capacities strictly positive by sufficiently increasing the value of μ, for all nonzero values of the transmissivity λ, and all thermal photon numbers ν.

Theorem 72. *Let $\lambda \in (0,1)$ and $\mu \in [0,1)$. In the absence of thermal noise, i.e. $\nu = 0$, it holds that*

$$Q(\lambda,\mu,\nu=0) > 0 \iff \sqrt{\mu} > \frac{\log_2\left(\frac{1}{\lambda}\right)-1}{\log_2\left(\frac{1}{\lambda}\right)+1}\,. \tag{7.5}$$

In addition, for all $\nu \geq 0$ it holds that

$$K(\lambda,\mu,\nu),\, Q_2(\lambda,\mu,\nu) > 0 \iff \sqrt{\mu} > \frac{\ln\left(\frac{1}{\lambda}\right)-\ln(1+\frac{1}{\nu})}{\ln\left(\frac{1}{\lambda}\right)+\ln(1+\frac{1}{\nu})}\,. \tag{7.6}$$

The following theorem, reported and proved in Theorem 92 in the Supplementary, provides the exact solution for the capacities in the absence of thermal noise.

Theorem 73. *Let $\lambda \in (0,1)$ and $\mu \in [0,1)$. In absence of thermal noise, i.e. $\nu = 0$, it holds that*

$$\begin{aligned} Q\left(\lambda,\mu,\nu=0\right) &= \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi}\, \max\left\{0, \log_2\left(\frac{\eta^{(\lambda,\mu)}(x)}{1-\eta^{(\lambda,\mu)}(x)}\right)\right\}, \\ Q_2\left(\lambda,\mu,\nu=0\right) &= K\left(\lambda,\mu,\nu=0\right) \\ &= \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi}\, \log_2\left(\frac{1}{1-\eta^{(\lambda,\mu)}(x)}\right), \end{aligned} \tag{7.7}$$

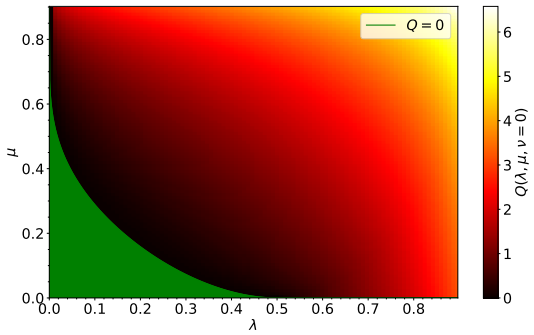
where

$$\eta^{(\lambda,\mu)}(x) := \lambda^{\frac{1-\mu}{1+\mu-2\sqrt{\mu}\cos(x/2)}} \quad \forall\, x \in [0,2\pi]\,. \tag{7.8}$$

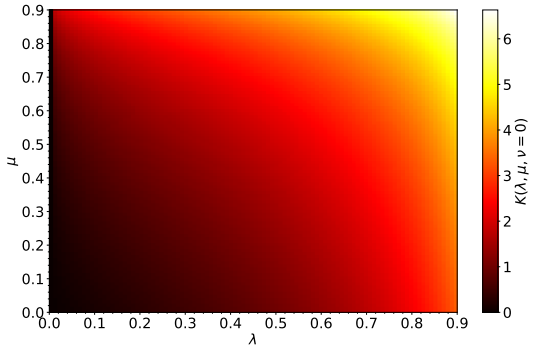
In Fig. 7.3, we plot the capacities given in (7.7). If $\mu = 0$ (i.e. in the memoryless case), all these capacities are equal to the corresponding capacities of the pure-loss channel. Upon plotting the capacities in Fig. 7.3, one observes that for any $\lambda \in (0,1]$ all the capacities $Q\left(\lambda,\mu,\nu=0\right)$, $Q_2\left(\lambda,\mu,\nu=0\right)$, and $K\left(\lambda,\mu,\nu=0\right)$ are monotonically increasing in $\mu \in [0,1]$. Hence, at least in the absence of thermal noise, as the memory parameter μ increases (corresponding to a decrease in the time interval between consecutive signals), quantum communication performances improve. It is reasonable to expect that such a monotonicity in μ holds even for $\nu > 0$.

Discussion

In this paper we analysed quantum communication, entanglement distribution, and quantum key distribution across optical fibres in the little-studied case where memory effects are present, and hence commonly employed approximations break down. It is generally believed that memory effects should improve the information transfer of a communication line, by offering a means to recover potentially lost information that would otherwise be unrecoverable in a memoryless scenario. Indeed, in a memoryless scenario, if photons are lost, they are irretrievably gone. However, in the presence of memory effects, there exists a probability that lost photons can interact with subsequent signals, allowing some information encoded in lost photons to persist within the fibre.



(a)



(b)

Figure 7.3: In **(a)** we plot the quantum capacity $Q(\lambda, \mu, \nu = 0)$ of the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu=0}\}_{n\in\mathbb{N}}$ with respect to $\lambda, \mu \in [0, 0.9]$ for $\nu = 0$. The green region is the one where the quantum capacity vanishes, which is given by all the points (λ, μ) such that $\sqrt{\mu} \leq \frac{\log_2\left(\frac{1}{\lambda}\right)-1}{\log_2\left(\frac{1}{\lambda}\right)+1}$, as established by Theorem 72. In **(b)** we plot the secret-key capacity $K(\lambda, \mu, \nu = 0)$ of the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu=0}\}_{n\in\mathbb{N}}$ with respect to $\lambda, \mu \in [0, 0.9]$ for $\nu = 0$. This capacity is exactly equal to the two-way quantum capacity $Q_2(\lambda, \mu, \nu = 0)$, as guaranteed by Theorem 73. The capacities plotted in **(a)** and in **(b)** are calculated by exploiting the expressions given in (7.7). In **(a)** and in **(b)** we have limited the range of λ and μ to the interval $[0, 0.9]$ because all the capacities Q, Q_2, and K diverge when either λ or μ approaches 1.

We first proposed a model of optical fibre with memory effects which overcomes problems of the LIM previously introduced in the literature [10, 11, 16–18]. Our model depends on three parameters: the transmissivity $\lambda \in [0, 1]$ of the optical fibre, the thermal noise $\nu \geq 0$, and the memory parameter $\mu \in [0, 1]$, with the latter being related to the time interval between subsequent signals. In particular, increasing μ has the operational meaning as a decrease in the time interval between subsequent signals.

In Theorem 73 we found the exact solution for the quantum capacity Q, the two-way quantum capacity Q_2, and the secret-key capacity K of the DIM channel in the absence of thermal noise ($\nu = 0$). These capacities are monotonically increasing in the memory parameter μ, meaning that memory effects improve quantum communication. It is worth noting that in the presence of thermal noise ($\nu > 0$) the capacities remain unknown even in the absence of memory effects ($\mu = 0$), as the capacities of the thermal attenuator are currently unknown.

The reader might question the meaningfulness of considering communication tasks assisted by two-way classical communication in a setting where the memory parameter μ, and thus the time interval between subsequent signals, is fixed. Indeed, in general, the rounds of classical communication between consecutive channel uses may vary during the communication protocol. Additionally, another problem is that, when the sender and receiver are very far apart, even a

single round of classical communication requires an excessively long waiting time which prevents the exploitation of memory effects altogether. However, in the case of Choi-simulable channels [28], it is meaningful to consider such communication tasks. This is because optimal entanglement and secret-key distribution strategies for Choi-simulable channels can be achieved by initially utilising the channel (with a constant and short time interval between subsequent input signals) to share multiple copies of the Choi state, followed by employing optimal distillation protocols (either for entanglement or secret-key) to distil the Choi state. Fortunately, our model is associated with the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$, which is Choi simulable because it is Gaussian [28].

In the absence of memory effects ($\mu = 0$), it is known that no quantum communication tasks can be achieved when the transmissivity is sufficiently low ($Q = 0$ for $\lambda \leq \frac{1}{2}$, and $Q_2 = K = 0$ for $\lambda \leq \frac{\nu}{\nu+1}$). However, in Theorem 72 we established that for any transmissivity $\lambda > 0$ and thermal noise $\nu \geq 0$, there exists a critical value of the memory parameter μ above which it becomes possible to achieve qubit distribution ($Q > 0$), entanglement distribution ($Q_2 > 0$), and secret-key distribution ($K > 0$). This result is particularly intriguing as it demonstrates that — at least within our model — memory effects provide an advantage, enabling quantum communication tasks to be performed in highly noisy regimes where it was previously considered impossible without the use of quantum repeaters. While this result is model dependent, it offers hope that memory effects could offer a concrete route to achieve efficient long-distance quantum communication with fewer quantum repeaters than previously believed. Specifically, this result bears resemblance to the "die-hard quantum communication" effect" observed in [10, 11] within their toy model of optical fibre with memory effects, which was limited to the analysis of Q only and does not consider K and Q_2. In our paper, we not only demonstrate the persistence of such an effect in a more realistic model, accounting for Q, Q_2, and K, but we also derive an analytical expression for the critical value of the memory parameter μ, which goes beyond the findings of [10, 11]. Note that, by relating μ to the temporal interval δt between subsequent input signals (e.g. $\mu = e^{-\delta t/t_E}$ with t_E being the thermalisation timescale), Theorem 72 can also be expressed in terms of the critical δt below which the above mentioned quantum communication tasks can be achieved.

Methods

Let $\mathcal{H}$ be an Hilbert space, let $\mathscr{L}(\mathcal{H})$ the space of linear operators on $\mathcal{H}$, and let $\mathfrak{S}(\mathcal{H})$ be the set of quantum states on $\mathcal{H}$. Given a quantum channel Φ and a sequence of quantum channels $\{\Phi_k\}_{k\in\mathbb{N}}$, we say that $\{\Phi_k\}_{k\in\mathbb{N}}$ *strongly converges* to Φ if

$$\lim_{k\to\infty} \|\Phi_k(\rho) - \Phi(\rho)\|_1 = 0 \quad \forall\, \rho \in \mathfrak{S}(\mathcal{H})\,, \tag{7.9}$$

where $\|\Theta\|_1 := \mathrm{Tr}\sqrt{\Theta^\dagger\Theta}$ denotes the trace norm of a linear operator Θ. We now define quantum memoryless and memory channels, and introduce relevant notations.

Definition 4. Let $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ be a family of quantum channels with $\Phi^{(n)} : \mathscr{L}(\mathcal{H}^{\otimes n}) \to \mathscr{L}(\mathcal{H}^{\otimes n})$ for all $n \in \mathbb{N}$. The family $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ is *memoryless* if there exists a quantum channel $\Phi : \mathscr{L}(\mathcal{H}) \to \mathscr{L}(\mathcal{H})$ such that $\Phi^{(n)} = \Phi^{\otimes n}$ for all $n \in \mathbb{N}$. In such a scenario, we will also refer to Φ as a memoryless channel. Moreover, given a capacity C (e.g. $C \in \{Q, Q_2, K\}$), the capacity C of the memoryless quantum channel Φ is denoted as $C(\Phi)$.

Definition 5. Let $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ be a family of quantum channels with $\Phi^{(n)} : \mathscr{L}(\mathcal{H}^{\otimes n}) \to \mathscr{L}(\mathcal{H}^{\otimes n})$ for all $n \in \mathbb{N}$. The family $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ is a *memory quantum channel* if it is not memoryless. In such a scenario, the term "n channel uses" corresponds to the quantum channel $\Phi^{(n)}$. Moreover, given a capacity C (e.g. $C \in \{Q, Q_2, K\}$), the capacity C of the quantum memory channel $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ is denoted as $C\left(\{\Phi^{(n)}\}_{n\in\mathbb{N}}\right)$.

For all $\lambda \in [0,1]$ the capacities Q [25–27], Q_2 [28], and K [28] of the pure-loss channel $\mathscr{E}_{\lambda,0}$ are given by:

$$\begin{aligned} Q(\mathscr{E}_{\lambda,0}) &= \begin{cases} \log_2\left(\frac{\lambda}{1-\lambda}\right) & \text{if } \lambda \in [\frac{1}{2},1] \ , \\ 0 & \text{if } \lambda \in [0,\frac{1}{2}] \ . \end{cases} \\ Q_2(\mathscr{E}_{\lambda,0}) &= K(\mathscr{E}_{\lambda,0}) = \log_2\left(\frac{1}{1-\lambda}\right) . \end{aligned} \tag{7.10}$$

In the following we will outline the key concepts that enabled us to derive the results stated in Section 7.1.2. To achieve this, we generalise methods introduced in [16–18]. Let us begin with the forthcoming Theorem 74, which is proved in Theorem 80 in the Supplementary.

Theorem 74. *Let $\lambda, \mu \in [0,1]$, $\nu \geq 0$, and $n \in \mathbb{N}$. There exists passive [24] n-mode unitary transformations $\mathscr{U}_1, \mathscr{U}_2$ such that*

$$\mathscr{U}_2 \circ \Phi^{(n)}_{\lambda,\mu,\nu} \circ \mathscr{U}_1 = \bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu} \, , \tag{7.11}$$

where the transmissivities $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\ldots,n}$ are the square of the singular values, indexed in increasing order, of the $n \times n$ real matrix $\bar{A}^{(n,\lambda,\mu)}$ whose (i,k) element is

$$\bar{A}^{(n,\lambda,\mu)}_{i,k} := \Theta(i-k)\sqrt{\lambda}\mu^{\frac{i-k}{2}} L^{(-1)}_{i-k}(-\ln \lambda) \tag{7.12}$$

for all $i,k \in \{1,2,\ldots,n\}$. Here, $L^{(-1)}_m$ is a generalised Laguerre polynomial, and

$$\Theta(x) := \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{7.13}$$

Consequently, any capacity of the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$ coincides with the corresponding capacity of $\{\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\}_{n\in\mathbb{N}}$.

The proof of (7.11) involves expressing the output annihilation operators $\{a_{i,M}\}_{i=1,2,\ldots,n}$ (see Figure 7.2) in terms of the input ones $\{a_i\}_{i=1,2,\ldots,n}$, performing a Bogoliubov transformation, and finally taking the continuum limit $M \to \infty$. As a consequence of (7.11), the quantum memory channel $\Phi^{(n)}_{\lambda,\mu,\nu}$, which models n uses of the optical fibre, is unitarily equivalent to a tensor product of n distinct thermal attenuators. Consequently, if Alice and Bob are linked by $\Phi^{(n)}_{\lambda,\mu,\nu}$ (resp. $\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}$), they can simulate $\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}$ (resp. $\Phi^{(n)}_{\lambda,\mu,\nu}$) through the application of $\mathscr{U}_1$ (resp. $\mathscr{U}_1^\dagger$) by Alice just before transmission, and $\mathscr{U}_2$ (resp. $\mathscr{U}_2^\dagger$) by Bob just after reception. This implies the capacity equivalence between $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$ and $\{\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\}_{n\in\mathbb{N}}$, as stated in the final part of Theorem 74. Notably, this equivalence holds even for energy-constrained capacities [12], thanks to the passivity of $\mathscr{U}_1$.

Let us analyse the behaviour of the transmissivities $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\ldots,n}$, as the number of uses of the optical fibre, n, approaches infinity. By plotting in Fig. 7.4 the points

$$\left\{ \left(2\pi \frac{j}{n} \, , \, \eta_j^{(n,\lambda,\mu)} \right) \, : \, j = 1,2,\ldots,n \right\} \tag{7.14}$$

on a two-dimensional plane, we observe that they converge for $n \to \infty$ to the graph of a function $\eta^{(\lambda,\mu)} : [0,2\pi] \to \mathbb{R}$, which we dub *effective transmissivity function*. We formally state this result in the forthcoming Theorem 75 and we provide its proof in Theorem 90 in the Supplementary by establishing a corollary of the *Avram–Parter theorem* [39, 40], a matrix analysis result about the asymptotic behaviour for $n \to \infty$ of the singular values of a $n \times n$ Toeplitz matrix. Notably, our corollary, reported in Theorem 89 in the Supplementary, seems to be unknown in the matrix analysis literature and we believe it may have independent interest.

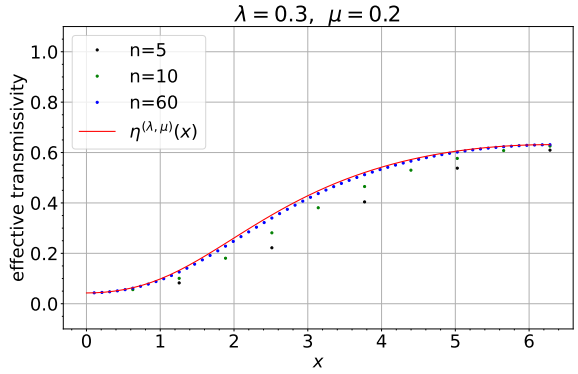


Figure 7.4: The red curve is the effective transmissivity function $\eta^{(\lambda,\mu)}(x)$, as defined in (7.16), plotted with respect $x \in [0, 2\pi]$ for $\lambda = 0.3$ and $\mu = 0.2$. For $n = 4$ (black points), $n = 10$ (green points), and $n = 60$ (blue points), we plot the n points $\left\{\left(2\pi\frac{j}{n}, \eta_j^{(n,\lambda,\mu)}\right) : j = 1, 2, \ldots, n\right\}$. We observe that these n points converge for $n \to \infty$ to the graph of the effective transmissivity function $\eta^{(\lambda,\mu)}$, as established by Theorem 75. Here, $\{\eta_j^{(n,\lambda,\mu)}\}_{j=1,2,\ldots,n}$ are the transmissivities defined in Theorem 74.

Theorem 75. *Let $\lambda \in [0, 1]$ and $\mu \in [0, 1)$. There exists a sequence $\{j_n\}_{n\in\mathbb{N}} \subseteq \mathbb{N}$ such that $0 \leq j_n \leq n$ for all $n \in \mathbb{N}$, $\lim_{n\to\infty} \frac{j_n}{n} = 0$, and*

$$\lim_{n\to\infty} \max\left\{\left|\eta_j^{(n,\lambda,\mu)} - \eta^{(\lambda,\mu)}\left(\frac{2\pi j}{n}\right)\right| : j \in \{j_n, \ldots, n\}\right\} = 0\,, \tag{7.15}$$

where the effective transmissivity function is defined as

$$\eta^{(\lambda,\mu)}(x) := \lambda^{\frac{1-\mu}{1+\mu-2\sqrt{\mu}\cos(x/2)}} \quad \forall\, x \in [0, 2\pi]\,. \tag{7.16}$$

As a consequence of the previous theorem and the fact that $\max_{x\in[0,2\pi]} \eta^{(\lambda,\mu)}(x) = \eta^{(\lambda,\mu)}(2\pi)$, the value $\eta^{(\lambda,\mu)}(2\pi)$ determines whether or not the capacities of our model are strictly positive. Hence, (7.2) and (7.3) imply the following theorem, which is proved in Theorem 91 in the Supplementary.

Theorem 76. *Let $\lambda \in (0, 1)$, $\mu \in [0, 1)$, and $\nu \geq 0$. Let $C(\lambda, \mu, \nu)$ be one of the following capacities of the quantum memory channel $\{\Phi_{\lambda,\mu,\nu}^{(n)}\}_{n\in\mathbb{N}}$: quantum capacity Q, two-way quantum capacity Q_2, or secret key capacity K. It holds that*

$$C(\lambda, \mu, \nu) > 0 \iff C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu}) > 0\,, \tag{7.17}$$

where $\eta^{(\lambda,\mu)}$ is reported in (7.16). In particular, in the absence of thermal noise, i.e. $\nu = 0$, it holds that

$$Q(\lambda, \mu, \nu = 0) > 0 \iff \sqrt{\mu} > \frac{\log_2\left(\frac{1}{\lambda}\right) - 1}{\log_2\left(\frac{1}{\lambda}\right) + 1}\,. \tag{7.18}$$

In addition, for all $\nu \geq 0$ it holds that

$$K(\lambda, \mu, \nu),\, Q_2(\lambda, \mu, \nu) > 0 \iff \sqrt{\mu} > \frac{\ln\left(\frac{1}{\lambda}\right) - \ln(1 + \frac{1}{\nu})}{\ln\left(\frac{1}{\lambda}\right) + \ln(1 + \frac{1}{\nu})}\,. \tag{7.19}$$

Now, let us provide a brief overview of the key ideas underlying the derivation of the capacities of $\{\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\}_{n\in\mathbb{N}}$. For additional technical details, see Theorem 92 in the Supplementary.

For $l, P \in \mathbb{N}$, consider Pl channel uses as divided into $P \to \infty$ groups, each containing $l \to \infty$ channel uses:

$$\bigotimes_{i=1}^{lP} \mathscr{E}_{\eta_i^{(Pl,\lambda,\mu)},\nu} = \bigotimes_{j=1}^{l} \bigotimes_{p=1}^{P} \mathscr{E}_{\eta_{(p-1)l+j}^{(Pl,\lambda,\mu)},\nu} . \tag{7.20}$$

Roughly speaking, Theorem 75 implies that for any $j \in \{1, 2, \ldots, l\}$ it holds that

$$\eta_{(p-1)l+j}^{(Pl,\lambda,\mu)} \simeq \eta^{(\lambda,\mu)}\left(2\pi \frac{p}{P}\right) \quad \text{for } l, P \to \infty , \tag{7.21}$$

meaning that lP channel uses, with $l, P \to \infty$, corresponds to l uses of the following P-mode thermal attenuator:

$$\bigotimes_{p=1}^{P} \mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi\frac{p}{P}),\nu} , \tag{7.22}$$

which is a memoryless quantum channel. Hence, for any capacity C (e.g. $C = Q, Q_2, K$) it holds that the capacity C of the quantum memory channel $\{\Phi_{\lambda,\mu,\nu}^{(n)}\}_{n\in\mathbb{N}}$ satisfies

$$C\left(\{\Phi_{\lambda,\mu,\nu}^{(n)}\}_{n\in\mathbb{N}}\right) \overset{\text{(i)}}{\approx} \lim_{P\to\infty} \frac{1}{P} C\left(\bigotimes_{p=1}^{P} \mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi\frac{p}{P}),\nu}\right) \tag{7.23}$$

$$\overset{\text{(ii)}}{\geq} \lim_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P} C\left(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi\frac{p}{P}),\nu}\right) \tag{7.24}$$

$$\overset{\text{(iii)}}{=} \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} C\left(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),\nu}\right) , \tag{7.25}$$

where: (i) comes from the fact that a single use of the P-mode attenuator in (7.22) is approximately equivalent to P uses of the optical fibre; (ii) is a consequence of the fact that Alice and Bob can independently employ the optimal communication strategy for each of the P single-mode channels that define the P-mode attenuator; in (iii) we have just introduced the Riemann integral. Moreover, the equality in (7.24) holds if C is additive, i.e. if C is such that for all transmissivities $\{\lambda_p\}_{p=1,2,\ldots,P}$ and all $P \in \mathbb{N}$ it holds that

$$C\left(\bigotimes_{p=1}^{P} \mathscr{E}_{\lambda_p,\nu}\right) = \sum_{p=1}^{P} C\left(\mathscr{E}_{\lambda_p,\nu}\right) . \tag{7.26}$$

Hence, leveraging the additivity and the expression in (7.10) of the capacities Q [41], Q_2 [28], and K [28] of the pure-loss channel, we can derive the precise values of the capacities of our model in the absence of thermal noise. This exact solution is provided by Theorem 77, which is proved in Theorem 92 in the Supplementary.

Theorem 77. *Let $\lambda \in (0, 1)$, $\mu \in [0, 1)$, $\nu \geq 0$. Let C be one of the following capacities: quantum capacity Q, two-way quantum capacity Q_2, or secret key capacity K. In addition, let $C(\lambda, \mu, \nu)$ be the capacity C of the quantum memory channel $\{\Phi_{\lambda,\mu,\nu}^{(n)}\}_{n\in\mathbb{N}}$. In absence of thermal noise, i.e. $\nu = 0$, it holds that*

$$C(\lambda, \mu, \nu = 0) = \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} C\left(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),0}\right) . \tag{7.27}$$

where $\eta^{(\lambda,\mu)}(x)$ is the effective transmissivity function expressed in (7.16). In particular,

$$\begin{aligned} Q(\lambda, \mu, \nu = 0) &= \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} \max\left\{0, \log_2\left(\frac{\eta^{(\lambda,\mu)}(x)}{1 - \eta^{(\lambda,\mu)}(x)}\right)\right\} , \\ Q_2(\lambda, \mu, \nu = 0) &= K(\lambda, \mu, \nu = 0) \\ &= \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} \log_2\left(\frac{1}{1 - \eta^{(\lambda,\mu)}(x)}\right) . \end{aligned} \tag{7.28}$$

Moreover, in the presence of thermal noise, i.e. $\nu > 0$, it holds that

$$C(\lambda, \mu, \nu) \geq \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} C\left(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),\nu}\right) . \tag{7.29}$$

Notation and preliminaries

Let $\mathcal{H}$ be an Hilbert space, let $\mathscr{L}(\mathcal{H})$ the space of linear operators on $\mathcal{H}$, and let $\mathfrak{S}(\mathcal{H})$ be the set of quantum states on $\mathcal{H}$. The trace norm of a linear operator Θ is defined by $\|\Theta\|_1 := \mathrm{Tr}\sqrt{\Theta^\dagger\Theta}$. A map $\Phi : \mathscr{L}(\mathcal{H}) \to \mathscr{L}(\mathcal{H})$ is a quantum channel if it is linear, completely positive, and trace preserving. Given a quantum channel Φ and a sequence of quantum channels $\{\Phi_k\}_{k\in\mathbb{N}}$, we say that $\{\Phi_k\}_{k\in\mathbb{N}}$ *strongly converges* to Φ when

$$\lim_{k\to\infty} \|\Phi_k(\rho) - \Phi(\rho)\|_1 = 0 \quad \forall\, \rho \in \mathfrak{S}(\mathcal{H}) . \tag{7.30}$$

We now define quantum memoryless and memory channels, and introduce relevant notations.

Definition 6. Let $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ be a family of quantum channels with $\Phi^{(n)} : \mathscr{L}(\mathcal{H}^{\otimes n}) \to \mathscr{L}(\mathcal{H}^{\otimes n})$ for all $n \in \mathbb{N}$. The family $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ is *memoryless* if there exists a quantum channel $\Phi : \mathscr{L}(\mathcal{H}) \to \mathscr{L}(\mathcal{H})$ such that $\Phi^{(n)} = \Phi^{\otimes n}$ for all $n \in \mathbb{N}$. In such a scenario, we will also refer to Φ as a memoryless channel.

Definition 7. Let $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ be a family of quantum channels with $\Phi^{(n)} : \mathscr{L}(\mathcal{H}^{\otimes n}) \to \mathscr{L}(\mathcal{H}^{\otimes n})$ for all $n \in \mathbb{N}$. The family $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ is a *memory quantum channel* if it is not memoryless. In such a scenario, the term "n channel uses" corresponds to the quantum channel $\Phi^{(n)}$.

Quantum information with continuous variables systems

Let us provide a brief overview of the formalism of quantum information with continuous variable systems [24]. In this context, we consider m modes of harmonic oscillators denoted by S_1, S_2, ..., S_m, which are associated with the Hilbert space $L^2(\mathbb{R}^m)$, which comprises all square-integrable complex-valued functions over $\mathbb{R}$. Each mode represents a single mode of electromagnetic radiation with a definite frequency and polarisation. For each $j = 1, 2, \ldots, m$, the annihilation operator a_j of mode S_j is defined as $a_j := \frac{\hat{x}_j + i\hat{p}_j}{\sqrt{2}}$, where $\hat{x}_j$ and $\hat{p}_j$ are the well-known position and momentum operators of mode S_j. The operator $a_j^\dagger$ is referred to as the creation operator of mode S_j. The operator $a_j^\dagger a_j$ corresponds to the photon number of mode S_j. The nth Fock state of mode S_j is defined as

$$|n\rangle_{S_j} := \frac{(a_j^\dagger)^n}{\sqrt{n!}} |0\rangle_{S_j} , \tag{7.31}$$

where $|0\rangle_{S_j}$ represents the vacuum state of mode S_j. The characteristic function $\chi_\rho : \mathbb{C}^m \to \mathbb{C}$ of a m-mode state $\rho \in \mathfrak{S}(L^2(\mathbb{R}^m))$ is defined as

$$\chi_\rho(\alpha) := \mathrm{Tr}\left[\rho D(\alpha)\right] , \tag{7.32}$$

where for all $\alpha \in \mathbb{C}^m$ the displacement operator $D(\alpha)$ is defined as

$$D(\alpha) := \exp\left[\sum_{j=1}^m (\alpha_j a_j^\dagger - \alpha_j^* a_j)\right] . \tag{7.33}$$

Any state ρ can be written in terms of its characteristic function as

$$\rho = \int \frac{\mathrm{d}^{2m}\alpha}{\pi^m} \chi_\rho(\alpha) D(-\alpha) , \tag{7.34}$$

and, consequently, quantum states and characteristic functions are in one-to-one correspondence. The following Lemma (see ([42, Lemma 4] or [43, Theorem 2])) will be useful for the following.

Lemma 78. *Let $m \in \mathbb{N}$, let $\{\sigma_k\}_{k\in\mathbb{N}} \subseteq \mathfrak{S}\left(L^2(\mathbb{R}^m)\right)$ be a sequence of m-mode states, and let $\sigma \in \mathfrak{S}\left(L^2(\mathbb{R}^m)\right)$ be a m-mode state. Then, $\{\sigma_k\}_{k\in\mathbb{N}}$ converges in trace norm to σ if and only if the sequence of characteristic functions $\{\chi_{\sigma_k}\}_{k\in\mathbb{N}}$ converges pointwise to the characteristic function χ_σ. In formula, it holds that*

$$\lim_{k\to\infty} \|\sigma_k - \sigma\|_1 = 0 \qquad \Longleftrightarrow \qquad \lim_{k\to\infty} \chi_{\sigma_k}(\alpha) = \chi_\sigma(\alpha) \quad \forall\, \alpha \in \mathbb{C}\,. \tag{7.35}$$

Remark 79. *Let us consider m single-mode systems with annihilation operators $a_1, a_2, \ldots, a_m$ and let us define the following vector of annihilation operators:*

$$\boldsymbol{a} \coloneqq (a_1, a_2, \ldots, a_m)^T\,. \tag{7.36}$$

For any $n \times n$ orthogonal matrix O one can define a unitary operator U_O on $L^2(\mathbb{R})^{\otimes m}$ such that $U_O^\dagger \mathbf{a} U_O = O\mathbf{a}$. The transformation $\mathscr{U}_O(\cdot) \coloneqq U_O \cdot U_O^\dagger$ is dubbed "passive" [24] and it preserves the total photon number, i.e.

$$U_O^\dagger \left(\sum_{j=1}^m a_j^\dagger a_j\right) U_O = \sum_{j=1}^m a_j^\dagger a_j\,. \tag{7.37}$$

Note that, at the level of characteristic functions, for any quantum state $\rho \in \mathfrak{S}\left(L^2(\mathbb{R}^m)\right)$ it holds that

$$\chi_{\mathscr{U}_O(\rho)}(\alpha) = \chi_\rho(O^T\alpha) \qquad \forall\, \alpha \in \mathbb{C}^m\,. \tag{7.38}$$

7.1.3 Beam splitter and thermal attenuator

Let $\mathcal{H}_S$ and $\mathcal{H}_E$ be single-mode systems and let a and b denote their annihilation operators, respectively. For all $\lambda \in [0,1]$ the unitary operator describing a beam splitter of transmissivity λ is

$$U_\lambda^{SE} \coloneqq \exp\left[\arccos\sqrt{\lambda}\left(a^\dagger b - a\, b^\dagger\right)\right]\,. \tag{7.39}$$

Under the beam splitter unitary, the annihilation operators a and b transform as

$$\begin{aligned}
\left(U_\lambda^{SE}\right)^\dagger a\, U_\lambda^{SE} &= \sqrt{\lambda}\, a + \sqrt{1-\lambda}\, b\,,\\
U_\lambda^{SE} a\, \left(U_\lambda^{SE}\right)^\dagger &= \sqrt{\lambda}\, a - \sqrt{1-\lambda}\, b\,,\\
\left(U_\lambda^{SE}\right)^\dagger b\, U_\lambda^{SE} &= -\sqrt{1-\lambda}\, a + \sqrt{\lambda}\, b\,,\\
U_\lambda^{SE} b\, \left(U_\lambda^{SE}\right)^\dagger &= \sqrt{1-\lambda}\, a + \sqrt{\lambda}\, b\,.
\end{aligned} \tag{7.40}$$

At the level of characteristic functions, the beam splitter transforms any two-mode state $\rho \in \mathfrak{S}(\mathcal{H}_S \otimes \mathcal{H}_E)$ as

$$\begin{aligned}
&\chi_{U_\lambda^{SE}\rho U_\lambda^{SE\dagger}}(\alpha,\beta) \coloneqq \mathrm{Tr}\left[U_\lambda^{SE}\rho U_\lambda^{SE^\dagger}\, D_S(\alpha)\otimes D_E(\beta)\right]\\
&= \chi_\rho\left(\sqrt{\lambda}\,\alpha - \sqrt{1-\lambda}\,\beta,\, \sqrt{1-\lambda}\,\alpha + \sqrt{\lambda}\,\beta\right) \qquad \alpha,\beta \in \mathbb{C}\,.
\end{aligned} \tag{7.41}$$

For all $\lambda \in [0,1]$ and $\nu \geq 0$, a thermal attenuator $\mathscr{E}_{\lambda,\nu} : \mathfrak{S}(\mathcal{H}_S) \to \mathfrak{S}(\mathcal{H}_S)$ is a quantum channel defined as follows:

$$\mathscr{E}_{\lambda,\nu}(\rho) \coloneqq \mathrm{Tr}_E\left[U_\lambda^{SE}\left(\rho^S \otimes \tau_\nu^E\right) U_\lambda^{SE\dagger}\right]\,, \tag{7.42}$$

Here, $\tau_\nu \in \mathfrak{S}(\mathcal{H}_E)$ denotes the thermal state with mean photon number equal to ν, which is defined by

$$\tau_\nu \coloneqq \frac{1}{\nu+1}\sum_{n=0}^\infty \left(\frac{\nu}{\nu+1}\right)^n |n\rangle\langle n|\,, \tag{7.43}$$

and its characteristic function reads

$$\chi_{\tau_\nu}(\alpha) := e^{-\frac{1}{2}(2\nu+1)|\alpha|^2} \quad \forall\, \alpha \in \mathbb{C}\,. \tag{7.44}$$

By exploiting (7.41) and (7.44), it can be shown that the thermal attenuator transforms any single-mode state ρ at the level of characteristic functions as follows:

$$\chi_{\mathscr{E}_{\lambda,\nu}(\rho)}(\alpha) := \chi_\rho\left(\sqrt{\lambda}\alpha\right) e^{-\frac{1}{2}(2\nu+1)(1-\lambda)|\alpha|^2} \quad \forall\, \alpha \in \mathbb{C}\,. \tag{7.45}$$

By exploiting (7.45) and the one-to-one correspondence between quantum states and characteristic functions, one can show that the following composition rule holds for any $\lambda_1, \lambda_2 \in [0,1]$ and $\nu \geq 0$:

$$\mathscr{E}_{\lambda_1,\nu} \circ \mathscr{E}_{\lambda_2,\nu} = \mathscr{E}_{\lambda_1\lambda_2,\nu}\,. \tag{7.46}$$

7.2 Quantum capacity, two-way quantum capacity, and secret-key capacity

In the framework of quantum Shannon theory [19, 20], the fundamental limitations of point-to-point quantum communication are established by the *capacities* of quantum channels. The capacities quantify the maximum amount of information that can be reliably transmitted per channel use in the asymptotic limit of many uses. Different notions of capacities have been defined, based on the type of information to be transmitted, such as qubits or secret-key bits, and the additional resources permitted in the protocol design, such as classical feedback. In this paper, we investigate three distinct capacities:

- the *quantum capacity* [19, 20], denoted as Q, is the maximum achievable rate of qubits that can be reliably transmitted through the channel in the limit of infinite channel uses. The rate of qubits is the ratio between the number of qubits transmitted and the number of uses of the channel.

- the *two-way quantum capacity* [20, Chapters 14 and 15], denoted as Q_2, is the maximum achievable rate of qubits that can be reliably transmitted through the channel in the limit of infinite channel uses by assuming that the sender Alice and the receiver Bob have free access to a public, noiseless, two-way classical communication line. Note that Q_2 is also the proper capacity for entanglement distribution. Indeed, an ebit (i.e. a two-qubit maximally entangled states) can teleport an arbitrary qubit [44], and the ability of transmitting an arbitrary qubit implies the ability of distributing an ebit.

- the *secret-key capacity* [20, Chapters 14 and 15], denoted as K, is the maximum achievable rate of secret-key bits that can be reliably transmitted through the channel in the limit of infinite channel uses by assuming that the sender Alice and the receiver Bob have free access to a public, noiseless, two-way classical communication line.

A rigorous definitions of capacity can be found in [12, 20]. Let C represent one of the three capacities mentioned above, namely $C \in \{Q, Q_2, K\}$. We will employ the following notation:

- The capacity C of a memoryless quantum channel Φ will be denoted as $C(\Phi)$.

- The capacity C of a quantum memory channel $\{\Phi^{(n)}\}_{n\in\mathbb{N}}$ will be denoted as $C\left(\{\Phi^{(n)}\}_{n\in\mathbb{N}}\right)$ (refer to Definition 7).

Hence, it follows that

$$C(\{\Phi^{\otimes n}\}_{n\in\mathbb{N}}) = C(\Phi)\,. \tag{7.47}$$

In addition, note that for any $\alpha > 0$ it holds that

$$C(\{\Phi^{\otimes \lceil \alpha n\rceil}\}_{n\in\mathbb{N}}) = \alpha C(\Phi)\,, \tag{7.48}$$

where $\lceil\cdot\rceil$ denotes the ceil function.

7.2.1 Quantum capacity, two-way quantum capacity, and secret-key capacity of the pure-loss channel and thermal attenuator

For all $\lambda \in [0,1]$ the capacities Q [25–27], Q_2 [28], and K [28] of the pure-loss channel $\mathscr{E}_{\lambda,0}$ are given by:

$$\begin{aligned} Q(\mathscr{E}_{\lambda,0}) &= \begin{cases} \log_2\left(\frac{\lambda}{1-\lambda}\right) & \text{if } \lambda \in [\frac{1}{2},1]\,, \\ 0 & \text{if } \lambda \in [0,\frac{1}{2}]\,. \end{cases} \\ Q_2(\mathscr{E}_{\lambda,0}) &= K(\mathscr{E}_{\lambda,0}) = \log_2\left(\frac{1}{1-\lambda}\right)\,. \end{aligned} \tag{7.49}$$

In particular, it holds that

$$\lambda \leq \frac{1}{2} \iff Q(\mathscr{E}_{\lambda,0}) = 0\,. \tag{7.50}$$

For $\lambda \in [0,1]$ and $\nu > 0$, the capacities Q, Q_2, and K of the thermal attenuator $\mathscr{E}_{\lambda,\nu}$ are not known, but bounds have been established: see [25, 28–33] for bounds on $Q(\mathscr{E}_{\lambda,\nu})$, while see [22, 25, 25, 28, 32, 34, 35, 45–51] for bounds on $Q_2(\mathscr{E}_{\lambda,\nu})$. Nevertheless, the region of zero two-way quantum capacity and zero secret-key capacity of the thermal attenuator have been determined and it reads [22]:

$$\lambda \leq \frac{\nu}{\nu+1} \iff Q_2(\mathscr{E}_{\lambda,\nu}) = K(\mathscr{E}_{\lambda,\nu}) = 0\,. \tag{7.51}$$

Although the exact region of zero quantum capacity of the thermal attenuator has not been determined, there exist known bounds on this region. In particular, it has been established that [25, 29, 52, 53]

$$\begin{aligned} \lambda &\leq \frac{\nu+\frac{1}{2}}{\nu+1} \implies Q(\mathscr{E}_{\lambda,\nu}) = 0\,, \\ \lambda &> \frac{1}{1+2^{-g(\nu)}} \implies Q(\mathscr{E}_{\lambda,\nu}) > 0\,, \end{aligned} \tag{7.52}$$

where

$$g(x) \coloneqq (x+1)\log_2(x+1) - x\log_2 x \qquad \forall\, x \geq 0. \tag{7.53}$$

7.3 A novel model of optical fibre with memory effects: the Delocalised Interaction Model

In this section we formulate a novel model of optical fibre with memory effects that we dub "Delocalised Interaction Model" (DIM). Consider an optical fibre with length L and transmissivity λ. We can imagine the optical fibre as a composition of $M \in \mathbb{N}$ infinitesimal optical fibres, each with length L/M and transmissivity $\lambda^{1/M}$ with $M \to \infty$. The signals transmitted by Alice, which propagate through the optical fibre, are single-mode of electromagnetic radiation with a definite frequency and polarisation. To model the noise affecting each signal, we employ the following approach. For all $j = 1, 2, \ldots, M$, the jth infinitesimal optical fibre is represented as a beam splitter with transmissivity $\lambda^{1/M}$ that interacts with a single-mode environment denoted as $E_1^{(j)}$. Additionally, each environment $E_1^{(j)}$ is influenced by both the output signal of the composition of the first $(j-1)$ infinitesimal optical fibres via the beam splitter interaction $U_{\lambda^{1/M}}$, and a remote reservoir R_j. The remote reservoir solely interacts with $E_1^{(j)}$ and attempts to reset its state to τ_ν through a thermalisation process, as illustrated in Fig. 7.5.

We describe the thermalisation process caused by the reservoir R_j on the environment $E_1^{(j)}$ as a thermal attenuator $\mathscr{E}_{\mu,\nu}$ with transmissivity $\mu \in [0,1]$ and thermal noise $\nu \geq 0$. The transmissivity μ serves as a *memory parameter* that relates to the time interval δt between consecutive signals from Alice and the characteristic thermalisation time t_E. Specifically, we can define $\mu \coloneqq \exp(-\delta t/t_E)$ as a possible relationship between these quantities.

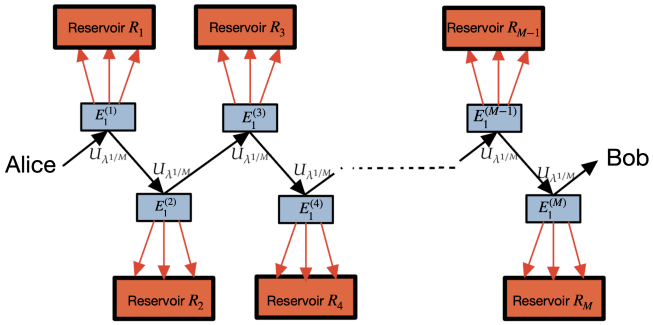


Figure 7.5: Depiction of our model of optical fibre with memory effects, dubbed "Delocalised Interaction Model" (DIM). As Alice's input signal travels across the optical fibre, it interacts with each of the M single-mode environments via the beam splitter unitary $U_{\lambda^{1/M}}$. For each $j = 1, 2, \ldots, M$, the environment $E_1^{(j)}$ undergoes a thermalisation process induced by the reservoir R_j, which attempts to restore the state of $E_1^{(j)}$ to its initial thermal state τ_ν. If the time interval between consecutive input signals is much longer than the thermalisation timescale, the memoryless assumption holds, and our model simplifies to that of a thermal attenuator $\mathscr{E}_{\lambda,\nu}$. However, in the opposite regime, where the time interval is shorter than the thermalisation timescale, each environment does not have sufficient time to return to its initial thermal state between consecutive signals. Consequently, in this case, when an input signal interacts with the environments $\{E_1^{(j)}\}_{j=1,2,\ldots,M}$, the state of each environment is no longer a thermal state τ_ν, but depends on the previously transmitted input signals, thereby introducing memory effects into the system.

Let us consider the special case of $\mu = 0$ to further enhance our understanding. In this case, the property $\mathscr{E}_{\mu=0,\nu}(\rho) = \tau_\nu$ holds for all single-mode states ρ. This means that when an input signal interacts with any environment $E_1^{(j)}$, the state of $E_1^{(j)}$ just before the interaction is a thermal state τ_ν, regardless of the previously sent input signals. As a result, in this case, we can represent each infinitesimal optical fibre as a thermal attenuator $\mathscr{E}_{\lambda^{1/M},\nu}$ with transmissivity $\lambda^{1/M}$. Finally, according to the composition rule (7.46), the composition of M thermal attenuators with transmissivity $\lambda^{1/M}$ is equal to a single thermal attenuator with transmissivity λ.

On the other hand, when $\mu > 0$, the environments $\{E_1^{(j)}\}_{j=1,2,\ldots,M}$ deviate from the initial thermal state $\tau_\nu^{\otimes M}$ at the moment of the interaction with Alice's signal, and their state becomes dependent on the previously transmitted input signals. This indicates that when $\mu > 0$, memory effects come into play. The interaction of each input signal with the environments introduces a memory effect, causing the state of the environments to be influenced by the history of the transmitted signals. Thus, the presence of non-zero μ leads to the emergence of memory effects in the system.

Now, let us define the quantum channel describing our model of optical fibre with memory effects. In our model, fixed $\lambda \in [0, 1]$, $\nu \geq 0$, and $\mu \in [0, 1]$, the output state of $n \in \mathbb{N}$ uses of the optical fibre corresponds to the output of the n-mode quantum channel $\Phi_{\lambda,\mu,\nu}^{(M,n)}$ whose interferometric representation is given in Fig. 7.6 (see also Fig. 7.2(c) in the main text), in the *strong* limit $M \to \infty$. In the forthcoming Theorem 80 we demonstrate the existence of the following strong limit:

$$\Phi_{\lambda,\mu,\nu}^{(n)} := \lim_{M\to\infty} \Phi_{\lambda,\mu,\nu}^{(M,n)} \qquad \text{(strongly)}\,, \tag{7.54}$$

and we give an explicit expression for it. The family of quantum channels $\{\Phi_{\lambda,\mu,\nu}^{(n)}\}_{n\in\mathbb{N}}$ defined in (7.54) is a quantum memory channel [12], which completely defines our new model of optical fibre with memory effects, dubbed "Delocalised Interaction Model" (DIM). In the DIM, when Alice transmits an n-mode state $\rho^{(n)} \in \mathfrak{S}(L^2(\mathbb{R})^{\otimes n})$ through n uses of the optical fibre, Bob receives n signals in the n-mode state $\Phi_{\lambda,\mu,\nu}^{(n)}(\rho^{(n)})$.

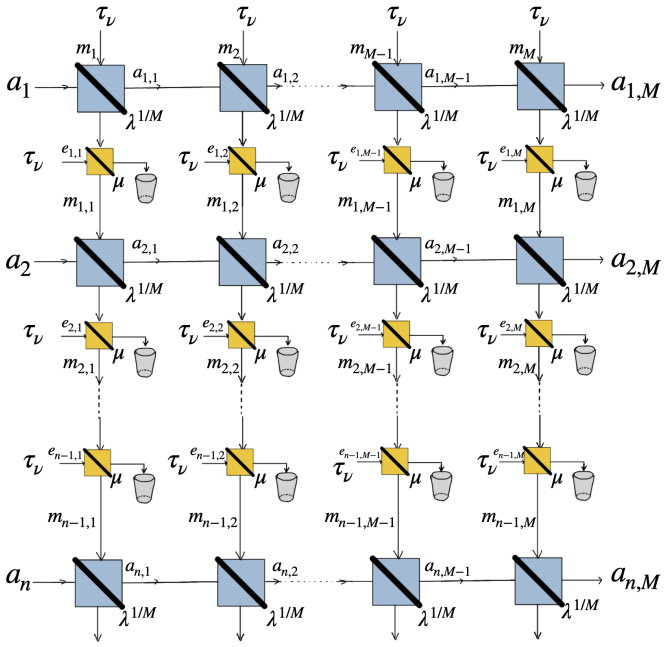


Figure 7.6: Depiction of the n-mode quantum channel $\Phi^{(M,n)}_{\lambda,\mu,\nu}$, which defines the "Delocalised Interaction Model" (DIM). Here, $M \in \mathbb{N}$ is the number of infinitesimal optical fibres, $n \in \mathbb{N}$ is the number of uses of the optical fibre, λ is the transmissivity of the optical fibre, ν is the thermal noise, and μ is the memory parameter. For all $i = 1, 2, \ldots, n$, the ith row of the figure, which consists of M blue beam splitters of transmissivity $\lambda^{1/M}$, represents the ith use of the optical fibre. For all $j = 1, 2, \ldots, M$, the jth column of the figure, which consists of n blue splitters of transmissivity $\lambda^{1/M}$ and n yellow beam splitters of transmissivity μ, represents the jth infinitesimal optical fibre. The yellow beam splitters, coupling the environments with the thermal state τ_ν, represent the thermalisation process $\mathscr{E}_{\mu,\nu}$. If $\mu = 0$ then $\Phi^{(M,n)}_{\lambda,\mu=0,\nu} = \mathscr{E}^{\otimes n}_{\lambda,\nu}$, meaning that the memoryless scenario is recovered. In the Heisenberg representation, $\Phi^{(M,n)}_{\lambda,\mu,\nu}$ maps the annihilation operator a_i (depicted on the left) of the ith input signal into the annihilation operators $a_{i,M}$ (depicted on the right) of the ith output signal for all $i = 1, 2, \ldots, n$. Our model of optical fibre with memory effects coincides with the *strong* limit $\Phi^{(n)}_{\lambda,\mu,\nu} = \lim_{M\to\infty} \Phi^{(M,n)}_{\lambda,\mu,\nu}$.

Theorem 80. *Let $\lambda \in [0,1]$, $\nu \geq 0$, $\mu \in [0,1]$, and $n \in \mathbb{N}$. The strong limit $\Phi^{(n)}_{\lambda,\mu,\nu} := \lim_{M\to\infty} \Phi^{(M,n)}_{\lambda,\mu,\nu}$ exists and it reads*

$$\Phi^{(n)}_{\lambda,\mu,\nu} = \mathscr{U}_{O_2^T} \circ \left(\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu} \right) \circ \mathscr{U}_{O_1} \,, \tag{7.55}$$

i.e.

$$\lim_{M\to\infty} \left\| \Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)}) - \mathscr{U}_{O_2^T} \circ \left(\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu} \right) \circ \mathscr{U}_{O_1}(\rho^{(n)}) \right\|_1 = 0 \qquad \forall\, \rho^{(n)} \in \mathfrak{S}(L^2(\mathbb{R}^{\otimes n})) \,. \tag{7.56}$$

Here,

- *$\mathscr{U}_{O_2^T}$ and $\mathscr{U}_{O_1}$ are passive unitary operators on $L^2(\mathbb{R}^{\otimes n})$ (see Remark 79 for the definition of passive unitary operators) associated with the orthogonal matrices O_2^T and O_1 (defined below), respectively.*

- *The transmissivities $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\ldots,n}$ of the n-mode thermal attenuator in (7.55) and the $n \times n$ orthogonal matrices O_2^T and O_1 are defined by means of the singular value decomposition of the $n \times n$ real matrix $\bar{A}^{(n,\lambda,\mu)}$ (defined below):*

$$\bar{A}^{(n,\lambda,\mu)} = O_2^T D O_1 \,, \tag{7.57}$$

where $D \coloneqq \mathrm{diag}(\sqrt{\eta_1^{(n,\lambda,\mu)}}, \sqrt{\eta_2^{(n,\lambda,\mu)}}, \ldots, \sqrt{\eta_n^{(n,\lambda,\mu)}})$ *is an* $n \times n$ *positive diagonal matrix with*

$$1 \geq \eta_n^{(n,\lambda,\mu)} \geq \eta_{n-1}^{(n,\lambda,\mu)} \geq \ldots \geq \eta_1^{(n,\lambda,\mu)} \geq 0\,. \tag{7.58}$$

- *For each* $i,k \in \{1,2,\ldots,n\}$*, the* (i,k) *element of the* $n \times n$ *real matrix* $\bar{A}^{(n,\lambda,\mu)}$ *is defined as*

$$\bar{A}_{i,k}^{(n,\lambda,\mu)} = \Theta(i-k)\,\sqrt{\lambda}\mu^{\frac{i-k}{2}} L_{i-k}^{(-1)}(-\ln\lambda)\,. \tag{7.59}$$

where $\Theta : \mathbb{R} \to \mathbb{R}$ *is the Heaviside Theta function defined as*

$$\Theta(x) \coloneqq \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{7.60}$$

and where $\{L_m^{(-1)}\}_{m\in\mathbb{N}}$ *are the generalised Laguerre polynomials defined as*

$$\begin{aligned} L_m^{(-1)}(x) &\coloneqq \sum_{l=1}^{m} \binom{m-1}{l-1} \frac{(-x)^l}{l!} \qquad \forall\, m \in \mathbb{N}^+\,, \\ L_0^{(-1)}(x) &\coloneqq 1\,, \end{aligned} \tag{7.61}$$

for all $x \in \mathbb{R}$*.*

(7.56) *establishes that the quantum memory channel* $\{\Phi_{\lambda,\mu,\nu}^{(n)}\}_{n\in\mathbb{N}}$ *is unitarily equivalent to* $\{\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\}_{n\in\mathbb{N}}$*. In particular, their (energy-constrained and unconstrained) capacities coincide.*

Before presenting the proof of Theorem 80, let us understand its meaning. Suppose Alice prepares an n-signal state $\rho^{(n)} \in \mathfrak{S}(L^2(\mathbb{R})^{\otimes n})$, applies the passive transformation $\mathscr{U}_{O_1^T}$, and then sends each of the n signals through the optical fibre. After receiving the signals, Bob applies the passive transformation $\mathscr{U}_{O_2}$. Consequently, Bob receives the state $\mathscr{U}_{O_2} \circ \Phi_{\lambda,\mu,\nu}^{(n)} \circ \mathscr{U}_{O_1^T}(\rho^{(n)})$ which can be expresses as

$$\mathscr{U}_{O_2} \circ \Phi_{\lambda,\mu,\nu}^{(n)} \circ \mathscr{U}_{O_1^T}(\rho^{(n)}) = \bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}(\rho^{(n)})\,, \tag{7.62}$$

thanks to Theorem 80. Therefore, by employing appropriate encoding and decoding passive unitary transformations, Alice and Bob can effectively communicate via

$$\left\{\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right\}_{n\in\mathbb{N}}\,, \tag{7.63}$$

instead of via $\{\Phi_{\lambda,\mu,\nu}^{(n)}\}_{n\in\mathbb{N}}$. To summarise, the quantum memory channel $\Phi_{\lambda,\mu,\nu}^{(n)}$, which models n uses of the optical fibre, is unitarily equivalent to a tensor product of n distinct thermal attenuators. These attenuators have the same thermal noise ν but differ in their transmissivities. This equivalence allows us to understand the behaviour of the quantum memory channel in terms of individual thermal attenuators operating on each signal independently. The transmissivities $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\ldots,n}$ can be calculated via the singular value decomposition of the matrix reported in (7.59). Notably, each element of this matrix depends solely on the difference between the row and column indices, making it a *Toeplitz matrix*. In Section 7.5 we will review important properties of Toeplitz matrices. Fortunately, these matrices allow for the calculation of the asymptotic distribution of their singular values as the dimension approaches infinity. Leveraging this observation, we will be able to analytically determine the transmissivities $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\ldots,n}$ as the number of uses of the optical fibre, n, approaches infinity.

Now, we are ready to proceed with the proof of Theorem 80, which follows and generalises the methods used in [16–18]

Proof of Theorem 80. In this proof, we will adopt the notation introduced in Fig. 7.6 and in Fig. 7.2(c) in the main text. Specifically, $n \in \mathbb{N}$ represents the number of uses of the optical fibre, while $M \in \mathbb{N}$ represents the number of infinitesimal optical fibres. Moreover, for each $i = 1, 2, \dots, n$ the annihilation operator associated with the ith input signal is denoted as a_i, and for each $j = 1, 2, \dots, M$ the annihilation operator associated with the jth single-mode environment $E_1^{(j)}$ is denoted as $b_1^{(j)}$. In addition, as shown in Fig. 7.2(c) in the main text, for each $i = 2, 3, \dots n$ and $j = 1, 2, \dots, M$ we denote as $E_i^{(j)}$ the single-mode environment associated with the thermal attenuator $\mathscr{E}_{\mu,\nu}$ that represents the thermalisation process acting on the jth environment $E_1^{(j)}$ right before the ith use of the fibre. As shown in Fig. 7.6, the annihilation operator of the environment $E_i^{(j)}$ is denoted as $b_i^{(j)}$.

Let $U_{\lambda,\mu}^{(n,M)}$ be the unitary operator implementing the physical representation of $\Phi_{\lambda,\mu,\nu}^{(M,n)}$ in Fig. 7.6:

$$\Phi_{\lambda,\mu,\nu}^{(M,n)}(\rho^{(n)}) = \mathrm{Tr}_E \left[U_{\lambda,\mu}^{(n,M)} \left(\rho^{(n)} \otimes \tau_\nu^{\otimes nM} \right) U_{\lambda,\mu}^{(n,M)\dagger} \right] \qquad \forall\, \rho^{(n)} \in \mathfrak{S}(L^2(\mathbb{R})^{\otimes n}) \,, \tag{7.64}$$

where Tr_E denotes the partial trace over all the environment systems ($E_i^{(j)}$ for all $i = 1, 2, \dots, n$ and all $j = 1, 2, \dots, M$). In addition, as shown in Fig. 7.6, for each $i = 1, 2, \dots, n$ let $a_{i,M} := U_{\lambda,\mu}^{(n,M)\dagger} a_i U_{\lambda,\mu}^{(n,M)}$ be the output of the ith annihilation operator in Heisenberg representation.

Since the unitary $U_{\lambda,\mu}^{(n,M)}$ is composed of only beam splitters (see Fig. 7.6), (7.40) implies that for all $i = 1, 2, \dots, M$ the output annihilation operator $a_{i,M}$ can be written as

$$a_{i,M} = \sum_{h=1}^{n} A_{i,h}^{(M,n,\lambda,\mu)} \, a_h + \sum_{l=1}^{M} \sum_{h=1}^{n} E_{l,i,h}^{(M,n,\lambda,\mu)} \, b_h^{(l)} \,, \tag{7.65}$$

where $A_{i,h}^{(M,n,\lambda,\mu)}$ and $E_{l,i,h}^{(M,n,\lambda,\mu)}$ are suitable real numbers such that

$$[a_{i,M}, (a_{j,M})^\dagger] = \delta_{i,j} \mathbb{1} \,. \tag{7.66}$$

Note that (7.65) holds because the transformations in (7.40) maps annihilation operators into annihilation operators (without introducing any creation operator). By defining the following vectors of annihilation operators

$$\begin{aligned} \mathbf{a} &:= (a_1, \, a_2, \dots, \, a_n)^T \,, \\ \mathbf{a_M} &:= (a_{1,M}, \, a_{2,M}, \dots, \, a_{n,M})^T \,, \\ \mathbf{b} &:= (b_1^{(1)}, \, b_1^{(2)}, \dots, \, b_1^{(M)}, \, b_2^{(1)}, \, b_2^{(2)}, \dots, \, b_2^{(M)}, \dots, \, b_n^{(1)}, \, b_n^{(2)}, \dots, \, b_n^{(M)})^T \,, \end{aligned} \tag{7.67}$$

one can rewrite (7.65) as

$$\mathbf{a_M} = A^{(M,n,\lambda,\mu)} \mathbf{a} + E^{(M,n,\lambda,\mu)} \mathbf{b} \,, \tag{7.68}$$

where $A^{(M,n,\lambda,\mu)}$ is an $n \times n$ real matrix and $E^{(M,n,\lambda,\mu)}$ is an $n \times nM$ real matrix which satisfy

$$A^{(M,n,\lambda,\mu)} A^{(M,n,\lambda,\mu)^T} + E^{(M,n,\lambda,\mu)} E^{(M,n,\lambda,\mu)^T} = \mathbb{1}_{n\times n} \,, \tag{7.69}$$

thanks to (7.66). Here, $\mathbb{1}_{n\times n}$ denotes the $n \times n$ identity matrix.

Below we will show that the limit of the matrix element $\lim_{M\to\infty} A_{i,h}^{(M,n,\lambda,\mu)}$ does exist and it is given by (7.59) for all $i, h = 1, 2, \dots n$. Hence, we can define the matrix $\bar{A}^{(n,\lambda,\mu)}$ whose (i, h) element is equal to

$$\bar{A}_{i,h}^{(n,\lambda,\mu)} := \lim_{M\to\infty} A_{i,h}^{(M,n,\lambda,\mu)} \,. \tag{7.70}$$

Now, we will demonstrate the validity of (7.56). For this purpose, let $\rho^{(n)} \in \mathfrak{S}(L^2(\mathbb{R})^{\otimes n})$ be an n-mode input state. The characteristic function of $\Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)})$ can be calculated as

$$\begin{aligned}
\chi_{\Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)})}(\mathbf{z}) &:= \mathrm{Tr}\left[\Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)})\, e^{\mathbf{a}^\dagger \mathbf{z} - \mathbf{z}^\dagger \mathbf{a}}\right] \\
&\overset{\text{(i)}}{=} \chi_{\rho^{(n)}}\left(A^{(M,n,\lambda,\mu)^T}\mathbf{z}\right) \chi_{\tau_\nu^{\otimes nM}}\left(E^{(M,n,\lambda,\mu)^T}\mathbf{z}\right) \\
&\overset{\text{(ii)}}{=} \chi_{\rho^{(n)}}\left(A^{(M,n,\lambda,\mu)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger E^{(M,n,\lambda,\mu)} E^{(M,n,\lambda,\mu)^T}\mathbf{z}} \\
&\overset{\text{(iii)}}{=} \chi_{\rho^{(n)}}\left(A^{(M,n,\lambda,\mu)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger (\mathbb{1}_{n\times n} - A^{(M,n,\lambda,\mu)} A^{(M,n,\lambda,\mu)^T})\mathbf{z}} \qquad \forall\, \mathbf{z} \in \mathbb{C}^n\,,
\end{aligned} \tag{7.71}$$

where: in (i) we exploited (7.64) and (7.68); in (ii) we used the expression in (7.44) of the characteristic function of the thermal state τ_ν; in (iii) we just exploited (7.69). Since $\lim_{M\to\infty} A^{(M,n,\lambda,\mu)}\mathbf{z} = \bar{A}^{(n,\lambda,\mu)}\mathbf{z}$ (thanks to (7.70)) and since every characteristic function is continuous [24, 54], it holds that

$$\lim_{M\to\infty} \chi_{\Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)})}(\mathbf{z}) = \chi_{\rho^{(n)}}\left(\bar{A}^{(n,\lambda,\mu)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger (\mathbb{1}_{n\times n} - \bar{A}^{(n,\lambda,\mu)}\bar{A}^{(n,\lambda,\mu)^T})\mathbf{z}}\,. \tag{7.72}$$

By applying the singular value decomposition, $\bar{A}^{(n,\lambda,\mu)}$ can be written as

$$\bar{A}^{(n,\lambda,\mu)} = O_2^T D O_1\,, \tag{7.73}$$

where O_1 and O_2 are $n \times n$ real orthogonal matrices, $D = \mathrm{diag}(s_1^{(n,\lambda,\mu)}, s_2^{(n,\lambda,\mu)}, \ldots, s_n^{(n,\lambda,\mu)})$ is an $n \times n$ positive diagonal matrix with

$$1 \geq s_n^{(n,\lambda,\mu)} \geq s_{n-1}^{(n,\lambda,\mu)} \geq \ldots \geq s_1^{(n,\lambda,\mu)} \geq 0\,, \tag{7.74}$$

where the fact that $s_n^{(n,\lambda,\mu)} \leq 1$ follows from (7.69). Let us define for all $i = 1, 2, \ldots, n$ the ith *transmissivity* $\eta_i^{(n,\lambda,\mu)}$ as the square of the ith diagonal element of D, i.e.

$$\eta_i^{(n,\lambda,\mu)} := (s_i^{(n,\lambda,\mu)})^2\,. \tag{7.75}$$

Note that $\eta_i^{(n,\lambda,\mu)}$ is a proper transmissivity because $0 \leq (s_i^{(n,\lambda,\mu)})^2 \leq 1$. In addition, for all $\mathbf{z} \in \mathbb{C}^n$ it holds that

$$\begin{aligned}
\lim_{M\to\infty} \chi_{\Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)})}(\mathbf{z}) &\overset{\text{(i)}}{=} \chi_{\rho^{(n)}}\left(O_1^T D O_2 \mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger (\mathbb{1}_{n\times n} - O_2^T D^2 O_2)\mathbf{z}} \\
&\overset{\text{(ii)}}{=} \chi_{\mathcal{U}_{O_1}(\rho^{(n)})}\left(D O_2 \mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger O_2^T (\mathbb{1}_{n\times n} - D^2) O_2 \mathbf{z}} \\
&\overset{\text{(iii)}}{=} \chi_{\mathcal{U}_{O_1}(\rho^{(n)})}\left(D O_2 \mathbf{z}\right) \chi_{\tau_\nu^{\otimes n}}\left(\sqrt{\mathbb{1}_{n\times n} - D^2} O_2 \mathbf{z}\right) \\
&\overset{\text{(iv)}}{=} \chi_{(\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}) \circ \mathcal{U}_{O_1}(\rho^{(n)})}\left(O_2 \mathbf{z}\right) \\
&\overset{\text{(v)}}{=} \chi_{\mathcal{U}_{O_2^T} \circ \left(\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right) \circ \mathcal{U}_{O_1}(\rho^{(n)})}\left(\mathbf{z}\right)\,,
\end{aligned} \tag{7.76}$$

where: in (i) we used (7.72) and (7.73); in (ii) we exploited Remark (79); (iii) is a consequence of the expression in (7.44) of the characteristic function of the thermal state τ_ν; (iv) comes from the expression in (7.45) of the characteristic function of the output state of the thermal attenuator; in (v) we exploited again Remark (79). Hence, (7.76) establishes that the characteristic function of $\Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)})_{M\in\mathbb{N}}$ converges pointwise for $M \to \infty$ to the characteristic function $\mathcal{U}_{O_2^T} \circ \left(\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right) \circ \mathcal{U}_{O_1}(\rho^{(n)})$. Consequently, Lemma (78) implies that

$$\lim_{M\to\infty} \left\| \Phi^{(M,n)}_{\lambda,\mu,\nu}(\rho^{(n)}) - \mathcal{U}_{O_2^T} \circ \left(\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right) \circ \mathcal{U}_{O_1}(\rho^{(n)}) \right\|_1 = 0\,, \tag{7.77}$$

i.e. we have proved (7.56).

Now, we only need to show that the limit $\lim_{M\to\infty} A_{i,h}^{(M,n,\lambda,\mu)}$ does exist and it is given by (7.59) for all $i,h=1,2,\dots n$. To proceed, we begin by calculating the elements of the matrix $A^{(M,n,\lambda,\mu)}$ defined in (7.68). Our goal is thus to express $a_{i,M}$ in terms of $a_1,a_2,\dots,a_n$.

In the following derivation, $m_{i,j}$ and $a_{i,j}$ denote the annihilation operators in Heisenberg representation introduced in Fig. 7.6. Specifically, $m_{i,j}$ denotes the Heisenberg evolution of the annihilation operator $b_1^{(j)}$ right before the $(i+1)$th use of the fibre. Moreover, $a_{i,j}$ denotes the Heisenberg evolution of the annihilation operator a_i of the ith input signal right after its transmission through the first j beam splitters of transmissivity $\lambda^{1/M}$. Additionally, we use the notations $a_{i,0}:=a_i$ and $m_{0,i}:=b_1^{(i)}$. By exploiting (7.40), one deduces that for all $i=1,2,\dots,n$ and all $j=1,2,\dots,M$ it holds that

$$\begin{aligned} a_{i,j} &= \sqrt{\lambda^{1/M}}\, a_{i,j-1} + \sqrt{1-\lambda^{1/M}}\, m_{i-1,j}\,, \\ m_{i,j} &= -\sqrt{1-\mu}\, b_{i+1}^{(j)} + \sqrt{\mu\lambda^{1/M}}\, m_{i-1,j} - \sqrt{\mu(1-\lambda^{1/M})}\, a_{i,j-1}\,. \end{aligned} \tag{7.78}$$

Consequently, $a_{i,M}$ can be expressed as a linear combination of the annihilation operators $\{a_i\}_i$ and $\{b_i^{(j)}\}_{i,j}$. In order to calculate the matrix $A^{(M,n,\lambda,\mu)}$, we only need to calculate the coefficients of $\{a_i\}_i$ in the expression of $a_{i,M}$, while we do not need to calculate the coefficients of $\{b_i^{(j)}\}_{i,j}$. Hence, we introduce the relation $A\simeq B$ between operators as follows:

$$A\simeq B \quad\Longleftrightarrow\quad A = B + \text{linear combination of } \{b_i^{(j)}\}_{i,j}. \tag{7.79}$$

Note that (7.78) implies that for all $i=1,2,\dots,n$ and all $j=1,2,\dots,M$ it holds that

$$a_{i,j} = \sqrt{\lambda^{j/M}}\, a_i + \sqrt{1-\lambda^{1/M}} \sum_{l=1}^{j} \lambda^{\frac{j-l}{2M}}\, m_{i-1,l} \tag{7.80}$$

and

$$\begin{aligned} m_{i,j} &\simeq \sqrt{\mu\lambda^{1/M}} m_{i-1,j} - \sqrt{\mu(1-\lambda^{1/M})} a_{i,j-1} \\ &= \sqrt{\mu\lambda^{1/M}} m_{i-1,j} - \sqrt{\mu(1-\lambda^{1/M})\lambda^{(j-1)/M}} a_i - \sqrt{\mu}(1-\lambda^{1/M}) \sum_{l=1}^{j-1} \lambda^{\frac{j-1-l}{2M}} m_{i-1,l}\,. \end{aligned} \tag{7.81}$$

From (7.80) we deduce that in order to express $a_{i,M}$ in terms of $\{a_i\}_i$, we now only need to express $\{m_{i,l}\}_{i,l}$ in terms of $\{a_i\}_i$ by solving the recurrence relation in (7.81). One can show that the latter implies that for all $i=1,2,\dots,n$ and all $j=2,3\dots,M$ it holds that

$$\begin{aligned} m_{i,1} &\simeq -\sqrt{\mu(1-\lambda^{1/M})} \sum_{l=1}^{i} \left(\mu\lambda^{1/M}\right)^{\frac{i-l}{2}}\, a_l\,, \\ m_{i,j} &\simeq \sqrt{\mu\lambda^{1/M}}\, m_{i-1,j} - \sqrt{\mu}\, m_{i-1,j-1} + \sqrt{\lambda^{1/M}}\, m_{i,j-1}\,. \end{aligned} \tag{7.82}$$

As a consequence of this, substituting terms $m_{l,j}$ with $l\le i$, through repeated application of the second equality, and finally dropping the term $m_{0,j}$, one can show that

$$m_{i,j} \simeq \lambda^{\frac{1}{2M}}\, m_{i,j-1} + (\lambda^{\frac{1}{2M}} - \lambda^{-\frac{1}{2M}}) \sum_{l=1}^{i-1} (\mu\lambda^{1/M})^{\frac{i-l}{2}}\, m_{l,j-1} \tag{7.83}$$

We stress that, here and in the following, summations are zero whenever the lower extreme is strictly larger than the upper extreme. By defining $\tilde m_{i,j} := \lambda^{-\frac{j}{2M}} (\mu\lambda^{1/M})^{-\frac{i}{2}}\, m_{i,j}$, the recurrence relation in (7.83) can be rewritten as

$$\tilde m_{i,j} \simeq \tilde m_{i,j-1} + (1-\lambda^{-\frac{1}{M}}) \sum_{l=1}^{i-1} \tilde m_{l,j-1}\,. \tag{7.84}$$

In order to solve this, let us introduce the following notation:

$$\mathscr{F}_{i,n}\left(\{\tilde{m}_{\cdot,1}\}\right) := \sum_{i_n=1}^{i-1} \sum_{i_{n-1}=1}^{i_n-1} \cdots \sum_{i_1=1}^{i_2-1} \tilde{m}_{i_1,1}\,. \tag{7.85}$$

By noting that the cardinality of the set $\{(i_2,\ldots,i_n)\in\mathbb{N}^{n-1} \,:\, i>i_n>i_{n-1}>\ldots>i_2>d\}$ is $\binom{i-d-1}{n-1}$, we deduce that

$$\mathscr{F}_{i,n}\left(\{\tilde{m}_{\cdot,1}\}\right) = \sum_{d=1}^{i-n} \binom{i-d-1}{n-1} \tilde{m}_{d,1}\,. \tag{7.86}$$

In addition, note that

$$\mathscr{F}_{i,n}\left(\{\tilde{m}_{\cdot,1}\}\right) = 0 \qquad \text{if } i\le n\,, \tag{7.87}$$

and

$$\sum_{l=1}^{i-1} \mathscr{F}_{l,n}\left(\{\tilde{m}_{\cdot,1}\}\right) = \mathscr{F}_{i,n+1}\left(\{\tilde{m}_{\cdot,1}\}\right)\,. \tag{7.88}$$

As a consequence, by exploiting (7.84), one can show by induction that for all $i=1,2,\ldots,n$ and all $j=1,2,\ldots,M$ it holds that

$$\tilde{m}_{i,j} \simeq \tilde{m}_{i,1} + \sum_{l=1}^{j-1} \binom{j-1}{l} (1-\lambda^{-1/M})^l \mathscr{F}_{i,l}\left(\{\tilde{m}_{\cdot,1}\}\right)\,, \tag{7.89}$$

and from this one can obtain

$$m_{i,j} \simeq \sum_{k=1}^{i} \mu^{\frac{i-k+1}{2}} \lambda^{\frac{j+i-k+1}{2M}} \sqrt{1-\lambda^{1/M}} \left[\sum_{l=0}^{\min(i-k,j-1)} \binom{j-1}{l} \binom{i-k}{l} (1-\lambda^{-1/M})^l \right] a_k \tag{7.90}$$

Consequently, (7.80) implies that

$$\begin{aligned} a_{i,M} &= \sqrt{\lambda} a_i + \sqrt{1-\lambda^{1/M}} \sum_{j=1}^{M} \lambda^{\frac{M-j}{2M}} m_{i-1,j} \\ &\simeq \sum_{k=1}^{n} \left[\sqrt{\lambda}\delta_{i,k} - \Theta(i-k-1)\,\mu^{\frac{i-k}{2}} \lambda^{\frac{M+i-k}{2M}} (1-\lambda^{1/M}) \sum_{j=1}^{M} \sum_{l=0}^{\min(i-k-1,j-1)} \binom{j-1}{l}\binom{i-k-1}{l} (1-\lambda^{-1/M})^l \right] a_k, \end{aligned} \tag{7.91}$$

where we have introduced the Heaviside function $\Theta(x)$ defined as $\Theta(x)=1$ if $x\ge 0$, and $\Theta(x)=0$ if $x<0$. Hence, we deduce that for all $i,k\in\{1,2,\ldots,n\}$ the (i,k) matrix element of $A^{(M,n,\lambda,\mu)}$ is

$$A^{(M,n,\lambda,\mu)}_{i,k} = \sqrt{\lambda}\delta_{i,k} - \Theta(i-k-1)\,\mu^{\frac{i-k}{2}} \lambda^{\frac{M+i-k}{2M}} (1-\lambda^{1/M}) \sum_{j=1}^{M} \sum_{l=0}^{\min(i-k-1,j-1)} \binom{j-1}{l}\binom{i-k-1}{l} (1-\lambda^{-1/M})^l. \tag{7.92}$$

The sum over j can be split in two parts, obtaining

$$\begin{aligned} &(1-\lambda^{1/M}) \sum_{j=1}^{M} \sum_{l=0}^{\min(i-k-1,j-1)} \binom{j-1}{l}\binom{i-k-1}{l} (1-\lambda^{-1/M})^l \\ &= (1-\lambda^{1/M}) \sum_{j=1}^{i-k} \sum_{l=0}^{j-1} \binom{j-1}{l}\binom{i-k-1}{l} (1-\lambda^{-1/M})^l + (1-\lambda^{1/M}) \sum_{j=i-k+1}^{M} \sum_{l=0}^{i-k-1} \binom{j-1}{l}\binom{i-k-1}{l} (1-\lambda^{-1/M})^l. \end{aligned} \tag{7.93}$$

The limit for $M\to\infty$ of the first piece is zero, while for the second we use that

$$\lim_{M\to\infty} (1-\lambda^{1/M}) \sum_{j=i-k+1}^{M} \binom{j-1}{l} (1-\lambda^{-1/M})^l = -\frac{(\ln\lambda)^{l+1}}{(l+1)!} \tag{7.94}$$

to obtain

$$\bar{A}_{i,k}^{(n,\lambda,\mu)} := \lim_{M\to\infty} A_{i,k}^{(M,n,\lambda,\mu)} = \sqrt{\lambda}\left[\delta_{i,k} + \Theta(i-k-1)\,\mu^{\frac{i-k}{2}} \sum_{l=1}^{i-k} \binom{i-k-1}{l-1} \frac{(\ln\lambda)^l}{l!}\right]. \tag{7.95}$$

By defining the generalised Laguerre polynomial $L_m^{(-1)}(x) := \sum_{l=1}^{m} \binom{m-1}{l-1} \frac{(-x)^l}{l!}$ for all $x \in \mathbb{R}$ and $m \in \mathbb{N}^+$ and $L_0^{(-1)}(x) := 1$, for all $i,k \in \{1,2,\ldots,n\}$ the (i,k) matrix element of $\bar{A}^{(n,\lambda,\mu)}$ in (7.95) can be rewritten as

$$\bar{A}_{i,k}^{(n,\lambda,\mu)} = \Theta(i-k)\,\sqrt{\lambda}\mu^{\frac{i-k}{2}} L_{i-k}^{(-1)}(-\ln\lambda)\,. \tag{7.96}$$

□

7.4 Problems of the "Localised" interaction model solved by the "Delocalised" one

Memory effects in optical fibres have been studied in [10, 11, 16–18]. As explained in the main text, the model examined in [10, 11, 16–18] corresponds to a specific case of the DIM we have presented here, specifically the case in which we set $M = 1$ (see Section 7.3 for the meaning of M). This corresponds to the configuration depicted in Fig.7.2(b) in the main text, where the interaction between the signal and the fibre is localised. For this reason, in the main text we dubbed the model analysed in [10, 11, 16–18] as "Localised Interaction Model" (LIM). The memory quantum channel characterising the LIM is $\{\Phi_{\lambda,\mu,\nu}^{(1,n)}\}_{n\in\mathbb{N}}$ (see Section 7.3 for its definition).

However, it is important to note that the LIM is not realistic because it fails to satisfy the following two essential properties that any "reasonable" model of an optical fibre should possess:

- *Property 1*: When the transmissivity λ is exactly equal to zero, no information can be transmitted;
- *Property 2*: The model is consistent under concatenation of optical fibres. In other words, the concatenation of k optical fibres, each with transmissivity $\lambda_1, \lambda_2, \ldots, \lambda_k$ and identical thermal noise ν, results in another optical fibre with a transmissivity equal to the product of the individual transmissivities $\lambda_1\lambda_2\ldots\lambda_k$, and with the same thermal noise ν.

Remark 81. *The LIM, which is characterised by the quantum memory channel $\{\Phi_{\lambda,\mu,\nu}^{(1,n)}\}_{n\in\mathbb{N}}$, fails to satisfy Property 1.*

Proof. The LIM is associated with the channel $\{\Phi_{\lambda,\mu,\nu}^{(1,n)}\}_{n\in\mathbb{N}}$ that transmits information even if λ is exactly zero. Indeed, for $\lambda = 0$ the channel $\Phi_{0,\mu,\nu}^{(1,n)}$ can be expressed in terms of the *quantum shift channel of order* 1, the thermal attenuator $\mathscr{E}_{\mu,\nu}$, and the unitary phase space inversion operation $\mathscr{V}(\cdot) := (-1)^{a^\dagger a} \cdot (-1)^{a^\dagger a}$. Specifically, if the state of the ith input signal is ρ, then the state of the $(i+1)$th output signal is $\mathscr{V} \circ \mathscr{E}_{\mu,\nu}(\rho)$, which is not constant with respect to ρ for $\mu > 0$. □

In particular, from the above proof we conclude that the LIM would lead to the paradoxical result that an optical fibre with zero transmissivity and a memory parameter $\mu > 0$ is equivalent to a memoryless optical fibre with transmissivity $\mu > 0$ with a certain time-delay δt (e.g. δt can be determined in terms of μ by $\mu = \exp(-\delta t/t_E)$), provided the output of the first signal is discarded.

Furthermore, the DIM with a finite value of M is also unrealistic because it fails to satisfy Property 1, as we now demonstrate.

Remark 82. *The DIM with a finite value of M, which is characterised by the quantum memory channel $\{\Phi_{\lambda,\mu,\nu}^{(M,n)}\}_{n\in\mathbb{N}}$, does not satisfy Property 1.*

Proof. The model with a finite value of M is associated with the channel $\{\Phi^{(M,n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$ that transmits information even if λ is exactly zero. Indeed, for $\lambda = 0$ the channel $\Phi^{(M,n)}_{0,\mu,\nu}$ can be expressed in terms of the *quantum shift channel of order* M, the thermal attenuator $\mathscr{E}_{\mu,\nu}$, and the unitary phase space inversion operation $\mathscr{V}^{M \bmod 2}$. Specifically, if the state of the ith input signal is ρ, then the state of the $(i+M)$th output signal is $\mathscr{V}^{M \bmod 2} \circ \mathscr{E}_{\mu,\nu}(\rho)$, which is not constant with respect to ρ for $\mu > 0$. □

From the above proof we conclude that the model with a finite value of M would lead to the paradoxical result that an optical fibre with zero transmissivity and a memory parameter $\mu > 0$ is equivalent to a memoryless optical fibre with transmissivity $\mu > 0$ with a time-delay $M\,\delta t$, provided the output of the first M signals is discarded.

Furthermore, the DIM with a finite value of M also fails to satisfy Property 2, meaning that it is not consistent under the composition of optical fibres. This can be proved as follows.

Remark 83. *The DIM with a finite value of M, which is characterised by the quantum memory channel $\{\Phi^{(M,n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$, does not satisfy Property 2.*

Proof. One can show that $\Phi^{(M,n)}_{\lambda_1,\mu,\nu} \circ \Phi^{(M,n)}_{\lambda_2,\mu,\nu}$ is not equal to $\Phi^{(M,n)}_{\lambda_1\lambda_2,\mu,\nu}$ in general. Indeed, if $\lambda_1 = \lambda_2 = 0$, then the channel $\Phi^{(M,n)}_{0,\mu,\nu} \circ \Phi^{(M,n)}_{0,\mu,\nu}$ is expressed in terms of a *quantum shift channel of order* $2M$, while the channel $\Phi^{(M,n)}_{0,\mu,\nu}$ is expressed in terms of a *quantum shift channel of order* M. □

Let us now show that our model of optical fibre with memory effects, which corresponds to the DIM with $M \to \infty$, does satisfy both Property 1 and Property 2. We recall that such a model is characterised by the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$, which is defined via the following strong limit

$$\Phi^{(n)}_{\lambda,\mu,\nu} := \lim_{M\to\infty} \Phi^{(M,n)}_{\lambda,\mu,\nu} \qquad (\text{strongly})\,, \tag{7.97}$$

meaning that we are taking into account the model in Fig. 7.5 and in Fig. 7.6 with an infinite number of environments.

Remark 84. *The DIM with $M \to \infty$, which is characterised by the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$, does satisfy Property 1.*

Proof. Fix an n-mode state ρ. For $\lambda = 0$, if $M \geq n$ it holds that $\Phi^{(M,n)}_{0,\mu,\nu}(\rho) = \tau_\nu^{\otimes n}$. Consequently, we have $\Phi^{(n)}_{0,\mu,\nu}(\rho) = \tau_\nu^{\otimes n}$, indicating that the output state of the fibre is independent of the input state. This implies that the fibre does not transmit any information. □

Remark 85. *The DIM with $M \to \infty$, which is characterised by the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$, does satisfy Property 2. Mathematically, this means that for all $n \in \mathbb{N}$ and all $\lambda_1, \lambda_2 \in [0,1]$ it holds that*

$$\Phi^{(n)}_{\lambda_1,\mu,\nu} \circ \Phi^{(n)}_{\lambda_2,\mu,\nu} = \Phi^{(n)}_{\lambda_1\lambda_2,\mu,\nu} \tag{7.98}$$

Proof. The validity of Property 2 becomes intuitive when observing Fig. 7.6 in the limit as $M \to \infty$. Nevertheless, let us provide a rigorous proof to establish its validity. Let us fix an n-mode state $\rho^{(n)}$. By exploiting (7.72), the characteristic function of $\Phi^{(n)}_{\lambda_1\lambda_2,\mu,\nu}(\rho^{(n)})$ is given by

$$\chi_{\Phi^{(n)}_{\lambda_1\lambda_2,\mu,\nu}(\rho^{(n)})}(\mathbf{z}) = \chi_{\rho^{(n)}}\left(\bar{A}^{(n,\lambda_1\lambda_2,\mu)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger(\mathbb{1}_{n\times n} - \bar{A}^{(n,\lambda_1\lambda_2,\mu)}\bar{A}^{(n,\lambda_1\lambda_2,\mu)^T})\mathbf{z}} \qquad \forall\,\mathbf{z}\in\mathbb{C}^n\,. \tag{7.99}$$

On the other hand, the characteristic function of $\Phi^{(n)}_{\lambda_1,\mu,\nu} \circ \Phi^{(n)}_{\lambda_2,\mu,\nu}(\rho^{(n)})$ is given by

$$\begin{aligned}
&\chi_{\Phi^{(n)}_{\lambda_1,\mu,\nu}\circ\Phi^{(n)}_{\lambda_2,\mu,\nu}(\rho^{(n)})}(\mathbf{z}) = \chi_{\Phi^{(n)}_{\lambda_2,\mu,\nu}(\rho^{(n)})}\left(\bar{A}^{(n,\lambda_1,\mu)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger(\mathbb{1}_{n\times n} - \bar{A}^{(n,\lambda_1,\mu)}\bar{A}^{(n,\lambda_1,\mu)^T})\mathbf{z}} \\
&= \chi_{\rho^{(n)}}\left(\bar{A}^{(n,\lambda_2,\mu)^T}\bar{A}^{(n,\lambda_1,\mu)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger\bar{A}^{(n,\lambda_1,\mu)}(\mathbb{1}_{n\times n} - \bar{A}^{(n,\lambda_2,\mu)}\bar{A}^{(n,\lambda_2,\mu)^T})\bar{A}^{(n,\lambda_1,\mu)^T}\mathbf{z}} e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger(\mathbb{1}_{n\times n} - \bar{A}^{(n,\lambda_1,\mu)}\bar{A}^{(n,\lambda_1,\mu)^T})\mathbf{z}} \\
&= \chi_{\rho^{(n)}}\left(\bar{A}^{(n,\lambda_2,\mu)^T}\bar{A}^{(n,\lambda_1,\mu)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger(\mathbb{1}_{n\times n} - \bar{A}^{(n,\lambda_1,\mu)}\bar{A}^{(n,\lambda_2,\mu)}\bar{A}^{(n,\lambda_2,\mu)^T}\bar{A}^{(n,\lambda_1,\mu)^T})\mathbf{z}} \qquad \forall\,\mathbf{z}\in\mathbb{C}^n\,.
\end{aligned} \tag{7.100}$$

Since quantum states and characteristic functions are in one-to-one correspondence, (7.99) and (7.100) imply that, in order to show (7.98), we only need to show that

$$\bar{A}^{(n,\lambda_1,\mu)}\bar{A}^{(n,\lambda_2,\mu)} = \bar{A}^{(n,\lambda_1\lambda_2,\mu)}\,. \tag{7.101}$$

By using (7.59), for all $i,k = 1,2,\ldots,n$ the element (i,k) of the matrix $\bar{A}^{(n,\lambda_1,\mu)}\bar{A}^{(n,\lambda_2,\mu)}$ can be expressed as

$$\begin{aligned}
(\bar{A}^{(n,\lambda_1,\mu)}\bar{A}^{(n,\lambda_2,\mu)})_{i,k} &= \sum_{j=0}^{n} \bar{A}^{(n,\lambda_1,\mu)}_{i,j}\,\bar{A}^{(n,\lambda_2,\mu)}_{j,k} \\
&= \sum_{j=0}^{n} \Theta(i-j)\,\sqrt{\lambda_1}\mu^{\frac{i-j}{2}} L^{(-1)}_{i-j}(-\ln\lambda_1)\,\Theta(j-k)\,\sqrt{\lambda_2}\mu^{\frac{j-k}{2}} L^{(-1)}_{j-k}(-\ln\lambda_2) \\
&= \Theta(k-i)\sqrt{\lambda_1\lambda_2}\mu^{\frac{i-k}{2}} \sum_{j=k}^{i} L^{(-1)}_{i-j}(-\ln\lambda_1)\,L^{(-1)}_{j-k}(-\ln\lambda_2) \\
&= \Theta(k-i)\sqrt{\lambda_1\lambda_2}\mu^{\frac{i-k}{2}} \sum_{j=0}^{i-k} L^{(-1)}_{i-k-j}(-\ln\lambda_1)\,L^{(-1)}_{j}(-\ln\lambda_2) \\
&\overset{\text{(i)}}{=} \Theta(k-i)\sqrt{\lambda_1\lambda_2}\mu^{\frac{i-k}{2}}\, L^{(-1)}_{i-k}\left(-\ln(\lambda_1\lambda_2)\right) \\
&= \bar{A}^{(n,\lambda_1\lambda_2,\mu)}_{i,k}\,,
\end{aligned} \tag{7.102}$$

where in (i) we exploited the addition formula for Laguerre polynomials. Therefore, (7.101) holds, which, as mentioned earlier, is sufficient to conclude the proof.

□

To summarise, while the DIM with a finite value of M does not satisfy Property 1 and Property 2, the DIM with $M \to \infty$ does. In particular, the DIM with $M \to \infty$ is more realistic than the LIM analysed in [10, 11, 16–18] (which corresponds to the DIM with $M = 1$).

7.5 Avram–Parter theorem and its consequences

In this section, we review the *Avram–Parter's theorem* [39, 40] and, by leveraging it, we establish a matrix-analysis result in the forthcoming Theorem 89 that will play a crucial role in the analysis of the quantum communication performances which are achievable within our model of optical fibre with memory effects. The Avram–Parter's theorem is a matrix-analysis result concerning the asymptotic behaviour of the singular values of an $n \times n$ *Toeplitz* matrix as $n \to \infty$ (see Theorem (86) below for its statement). Such a theorem can be seen as a generalisation of the so-called Szegő theorem [55, 56], which pertains to eigenvalues instead of singular values. Importantly, our Theorem 89, derived by exploiting the Avram–Parter's theorem, appears to be novel in the field of matrix analysis and we believe that it may hold independent interest. We begin with the definition of Toeplitz matrices [57].

Definition 8. an $n \times n$ real matrix A is called a Toeplitz matrix if there exist real numbers $\{a_j\}_j$ such that the elements of A satisfy

$$A_{i,k} = a_{i-k} \qquad \forall\, i,k \in \{1,2,\ldots,n\}\,. \tag{7.103}$$

In addition, a *circulant matrix* is a square matrix that has all of its row vectors composed of the same elements, with each row vector rotated one element to the right relative to the preceding row vector. Circulant matrices have the advantageous property that their spectral and singular value decompositions can be analytically calculated. In a sense, infinite-dimensional Toeplitz matrices can be viewed as infinite-dimensional circulant matrices. This intuitive idea forms the

basis for the proof of the so-called Szegő theorem [55, 56], which pertains to eigenvalues, as well as the Avram–Parter theorem [39, 40], which pertains to singular values. For an understanding of topics related to Toeplitz matrices, Szegő theorem, and Avram–Parter theorem, we recommend referring to the references [57–61]. It is worth noting that the Szegő theorem has been previously applied in the literature of quantum information theory, specifically in works such as [16–18, 62].

Theorem 86 *[Avram–Parter's Theorem]. Let $\{a_k\}_{k\in\mathbb{Z}}$ be a sequence of real numbers. For all $n\in\mathbb{N}$ let $T^{(n)}$ be the $n\times n$ Toeplitz matrix with elements $T^{(n)}_{k,j}:=a_{k-j}$ for all $k,j\in\{1,2,\dots,n\}$. Let $\{s^{(n)}_j\}_{j=1,2,\dots,n}$ be the singular values of the matrix $T^{(n)}$ ordered in increasing order in j. Assume that the function $s:[0,2\pi]\to\mathbb{C}$ defined by*

$$s(x):=\left|\sum_{k=-\infty}^{+\infty}a_k\,e^{\frac{ikx}{2}}\right| \qquad \forall\, x\in[0,2\pi]\,, \tag{7.104}$$

is bounded. Then for all continuous function $F:\mathbb{R}\to\mathbb{R}$ with bounded support it holds that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}F\left(s^{(n)}_j\right)=\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}F\left(s(x)\right)\,. \tag{7.105}$$

Lemma 87. *By using the notations of Theorem 86, for all $n\in\mathbb{N}$ it holds that all the singular values of $T^{(n)}$ are upper bounded by $\sup_{x\in[0,2\pi]}s(x)$, i.e.*

$$s^{(n)}_n\leq\sup_{x\in[0,2\pi]}s(x) \qquad \forall\, n\in\mathbb{N}\,. \tag{7.106}$$

Proof. The maximum singular value of the matrix $T^{(n)}$ satisfy

$$\begin{aligned}
s^{(n)}_n&=\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}|\mathbf{v}^TT^{(n)}\mathbf{w}|=\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}\left|\sum_{k,j=1}^{n}v_k^*T^{(n)}_{k,j}w_j\right|=\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}\left|\sum_{k,j=1}^{n}v_k^*w_ja_{k-j}\right|\\
&=\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}\left|\sum_{k,j=1}^{n}v_k^*w_j\int_0^{4\pi}\frac{\mathrm{d}x}{4\pi}e^{-i(k-j)\frac{x}{2}}\sum_{l=-\infty}^{+\infty}a_le^{il\frac{x}{2}}\right|\\
&\leq\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}\int_0^{4\pi}\frac{\mathrm{d}x}{4\pi}\left|\left(\sum_{k=1}^{n}v_ke^{ik\frac{x}{2}}\right)^*\left(\sum_{j=1}^{n}w_je^{ij\frac{x}{2}}\right)\right|\left|\sum_{l=-\infty}^{+\infty}a_le^{il\frac{x}{2}}\right|\\
&=\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}\int_0^{4\pi}\frac{\mathrm{d}x}{4\pi}\left|\left(\sum_{k=1}^{n}v_ke^{ik\frac{x}{2}}\right)^*\left(\sum_{j=1}^{n}w_je^{ij\frac{x}{2}}\right)\right|s(x)\\
&\leq\left(\sup_{x\in[0,4\pi]}s(x)\right)\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}\left|\left(\sum_{k=1}^{n}v_ke^{ikx}\right)^*\left(\sum_{j=1}^{n}w_je^{ijx}\right)\right|\\
&\leq\left(\sup_{x\in[0,4\pi]}s(x)\right)\max_{\substack{\mathbf{v},\mathbf{w}\in\mathbb{C}^n,\\ \mathbf{v}^T\mathbf{v}=\mathbf{w}^T\mathbf{w}=1}}\sqrt{\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}\left|\sum_{k=1}^{n}v_ke^{ikx}\right|^2}\sqrt{\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}\left|\sum_{j=1}^{n}w_je^{ijx}\right|^2}\\
&=\sup_{x\in[0,4\pi]}s(x)\,.
\end{aligned} \tag{7.107}$$

Since $\{a_k\}_k$ are real numbers, the function $s(x):=\left|\sum_{k=-\infty}^{+\infty}a_k\,e^{\frac{ikx}{2}}\right|$ satisfies $s(4\pi-x)=s(x)$. Consequently, it holds that

$$\sup_{x\in[0,4\pi]}s(x)=\sup_{x\in[0,2\pi]}s(x)\,. \tag{7.108}$$

Hence, this concludes the proof. □

Lemma 88. *For all $n \in \mathbb{N}$ let $\{s_j^{(n)}\}_{j\in\{1,2,\dots,n\}}$ be n real numbers that are non-decreasing in j and uniformly bounded, i.e. there exists $M > 0$ such that $|s_n^{(n)}| \leq M$ for all $n \in \mathbb{N}$. Assume that there exists a monotonically non-decreasing continuous function $s : [0, 2\pi] \to \mathbb{R}$ such that the following relation holds for any continuous function $F : \mathbb{R} \to \mathbb{R}$ with bounded support:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} F\left(s_j^{(n)}\right) = \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} F\left(s(x)\right) . \tag{7.109}$$

Then, it holds that

$$\lim_{n\to\infty} \max\left\{ \left| s_j^{(n)} - s\left(\frac{2\pi j}{n}\right) \right| : j \in \{1, 2, \dots, n\}, \quad s_j^{(n)} \in [s(0), s(2\pi)] \right\} = 0 \tag{7.110}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \left| \left\{ j : j \in \{1, 2, \dots, n\}, \quad s_j^{(n)} \notin [s(0), s(2\pi)] \right\} \right| = 0 . \tag{7.111}$$

Proof. For each $n \in \mathbb{N}$, let

$$\mu_n := \frac{1}{n} \sum_{j=1}^{n} \delta_{s_j^{(n)}} , \tag{7.112}$$

where for any $x \in \mathbb{R}$, δ_x is the Dirac delta probability distribution centered in x. From (7.109), the sequence of probability distributions μ_n converges weakly to the probability distribution μ with cumulative distribution function C [63], which is defined by the following equation:

$$C^{-1}(y) = s(2\pi y) \qquad \forall\, y \in [0, 1] , \tag{7.113}$$

where C^{-1} is the inverse function of C. For any $n \in \mathbb{N}$, let C_n be the cumulative distribution function of μ_n, which satisfies

$$C_n\left(s_j^{(n)}\right) = \frac{j}{n} \qquad \forall\, j = 1, \dots, n . \tag{7.114}$$

We have

$$\begin{aligned}
&\limsup_{n\to\infty} \max\left\{ \left| s_j^{(n)} - s\left(\frac{2\pi\, j}{n}\right) \right| : j = 1, \dots, n\,,\ s(0) \leq s_j^{(n)} \leq s(2\pi) \right\} \\
&= \limsup_{n\to\infty} \max\left\{ \left| s_j^{(n)} - s\left(2\,\pi\, C_n\left(s_j^{(n)}\right)\right) \right| : j = 1, \dots, n\,,\ s(0) \leq s_j^{(n)} \leq s(2\pi) \right\} \\
&\leq \limsup_{n\to\infty} \max\left\{ \left| \lambda - C^{-1}\left(C_n(\lambda)\right) \right| : s(0) \leq \lambda \leq s(2\pi) \right\} .
\end{aligned} \tag{7.115}$$

Since C is continuous and μ_n converges weakly to μ, C_n converges uniformly to C [63]. Therefore, $C^{-1} \circ C_n$ converges uniformly to the identity function on the interval $[s(0), s(2\pi)]$, and the claim (7.110) follows. Moreover, for any $n \in \mathbb{N}$, let

$$j_n = \max\left\{ j = 1, \dots, n : s_j^{(n)} < s(0) \right\} = \left| \left\{ j = 1, \dots, n : s_j^{(n)} < s(0) \right\} \right| . \tag{7.116}$$

We have

$$\limsup_{n\to\infty} \frac{j_n}{n} = \limsup_{n\to\infty} C_n\left(s_{j_n}^{(n)}\right) \leq \limsup_{n\to\infty} C_n\left(s(0)\right) = C(s(0)) = 0 . \tag{7.117}$$

Analogously, one can show that

$$\limsup_{n\to\infty} \frac{1}{n} \left| \left\{ j = 1, \dots, n : s_j^{(n)} > s(2\pi) \right\} \right| = 0 , \tag{7.118}$$

and the claim (7.111) follows.

□

Theorem 89. *Let $\{a_k\}_{k\in\mathbb{Z}}$ be a sequence of real numbers. For all $n \in \mathbb{N}$ let $T^{(n)}$ be the $n \times n$ Toeplitz matrix with elements $T^{(n)}_{k,j} := a_{k-j}$ for all $k, j \in \{1, 2, \dots, n\}$. Let $\{s^{(n)}_j\}_{j=1,2,\dots,n}$ be the singular values of the matrix $T^{(n)}$ ordered in increasing order in j. Assume that the function $s : [0, 2\pi] \to \mathbb{C}$ defined by*

$$s(x) := \left| \sum_{k=-\infty}^{+\infty} a_k \, e^{\frac{ikx}{2}} \right| \qquad \forall\, x \in [0, 2\pi]\,, \tag{7.119}$$

is monotonically non-decreasing and continuous. Then, the following notion of convergence of the singular values $\{s^{(n)}_j\}_{j=1,2,\dots,n}$ to the function $s(\cdot)$ holds:

$$\lim_{n\to\infty} \max\left\{ \left| s^{(n)}_j - s\left(\frac{2\pi j}{n}\right) \right| : \; j \in \{k_n, k_n + 1, \dots, n\} \right\} = 0\,, \tag{7.120}$$

where $\{k_n\}_{n\in\mathbb{N}} \subseteq \mathbb{N}$ is a suitable sequence such that $k_n \leq n$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \frac{k_n}{n} = 0$. In particular, for all $x \in (0, 1]$ it holds that

$$\lim_{n\to\infty} s^{(n)}_{\lfloor xn \rfloor} = s(2\pi x)\,, \tag{7.121}$$

where $\lfloor\cdot\rfloor$ is the floor function. Furthermore, as n approaches infinity, the fraction of singular values $\{s^{(n)}_j\}_{j=1,2,\dots,n}$ that fall outside the image of the function $s(\cdot)$ becomes negligible. Mathematically, this can be expressed as

$$\lim_{n\to\infty} \frac{1}{n} \left| \left\{ j : \; j \in \{1, 2, \dots, n\}, \quad s^{(n)}_j \notin [s(0), s(2\pi)] \right\} \right| = 0\,. \tag{7.122}$$

More specifically, the inequality $s^{(n)}_n \leq s(2\pi)$ holds for any $n \in \mathbb{N}$.

Proof. This is a direct consequence of Theorem 86, Lemma 87, and Lemma 88. □

Analogously to Theorem 89, a similar statement can be made regarding the convergence of the eigenvalues of $T^{(n)}$ as n approaches infinity, leveraging the Szegő Theorem [55, 56].

7.6 Capacities of the Delocalised Interaction Model

In this section, we derive the exact solutions for the quantum capacity Q, the two-way quantum capacity Q_2, and the secret-key capacity K of the DIM in the absence of thermal noise ($\nu = 0$). Furthermore, we investigate the parameter regions of noise where the capacities of our model are strictly positive. This characterisation allows us to determine the regions where qubit distribution, two-way entanglement distribution, and quantum key distribution can be achieved.

Let us recall that, in our model, the channel $\Phi^{(n)}_{\lambda,\mu,\nu}$ corresponds to n uses of the optical fibre. In addition, in Theorem 80 we proved that $\Phi^{(n)}_{\lambda,\mu,\nu}$ is unitarily equivalent to the following channel:

$$\left\{ \bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu} \right\}_{n\in\mathbb{N}}, \tag{7.123}$$

which is the tensor product of n distinct thermal attenuators. The transmissivities $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\dots,n}$ of these attenuators are the square of the singular values of the Toeplitz matrix reported in (7.59). Hence, we can exploit here the properties of Toeplitz matrices, that we discussed in Section 7.5, to analyse the performance of quantum communication tasks over optical fibres with memory effects. We begin by establishing the forthcoming Theorem 90, which, roughly speaking, establishes that by plotting the points

$$\left\{ \left(2\pi \frac{j}{n}\,,\, \eta_j^{(n,\lambda,\mu)} \right) \quad \text{for } j = 1, 2, \dots, n \right\}$$

on a two-dimensional plane, they converge for $n \to \infty$ to the graph of a certain function $\eta^{(\lambda,\mu)} : [0, 2\pi] \to \mathbb{R}$ reported in (7.125), which we dub *effective transmissivity function*.

Theorem 90. *Let $\lambda \in [0,1]$ and $\mu \in [0,1)$. There exists a sequence $\{j_n\}_{n\in\mathbb{N}} \subseteq \mathbb{N}$ such that $j_n \leq n$ for all $n \in \mathbb{N}$, $\lim_{n\to\infty} \frac{j_n}{n} = 0$, and*

$$\lim_{n\to\infty} \max\left\{ \left| \eta_j^{(n,\lambda,\mu)} - \eta^{(\lambda,\mu)}\left(\frac{2\pi j}{n}\right) \right| : j \in \{j_n, \dots, n\} \right\} = 0\,, \tag{7.124}$$

where $\eta^{(\lambda,\mu)}$ is the effective transmissivity function defined by

$$\eta^{(\lambda,\mu)}(x) := \lambda^{\frac{1-\mu}{1+\mu-2\sqrt{\mu}\cos(x/2)}} \quad \forall\, x \in [0,2\pi]\,. \tag{7.125}$$

In particular, for all $x \in (0,1]$ it holds that

$$\lim_{n\to\infty} \eta_{\lfloor xn \rfloor}^{(n,\lambda,\mu)} = \eta^{(\lambda,\mu)}(2\pi x)\,, \tag{7.126}$$

where $\lfloor\cdot\rfloor$ is the floor function. In addition, all the transmissivities are uniformly bounded by $\eta^{(\lambda,\mu)}(2\pi)$, i.e.

$$\eta_n^{(n,\lambda,\mu)} \leq \eta^{(\lambda,\mu)}(2\pi) = \lambda^{\frac{1-\sqrt{\mu}}{1+\sqrt{\mu}}} \qquad \forall\, n \in \mathbb{N}\,. \tag{7.127}$$

Proof. The transmissivities $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\dots,n}$ are defined by

$$\eta_i^{(n,\lambda,\mu)} := (s_i^{(n,\lambda,\mu)})^2\,, \tag{7.128}$$

where $\{s_i^{(n,\lambda,\mu)}\}_{i=1,2,\dots,n}$ are the singular values, ordered in increasing order in j, of the $n \times n$ Toeplitz matrix $\bar{A}^{(n,\lambda,\mu)}$ reported in (7.59). For all $n \in \mathbb{N}$ and all $i,k \in \{1,2,\dots,n\}$, the (i,k) element of $\bar{A}^{(n,\lambda,\mu)}$ can be expressed as

$$\bar{A}_{i,k}^{(n,\lambda,\mu)} = a_{i-k}^{(\lambda,\mu)} \tag{7.129}$$

where

$$a_j^{(\lambda,\mu)} := \Theta(j)\,\sqrt{\lambda}\mu^{\frac{j}{2}} L_j^{(-1)}(-\ln\lambda) \qquad \forall\, j \in \mathbb{Z}\,. \tag{7.130}$$

By using that the generalised Laguerre polynomials $\{L_k^{(-1)}\}_{k\in\mathbb{N}}$ satisfies

$$\sum_{k=0}^{\infty} L_k^{(-1)}(x)\, w^k = e^{\frac{wx}{w-1}} \qquad \forall\, x \in \mathbb{R}\,, \forall\, w \in \mathbb{C} \text{ such that } |w| < 1\,, \tag{7.131}$$

one can show that

$$\left| \sum_{k=-\infty}^{+\infty} a_k^{(\lambda,\mu)}\, e^{\frac{ikx}{2}} \right|^2 = \lambda^{\frac{1-\mu}{1+\mu-2\sqrt{\mu}\cos(x/2)}} \qquad \forall\, x \in [0,2\pi]\,. \tag{7.132}$$

Consequently, by applying Theorem 89, the proof of (7.127) is complete. Moreover, Theorem 89 implies that there exists a sequence $\{j_n\}_{n\in\mathbb{N}} \subseteq \mathbb{N}$ such that $j_n \leq n$ for all $n \in \mathbb{N}$, $\lim_{n\to\infty} \frac{j_n}{n} = 0$, and

$$\lim_{n\to\infty} \max\left\{ \left| \sqrt{\eta_j^{(n,\lambda,\mu)}} - \sqrt{\eta^{(\lambda,\mu)}\left(\frac{2\pi j}{n}\right)} \right| : j \in \{j_n, \dots, n\} \right\} = 0\,. \tag{7.133}$$

Since $\{\eta_i^{(n,\lambda,\mu)}\}_{i=1,2,\dots,n}$ and $\eta^{(\lambda,\mu)}$ are both bounded, with lower bound 0 and upper bound 1, the limit in (7.124) follows. □

In the forthcoming Theorem 90 we study the parameter region of zero capacities of our model of optical fibre with memory effects.

Theorem 91. *Let $\lambda \in (0,1]$, $\mu \in [0,1)$, and $\nu \geq 0$. Let $C(\lambda,\mu,\nu)$ be one of the following capacities of the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$: quantum capacity Q, two-way quantum capacity Q_2, or secret key capacity K. It holds that*

$$C(\lambda,\mu,\nu) > 0 \qquad \Longleftrightarrow \qquad C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu}) > 0\,, \tag{7.134}$$

where $\eta^{(\lambda,\mu)}$ is reported in (7.125). In particular, in the absence of thermal noise, i.e. $\nu = 0$, it holds that

$$Q(\lambda,\mu,\nu=0) > 0 \qquad \Longleftrightarrow \qquad \sqrt{\mu} > \frac{\log_2\left(\frac{1}{\lambda}\right) - 1}{\log_2\left(\frac{1}{\lambda}\right) + 1}\,. \tag{7.135}$$

In addition, for all $\nu \geq 0$ it holds that

$$\begin{aligned}
K(\lambda,\mu,\nu),\, Q_2(\lambda,\mu,\nu) > 0 \qquad &\Longleftrightarrow \qquad \sqrt{\mu} > \frac{\ln\left(\frac{1}{\lambda}\right) - \ln(1+\frac{1}{\nu})}{\ln\left(\frac{1}{\lambda}\right) + \ln(1+\frac{1}{\nu})}\,,\\
Q(\lambda,\mu,\nu) > 0 \qquad &\Longrightarrow \qquad \sqrt{\mu} > \frac{\ln\left(\frac{1}{\lambda}\right) - \ln\left(1+\frac{1}{2\nu+1}\right)}{\ln\left(\frac{1}{\lambda}\right) + \ln\left(1+\frac{1}{2\nu+1}\right)}\,,\\
Q(\lambda,\mu,\nu) > 0 \qquad &\Longleftarrow \qquad \sqrt{\mu} > \frac{\ln\left(\frac{1}{\lambda}\right) - \ln\left(1+2^{-g(\nu)}\right)}{\ln\left(\frac{1}{\lambda}\right) + \ln\left(1+2^{-g(\nu)}\right)}\,.
\end{aligned} \tag{7.136}$$

Proof. Theorem 80 establishes that the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$ is unitarily equivalent to the quantum memory $\left\{\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right\}_{n\in\mathbb{N}}$. Hence, their capacity C is identical. Let us begin with the proof of the following implication:

$$C(\lambda,\mu,\nu) > 0 \Longrightarrow C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu}) > 0\,. \tag{7.137}$$

(7.127) establishes that all the transmissivities $\{\eta_n^{(n,\lambda,\mu)}\}_{i=1,2,\ldots,n}$ are upper bounded by $\eta^{(\lambda,\mu)}(2\pi)$. Hence, since the composition rule (7.46) holds, Alice and Bob can simulate $\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}$ by using n uses of the memoryless channel $\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu}$. This implies that $C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu}) \geq C(\lambda,\mu,\nu)$, which constitutes a proof of (7.137). Now, let us show that

$$C(\lambda,\mu,\nu) > 0 \Longleftarrow C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu}) > 0\,. \tag{7.138}$$

Assume $C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu}) > 0$. Since the function $x \longmapsto C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi x),\nu})$ is continuous in $x \in [0,1]$, there exists $\bar{x} \in [0,1]$ such that $C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi x),\nu}) > 0$ for all $x \in [\bar{x},1]$. Observe that Theorem 90 implies that

$$\lim_{n\to\infty} \eta^{(n,\lambda,\mu)}_{\lfloor(\frac{1+\bar{x}}{2})n\rfloor} = \eta^{(\lambda,\mu)}\left(2\pi\left(\frac{1+\bar{x}}{2}\right)\right) > \eta^{(\lambda,\mu)}(2\pi\bar{x})\,, \tag{7.139}$$

where $\lfloor\cdot\rfloor$ is the floor function. Consequently, if n is sufficiently large it holds that

$$\eta_j^{(n,\lambda,\mu)} > \eta^{(\lambda,\mu)}(2\pi\bar{x}) \qquad \forall\, j \in \left\{\left\lfloor\left(\frac{1+\bar{x}}{2}\right)n\right\rfloor, \ldots, n\right\}\,. \tag{7.140}$$

Hence, since the composition rule (7.46) holds, Alice and Bob can simulate $n - \left\lfloor\left(\frac{1+\bar{x}}{2}\right)n\right\rfloor + 1$ uses of the memoryless channel $\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi)}$ by using the channel $\bigotimes_{i=1}^n \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}$, for n sufficiently large. Hence, since

$$\lim_{n\to\infty} \frac{n - \left\lfloor\left(\frac{1+\bar{x}}{2}\right)n\right\rfloor + 1}{n} = \frac{1-\bar{x}}{2}\,, \tag{7.141}$$

it holds that

$$C(\lambda,\mu,\nu) \geq \left(\frac{1-\bar{x}}{2}\right) C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi)}) > 0\,, \tag{7.142}$$

i.e. the implication in (7.138) is proved. Hence, we have proved that $C(\lambda,\mu,\nu)>0$ if and only if $C(\mathscr{E}_{\eta^{(\lambda,\mu)}(2\pi),\nu})>0$. As a consequence of this fact, by using that

$$\eta^{(\lambda,\mu)}(2\pi)=\lambda^{\frac{1-\sqrt{\mu}}{1+\sqrt{\mu}}}, \tag{7.143}$$

and by exploiting the bounds on the parameter region of zero capacities reported in (7.50), (7.51), and (7.52), one can show the validity of (7.135) and (7.136). □

Now, let us discuss why Theorem 91 is interesting. In the absence of memory effects ($\mu=0$), it is known that no quantum communication tasks can be achieved when the transmissivity is sufficiently low ($Q=0$ for $\lambda\leq\frac{1}{2}$, and $Q_2=K=0$ for $\lambda\leq\frac{\nu}{\nu+1}$). However, in Theorem 91 we established that for any transmissivity $\lambda>0$ and thermal noise $\nu\geq 0$, there exists a critical value of the memory parameter μ above which, it becomes possible to achieve qubit distribution ($Q>0$), entanglement distribution ($Q_2>0$), and secret-key distribution ($K>0$). This result is particularly intriguing as it demonstrates that memory effects provide an advantage, enabling quantum communication tasks in highly noisy regimes that were previously deemed impossible. Specifically, this phenomenon bears resemblance to the "die-hard quantum communication" phenomenon observed in [10, 11] within their toy model of optical fibre with memory effects, which was limited to the analysis of Q only and does not consider K and Q_2. In our paper, we not only demonstrate the persistence of such a phenomenon in a more realistic model, accounting for Q, Q_2, and K, but we also derive an analytical expression for the critical value of the memory parameter μ, which goes beyond the findings of [10, 11]. Note that, by relating μ to the temporal interval δt between subsequent input signals (e.g. $\mu=e^{-\delta t/t_E}$ with t_E being the thermalisation timescale), Theorem 91 can also be expressed in terms of the critical δt below which the above mentioned quantum communication tasks can be achieved.

In the forthcoming Theorem 92 we calculate the quantum capacity Q, the two-way quantum capacity Q_2, and the secret-key capacity K of our model in the absence of thermal noise ($\nu=0$). It is worth noting that in the presence of thermal noise ($\nu>0$), the capacities remain unknown even in the absence of memory effects ($\mu=0$), as the capacities of the thermal attenuator are currently unknown.

Theorem 92. *Let $\lambda\in[0,1)$, $\mu\in[0,1)$, $\nu\geq 0$. Let C be one of the following capacities: quantum capacity Q, two-way quantum capacity Q_2, or secret key capacity K. In addition, let $C(\lambda,\mu,\nu)$ be the capacity C of the quantum memory channel $\{\Phi^{(n)}_{\lambda,\mu,\nu}\}_{n\in\mathbb{N}}$. In absence of thermal noise, i.e. $\nu=0$, it holds that*

$$C(\lambda,\mu,\nu=0)=\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}C\left(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),0}\right). \tag{7.144}$$

where $\eta^{(\lambda,\mu)}(x)$ is the effective transmissivity function expressed in (7.125). In particular,

$$\begin{aligned} Q(\lambda,\mu,\nu=0)&=\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}\max\left\{0,\log_2\left(\frac{\eta^{(\lambda,\mu)}(x)}{1-\eta^{(\lambda,\mu)}(x)}\right)\right\},\\ Q_2(\lambda,\mu,\nu=0)&=K(\lambda,\mu,\nu=0)\\ &=\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}\log_2\left(\frac{1}{1-\eta^{(\lambda,\mu)}(x)}\right). \end{aligned} \tag{7.145}$$

Moreover, in the presence of thermal noise, i.e. $\nu>0$, it holds that

$$C(\lambda,\mu,\nu)\geq\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}C\left(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),\nu}\right). \tag{7.146}$$

Proof. Let us fix $P\in\mathbb{N}$ with $P\geq 2$. In addition, for all $n\in\mathbb{N}$ and all $i\in\mathbb{N}$ with $i\geq n$, let us define

$$\eta_i^{(n,\lambda,\mu)}:=\eta_n^{(n,\lambda,\mu)}. \tag{7.147}$$

By theorem 90, there exists a sequence $\{j_k\}_{k\in\mathbb{N}} \subseteq \mathbb{N}$ such that $j_k \leq k$ for all $k \in \mathbb{N}$, $\lim_{k\to\infty} \frac{j_k}{k} = 0$, and

$$\max\left\{\left|\eta_j^{(n,\lambda,\mu)} - \eta^{(\lambda,\mu)}\left(\frac{2\pi j}{n}\right)\right| : j \in \{j_n, \ldots, n\}\right\} < \frac{1}{P} \tag{7.148}$$

for n sufficiently large. In particular,

$$\begin{aligned} \eta_{\lceil \frac{n}{P} \rceil p}^{(n,\lambda,\mu)} &< \eta^{(\lambda,\mu)}\left(\frac{2\pi p}{P}\right) + \frac{1}{P} \qquad \forall\, p \in \{1, 2, \ldots, P\}\,, \\ \eta_{\lfloor \frac{n}{P} \rfloor p+1}^{(n,\lambda,\mu)} &> \eta^{(\lambda,\mu)}\left(\frac{2\pi p}{P}\right) - \frac{1}{P} \qquad \forall\, p \in \{1, 2, \ldots, P-1\}\,, \end{aligned} \tag{7.149}$$

for n sufficiently large, where $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ denote the ceil and floor function, respectively. It holds that

$$C(\lambda,\mu,\nu) \overset{\text{(i)}}{=} C\left(\left\{\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right\}_{n\in\mathbb{N}}\right) \tag{7.150}$$

$$\overset{\text{(ii)}}{\leq} C\left(\left\{\bigotimes_{i=1}^{\lceil \frac{n}{P} \rceil P} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right\}_{n\in\mathbb{N}}\right) \tag{7.151}$$

$$\overset{\text{(iii)}}{\leq} C\left(\left\{\left(\bigotimes_{p=1}^{P} \mathscr{E}_{\eta_{\lceil \frac{n}{P} \rceil p}^{(n,\lambda,\mu)},\nu}\right)^{\otimes \lceil \frac{n}{P} \rceil}\right\}_{n\in\mathbb{N}}\right) \tag{7.152}$$

$$\overset{\text{(iv)}}{\leq} C\left(\left\{\left(\bigotimes_{p=1}^{P} \mathscr{E}_{\min\{1,\eta^{(\lambda,\mu)}(2\pi\frac{p}{P})+\frac{1}{P}\},\nu}\right)^{\otimes \lceil \frac{n}{P} \rceil}\right\}_{n\in\mathbb{N}}\right) \tag{7.153}$$

$$\overset{\text{(v)}}{=} \frac{1}{P} C\left(\bigotimes_{p=1}^{P} \mathscr{E}_{\min\{1,\eta^{(\lambda,\mu)}(2\pi\frac{p}{P})+\frac{1}{P}\},\nu}\right), \tag{7.154}$$

where: in (i) we applied Theorem 80; (ii) is a consequence of the fact that

$$\bigotimes_{i=1}^{n} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu} \tag{7.155}$$

can be simulated via

$$\bigotimes_{i=1}^{\lceil \frac{n}{P} \rceil P} \mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\,, \tag{7.156}$$

by using only the first n channel uses and discarding the remaining $\left\lceil \frac{n}{P} \right\rceil P - n$ ones; in (iii) we exploited the composition rule in (7.46) and the fact that the transmissivity $\eta_i^{(Pl,\lambda,\mu)}$ is monotonically non-decreasing in i; in (iv) we made the assumption that n is sufficiently large, which can be done without loss of generality because the capacities are defined in the limit as the number of channel uses n approaches infinity, and in addition we employed the inequality stated in (7.149); in (v) we exploited (7.48) and the notation introduced in Section 7.48 to denote the capacities of memory

and memoryless quantum channels. Analogously, it holds that

$$C(\lambda,\mu,\nu) = C\left(\left\{\bigotimes_{i=1}^{n}\mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right\}_{n\in\mathbb{N}}\right) \tag{7.157}$$

$$\overset{\text{(vi)}}{\geq} C\left(\left\{\bigotimes_{i=1}^{\left\lfloor\frac{n}{P}\right\rfloor P}\mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu}\right\}_{n\in\mathbb{N}}\right) \tag{7.158}$$

$$\geq C\left(\left\{\left(\bigotimes_{p=1}^{P-1}\mathscr{E}_{\eta^{(n,\lambda,\mu)}_{\left\lfloor\frac{n}{P}\right\rfloor p+1},\nu}\right)^{\otimes\left\lfloor\frac{n}{P}\right\rfloor}\right\}_{n\in\mathbb{N}}\right) \tag{7.159}$$

$$\geq C\left(\left\{\left(\bigotimes_{p=1}^{P}\mathscr{E}_{\max\{0,\eta^{(\lambda,\mu)}(2\pi\frac{p}{P})-\frac{1}{P}\},\nu}\right)^{\otimes\left\lfloor\frac{n}{P}\right\rfloor}\right\}_{n\in\mathbb{N}}\right) \tag{7.160}$$

$$= \frac{1}{P}C\left(\bigotimes_{p=1}^{P}\mathscr{E}_{\max\{0,\eta^{(\lambda,\mu)}(2\pi\frac{p}{P})-\frac{1}{P}\},\nu}\right), \tag{7.161}$$

where (vi) is a consequence of the fact that

$$\bigotimes_{i=1}^{\left\lfloor\frac{n}{P}\right\rfloor P}\mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu} \tag{7.162}$$

can be simulated via

$$\bigotimes_{i=1}^{n}\mathscr{E}_{\eta_i^{(n,\lambda,\mu)},\nu} \tag{7.163}$$

by using only the first $\left\lfloor\frac{n}{P}\right\rfloor P$ channel uses and discarding the remaining $n-\left\lfloor\frac{n}{P}\right\rfloor P$ ones. Consequently, the capacity $C(\lambda,\mu,\nu)$ can be bounded as

$$\frac{1}{P}C\left(\bigotimes_{p=1}^{P-1}\mathscr{E}_{\max(0,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)-\frac{1}{P}),\nu}\right) \leq C(\lambda,\mu,\nu) \leq \frac{1}{P}C\left(\bigotimes_{p=1}^{P}\mathscr{E}_{\min(1,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)+\frac{1}{P}),\nu}\right) \tag{7.164}$$

for all $P\in\mathbb{N}$ with $P\geq 2$. In particular

$$\limsup_{P\to\infty}\frac{1}{P}C\left(\bigotimes_{p=1}^{P-1}\mathscr{E}_{\max(0,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)-\frac{1}{P}),\nu}\right) \leq C(\lambda,\mu,\nu) \leq \liminf_{P\to\infty}\frac{1}{P}C\left(\bigotimes_{p=1}^{P}\mathscr{E}_{\min(1,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)+\frac{1}{P}),\nu}\right) \tag{7.165}$$

Consequently, it holds that

$$C(\lambda,\mu,\nu) \geq \limsup_{P\to\infty}\frac{1}{P}C\left(\bigotimes_{p=1}^{P-1}\mathscr{E}_{\max(0,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)-\frac{1}{P}),\nu}\right) \tag{7.166}$$

$$\overset{\text{(i)}}{\geq} \limsup_{P\to\infty}\frac{1}{P}\sum_{p=1}^{P-1}C\left(\mathscr{E}_{\max(0,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)-\frac{1}{P}),\nu}\right) \tag{7.167}$$

$$\overset{\text{(ii)}}{=} \int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}C\left(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),\nu}\right), \tag{7.168}$$

where: (i) is a consequence of the fact that Alice and Bob can independently employ the optimal communication strategy for each of the P single-mode channels that define the P-mode attenuator $\bigotimes_{p=1}^{P-1}\mathscr{E}_{\max(0,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)-\frac{1}{P}),\nu}$; in (ii) we exploited the fact that the function $\lambda\longmapsto C(\mathscr{E}_{\lambda,0})$ is continuous in $\lambda\in[0,1)$ and that

$$\lambda^{\frac{1}{1-\mu}} \leq \eta^{(\lambda,\mu)}(x) \leq \lambda^{\frac{1-\sqrt{\mu}}{1+\sqrt{\mu}}} < 1\,, \tag{7.169}$$

as established by (7.127). Hence, we have proved (7.146).

Now, let us assume that the thermal noise is zero, i.e. $\nu = 0$. Under this assumption, we can exploit the additivity of the capacity C of the pure-loss channels, i.e. C is such that for all transmissivities $\{\lambda_p\}_{p=1,2,\ldots,P}$ and all $P \in \mathbb{N}$ it holds that

$$C\left(\bigotimes_{p=1}^{P} \mathscr{E}_{\lambda_p,\nu}\right) = \sum_{p=1}^{P} C\left(\mathscr{E}_{\lambda_p,\nu}\right) . \tag{7.170}$$

Indeed, the capacities Q [41], Q_2 [28], and K [28] of the pure-loss channel are additive. Hence, (7.165) implies that:

$$\limsup_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P-1} C\left(\mathscr{E}_{\max(0,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)-\frac{1}{P}),0}\right) \leq C(\lambda,\mu,0) \leq \liminf_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P} C\left(\mathscr{E}_{\min(1,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)+\frac{1}{P}),0}\right) . \tag{7.171}$$

By exploiting that $\eta^{(\lambda,\mu)}(2\pi) < 1$ and the fact that the function $\lambda \longmapsto C(\mathscr{E}_{\lambda,0})$ is continuous in $\lambda \in [0,1)$, we have that the left-hand side and the right-hand side of (7.171) converge to the same quantity:

$$\begin{aligned} \liminf_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P} C\left(\mathscr{E}_{\min(1,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)+\frac{1}{P}),0}\right) &= \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} C(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),0}) , \\ \limsup_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P-1} C\left(\mathscr{E}_{\max(0,\eta^{(\lambda,\mu)}\left(2\pi\frac{p}{P}\right)-\frac{1}{P}),0}\right) &= \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} C(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),0}) . \end{aligned} \tag{7.172}$$

Hence, in the absence of thermal noise, the capacity $C(\lambda,\mu,0)$ of our quantum memory channel is given by

$$C(\lambda,\mu,0) = \int_0^{2\pi} \frac{\mathrm{d}x}{2\pi} C(\mathscr{E}_{\eta^{(\lambda,\mu)}(x),0}) . \tag{7.173}$$

Consequently, by leveraging the expression of the capacities Q [41], Q_2 [28], and K [28] of the pure-loss channel reported in (7.49), we obtain (7.145).

□

7.7 Capacities of the Localised Interaction Model

In this section, we provide an expression for the capacities of the "Localised Interaction Model" (LIM) of references [16–18], which has been reviewed both in the main text and in Section 7.3.

Let $Q^{\mathrm{LIM}}(\lambda,\mu,\nu)$, $Q_2^{\mathrm{LIM}}(\lambda,\mu,\nu)$, and $K^{\mathrm{LIM}}(\lambda,\mu,\nu)$ be the quantum capacity, two-way quantum capacity, and secret-key capacity of the LIM with transmissivity λ, memory parameter μ, and thermal noise ν. The effective transmissivity function of the LIM is given by [16, Eq. (5)]:

$$\eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x) = \frac{\mu+\lambda-2\sqrt{\mu\lambda}\cos\left(\frac{x}{2}\right)}{1+\mu\lambda-2\sqrt{\mu\lambda}\cos\left(\frac{x}{2}\right)} \qquad \forall\, x \in [0,2\pi] . \tag{7.174}$$

Note that $\eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x)$ is monotonically increasing in x and thus achieves is maximum in $x = 2\pi$:

$$\max_{x\in[0,2\pi]} \eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x) = \eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x=2\pi) = \left(\frac{\sqrt{\mu}+\sqrt{\lambda}}{1+\sqrt{\mu\lambda}}\right)^2 . \tag{7.175}$$

We can apply the same reasoning used in Theorem 91 in order to determine the parameter region where the capacities are strictly positive. Specifically, in the absence of thermal noise ($\nu = 0$),

the quantum capacity $Q^{\mathrm{LIM}}(\lambda,\mu,\nu=0)$ is strictly positive if and only if $\eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x=2\pi)>\frac{1}{2}$. This condition can be expressed as follows:

$$Q^{\mathrm{LIM}}(\lambda,\mu,\nu=0)>0 \qquad \iff \qquad \sqrt{\mu}>\frac{1-\sqrt{2\lambda}}{\sqrt{2}-\sqrt{\lambda}}\,. \tag{7.176}$$

Additionally, the two-way quantum capacity $Q_2^{\mathrm{LIM}}(\lambda,\mu,\nu)$ and secret-key capacity $K^{\mathrm{LIM}}(\lambda,\mu,\nu)$ are strictly positive if and only if $\eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x=2\pi)>\frac{\nu}{\nu+1}$. This condition can be expressed as follows:

$$Q^{\mathrm{LIM}}(\lambda,\mu,\nu),K^{\mathrm{LIM}}(\lambda,\mu,\nu)>0 \qquad \iff \qquad \sqrt{\mu}>\frac{\sqrt{\nu}-\sqrt{\lambda(\nu+1)}}{\sqrt{\nu+1}-\sqrt{\lambda\nu}}\,. \tag{7.177}$$

The quantum capacity $Q^{\mathrm{LIM}}(\lambda,\mu,0)$ of the LIM in the absence of thermal noise ($\nu=0$) is given by [16]:

$$Q^{\mathrm{LIM}}(\lambda,\mu,0)=\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}\max\left(0,\log_2\left(\frac{\eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x)}{1-\eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x)}\right)\right). \tag{7.178}$$

One can numerically verify that $Q^{\mathrm{LIM}}(\lambda,\mu,\nu=0)$ is an increasing function of μ for each assigned transmissivity value λ, implying that memory effects can improve quantum communication.

Furthermore, by exploiting the same reasoning used in Theorem 92, we can directly derive an expression for the two-way quantum and the secret-key capacity of the LIM in the absence of thermal noise. Namely, they are given by:

$$K^{\mathrm{LIM}}(\lambda,\mu,0)=Q_2^{\mathrm{LIM}}(\lambda,\mu,0)=\int_0^{2\pi}\frac{\mathrm{d}x}{2\pi}\log_2\left(\frac{1}{1-\eta_{\mathrm{LIM}}^{(\lambda,\mu)}(x)}\right). \tag{7.179}$$

7.8 Transversal attenuation

A possible limitation of the DIM introduced before is the fact that the environments are taken to be single-mode. This would not be an issue in absence of memory effects, since one could always select a single-mode component of the environment as the effective state interacting with the input. Instead, memory effects allow for more complex dynamics. One could easily extend our model to capture a memoryless contribution to the noise, by considering the case where between each interaction with $E_1^{(j)}$ and $E_1^{(j+1)}$ there is an additional thermal loss of attenuation $\gamma^{1/M}$ and environment state τ_ν, but without memory effects, as depicted in Fig. 7.7. In this way, we are allowing for the presence of an additional mechanism of interaction with the environment which happens on a faster timescale. By taking $\gamma^{1/M}$ to be uniform along the several uses of the transmission line, the action on the annihilation operators of the input modes turns out to be a simple rescaling by $\sqrt{\gamma^{1/M}}$, which commutes with the original interaction. The net effect is to place identical thermal attenuators at each mode, before or after the memory channel.

This simple model captures the possibility of having the communication line interacting with a multi-mode system, where effectively there are two independent noise effects acting on different timescales. For example, one may think of the memory effects to be due to scattering and the memoryless noise to be due to absorption. Since at the end we will see that the memory channel can be written as a tensor product of attenuators, by the composition rules for attenuators, the capacities of this enhanced model are also completely computable. In particular, it follows that the advantage on the quantum capacity might remain, if γ is sufficiently large. If $\nu=0$, the advantage due to memory effects on one-way qubit transmission remains if and only if $\gamma>\frac{1}{2}$, while the two-way quantum capacity can be non-zero as soon as $\gamma\geq\frac{\nu}{\nu+1}$.

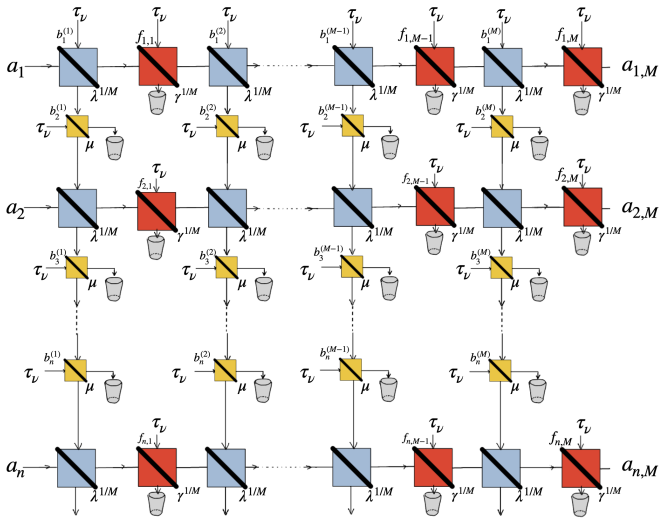


Figure 7.7: Depiction of the quantum memory channel which models both memory and memoryless noise. In contrast to the model presented in Figure 7.6, here we have also the red beam splitters of transmissivity $\gamma^{1/M}$ which models the presence of memoryless contribution to the noise. This addition, while it does not substantially affect the mathematical structure of the model, captures the possibility of having different kinds of noise effects on the fibre at the same time, such as scattering, which could enable memory effects, and absorption, which may be memoryless.

In concrete, we modify the model by introducing new modes $\mathbf{f}$ as depicted in Fig. 7.7, i.e.

$$\begin{aligned}
\mathbf{a} &:= (a_1,\, a_2, \ldots,\, a_n)^T\,, \\
\mathbf{a_M} &:= (a_{1,M},\, a_{2,M}, \ldots,\, a_{n,M})^T\,, \\
\mathbf{b} &:= (b_1^{(1)},\, b_1^{(2)}, \ldots,\, b_1^{(M)},\, b_2^{(1)},\, b_2^{(2)}, \ldots,\, b_2^{(M)}, \ldots,\, b_n^{(1)},\, b_n^{(2)}, \ldots,\, b_n^{(M)})^T\,, \\
\mathbf{f} &:= (f_{1,1}, f_{1,2}, \ldots,\, f_{1,M-1}, f_{1,M}, \ldots,\, f_{n-1,1},\, f_{n,2}, \ldots,\, f_{n,M-1}, f_{n,M})^T\,,
\end{aligned} \tag{7.180}$$

one can modify (7.68) as

$$\mathbf{a_M} = A^{(M,n,\lambda,\mu,\gamma)}\mathbf{a} + E^{(M,n,\lambda,\mu,\gamma)}\mathbf{b} + F^{(M,n,\lambda,\mu,\gamma)}\mathbf{f}\,, \tag{7.181}$$

where $A^{(M,n,\lambda,\mu,\gamma)}$ is an $n \times n$ real matrix, $E^{(M,n,\lambda,\mu,\gamma)}$ is an $n \times nM$ real matrix, and $F^{(M,n,\lambda,\mu,\gamma)}$ is an $n \times nM$ real matrix which satisfy

$$A^{(M,n,\lambda,\mu,\gamma)}A^{(M,n,\lambda,\mu,\gamma)^T} + E^{(M,n,\lambda,\mu,\gamma)}E^{(M,n,\lambda,\mu,\gamma)^T} + F^{(M,n,\lambda,\mu,\gamma)}F^{(M,n,\lambda,\mu,\gamma)^T} = \mathbb{1}_{n\times n}\,, \tag{7.182}$$

thanks to (7.66). By following the same steps of the proof of Theorem 80, we get again

$$\begin{aligned}
\chi_{\Phi^{(M,n)}_{\lambda,\mu,\nu,\gamma}(\rho^{(n)})}(\mathbf{z}) &:= \mathrm{Tr}\left[\Phi^{(M,n)}_{\lambda,\mu,\nu,\gamma}(\rho^{(n)})\, e^{\mathbf{a}^\dagger\mathbf{z}-\mathbf{z}^\dagger\mathbf{a}}\right] \\
&= \chi_{\rho^{(n)}}\left(A^{(M,n,\lambda,\mu,\gamma)^T}\mathbf{z}\right)\chi_{\tau_\nu^{\otimes nM}}\left(E^{(M,n,\lambda,\mu,\gamma)^T}\mathbf{z}\right)\chi_{\tau_\nu^{\otimes nM}}\left(F^{(M,n,\lambda,\mu,\gamma)^T}\mathbf{z}\right) \\
&= \chi_{\rho^{(n)}}\left(A^{(M,n,\lambda,\mu,\gamma)^T}\mathbf{z}\right) e^{-(\nu+\frac{1}{2})\mathbf{z}^\dagger(\mathbb{1}_{n\times n}-A^{(M,n,\lambda,\mu,\gamma)}A^{(M,n,\lambda,\mu,\gamma)^T})\mathbf{z}} \qquad \forall\,\mathbf{z}\in\mathbb{C}^n\,,
\end{aligned} \tag{7.183}$$

And the analysis continues as in the proof of Theorem 80. We are left to specify $A^{(M,n,\lambda,\mu,\gamma)}$ in the modified model. By exploiting (7.40) and the notation in Fig. 7.7, we derive the following passive transformation:

$$\begin{aligned}
a_{i,j} &= \sqrt{\gamma^{1/M}}\sqrt{\lambda^{1/M}}\, a_{i,j-1} + \sqrt{\gamma^{1/M}}\sqrt{1-\lambda^{1/M}}\, m_{i-1,j} + \sqrt{1-\gamma^{1/M}}\, f_{i,j-1}\,, \\
m_{i,j} &= -\sqrt{1-\mu}\, b^{(j)}_{i+1} + \sqrt{\mu\lambda^{1/M}}\, m_{i-1,j} - \sqrt{\mu(1-\lambda^{1/M})}\, a_{i,j-1}\,,
\end{aligned} \tag{7.184}$$

where the transmissivity γ is associated to memoryless noise. By defining $\tilde{a}_{i,j} \coloneqq \gamma^{-j/M} a_{i,j}$, and $\tilde{m}_{i,j} \coloneqq \gamma^{-(j-1)/M} m_{i,j}$ we obtain

$$\begin{aligned} \tilde{a}_{i,j} &\simeq \sqrt{\lambda^{1/M}}\, \tilde{a}_{i,j-1} + \sqrt{1-\lambda^{1/M}}\, \tilde{m}_{i-1,j}\,, \\ \tilde{m}_{i,j} &\simeq \sqrt{\mu\lambda^{1/M}}\, \tilde{m}_{i-1,j} - \sqrt{\mu(1-\lambda^{1/M})}\, \tilde{a}_{i,j-1}\,, \end{aligned} \tag{7.185}$$

which are the recurrence relations for the model at $\gamma = 1$. We immediately get

$$A^{(M,n,\lambda,\mu,\gamma)} = \sqrt{\gamma} A^{(M,n,\lambda,\mu,1)} = \sqrt{\gamma} A^{(M,n,\lambda,\mu)} \tag{7.186}$$

and therefore

$$\bar{A}^{(n,\lambda,\mu,\gamma)}_{i,h} \coloneqq \lim_{M\to\infty} A^{(M,n,\lambda,\mu,\gamma)}_{i,h} = \sqrt{\gamma} \bar{A}^{(n,\lambda,\mu)}_{i,h}\,. \tag{7.187}$$

This means that this modification of the model is equivalent to placing a thermal attenuator with transmissivity γ before or after the memory channel. Mathematically, the square of the absolute values of the singular values of $\bar{A}^{(n,\lambda,\mu,\gamma)}_{i,h}$ are obtained by rescaling those of $\bar{A}^{(n,\lambda,\mu)}_{i,h}$ by γ. It follows that as long as $\gamma > 1/2$ and for $\nu = 0$, there are values of μ and λ such that the quantum capacity is still non-zero. In particular $\gamma\lambda$, which can be thought as the effective attenuation on the fibre, could be less than $1/2$ while the quantum capacities are still non zero. A similar observation holds for the two-way capacities. If there were no memory effects, the two-way quantum capacity would be zero if and only if $\gamma\lambda \leq \frac{\nu}{\nu+1}$, while in our model it can be still non-zero even if the inequality is satisfied.

Bibliography

[1] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010.

[2] S. Wehner, D. Elkouss, and R. Hanson. Quantum internet: A vision for the road ahead. *Science*, 362:eaam9288, 10 2018. doi: 10.1126/science.aam9288.

[3] S. Pirandola et al. Advances in quantum cryptography. *Advances in Optics and Photonics*, 12(4):1012–1236, 2020. doi: 10.1364/AOP.361502. URL http://opg.optica.org/aop/abstract.cfm?URI=aop-12-4-1012.

[4] C. H. Bennett. Quantum cryptography: public key distribution and coin tossing. In *Proc. IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, pages 175–179, 1984.

[5] J. S. Sidhu et al. Advances in space quantum communications. *IET Quantum Communication*, 2(4):182–217, 2021. doi: https://doi.org/10.1049/qtc2.12015.

[6] J. I. Cirac, A. K. Ekert, S. F. Huelga, and C. Macchiavello. Distributed quantum computation over noisy channels. *Physical Review A*, 59:4249–4254, 1999. doi: 10.1103/PhysRevA.59.4249. URL https://link.aps.org/doi/10.1103/PhysRevA.59.4249.

[7] A. Broadbent, J. Fitzsimons, and E. Kashefi. Universal blind quantum computation. *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, 2009. doi: 10.1109/focs.2009.36. URL http://dx.doi.org/10.1109/FOCS.2009.36.

[8] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller. Quantum repeaters: The role of imperfect local operations in quantum communication. *Physical Review Letters*, 81:5932–5935, 1998. doi: 10.1103/PhysRevLett.81.5932.

[9] W. J. Munro, K. Azuma, K. Tamaki, and K. Nemoto. Inside quantum repeaters. *IEEE Journal of Selected Topics in Quantum Electronics*, 21(3):78–90, 2015. doi: 10.1109/JSTQE.2015.2392076.

[10] F. A. Mele, L. Lami, and V. Giovannetti. Restoring quantum communication efficiency over high loss optical fibers. *Physical Review Letters*, 129:180501, 2022. doi: 10.1103/PhysRevLett.129.180501. URL https://link.aps.org/doi/10.1103/PhysRevLett.129.180501.

[11] F. A. Mele, L. Lami, and V. Giovannetti. Quantum optical communication in the presence of strong attenuation noise. *Physical Review A*, 106:042437, 2022. doi: 10.1103/PhysRevA.106.042437. URL https://link.aps.org/doi/10.1103/PhysRevA.106.042437.

[12] F. Caruso, V. Giovannetti, C. Lupo, and S. Mancini. Quantum channels and memory effects. *Reviews of Modern Physics*, 86:1203–1259, 2014. doi: 10.1103/RevModPhys.86.1203.

[13] V. Giovannetti. A dynamical model for quantum memory channels. *J. Phys. A: Math. Gen.*, 38(50):10989–11005, 2005.

[14] K. Banaszek, A. Dragan, W. Wasilewski, and C. Radzewicz. Experimental demonstration of entanglement-enhanced classical communication over a quantum channel with correlated noise. *Physical Review Letters*, 92:257901, 2004. doi: 10.1103/PhysRevLett.92.257901. URL https://link.aps.org/doi/10.1103/PhysRevLett.92.257901.

[15] J. Ball, A. Dragan, and K. Banaszek. Exploiting entanglement in communication channels with correlated noise. *Physical Review A*, 69:042324, 2004.

[16] C. Lupo, V. Giovannetti, and S. Mancini. Capacities of lossy bosonic memory channels. *Phys. Rev. Lett.*, 104:030501, Jan 2010. doi: 10.1103/PhysRevLett.104.030501. URL https://link.aps.org/doi/10.1103/PhysRevLett.104.030501.

[17] C. Lupo, V. Giovannetti, and S. Mancini. Memory effects in attenuation and amplification quantum processes. *Phys. Rev. A*, 82:032312, Sep 2010. doi: 10.1103/PhysRevA.82.032312. URL https://link.aps.org/doi/10.1103/PhysRevA.82.032312.

[18] G. De Palma, A. Mari, and V. Giovannetti. Classical capacity of gaussian thermal memory channels. *Phys. Rev. A*, 90:042312, Oct 2014. doi: 10.1103/PhysRevA.90.042312. URL https://link.aps.org/doi/10.1103/PhysRevA.90.042312.

[19] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2nd edition, 2017. ISBN 9781316813300.

[20] S. Khatri and M. M. Wilde. Principles of quantum communication theory: A modern approach, 2020. URL https://arxiv.org/abs/2011.04672.

[21] L. Lami, M. B. Plenio, V. Giovannetti, and A. S. Holevo. Bosonic quantum communication across arbitrarily high loss channels. *Physical Review Letters*, 125:110504, 2020. doi: 10.1103/PhysRevLett.125.110504.

[22] F. A. Mele, L. Lami, and V. Giovannetti. Maximum tolerable excess noise in CV-QKD and improved lower bound on two-way capacities. *Preprint arXiv:2303.12867*, 2023. doi: 10.48550/arXiv.2303.12867.

[23] C. M. Caves and P. D. Drummond. Quantum limits on bosonic communication rates. *Reviews of Modern Physics*, 66:481–537, 1994. doi: 10.1103/RevModPhys.66.481. URL https://link.aps.org/doi/10.1103/RevModPhys.66.481.

[24] A. Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, Taylor & Francis Group, Boca Raton, USA, 2017.

[25] A. S. Holevo and R. F. Werner. Evaluating capacities of bosonic Gaussian channels. *Physical Review A*, 63:032312, 2001. doi: 10.1103/PhysRevA.63.032312.

[26] V. Giovannetti, S. Lloyd, L. Maccone, and P. W. Shor. Entanglement assisted capacity of the broadband lossy channel. *Physical Review Letters*, 91:047901, 2003. doi: 10.1103/PhysRevLett.91.047901. URL https://link.aps.org/doi/10.1103/PhysRevLett.91.047901.

[27] V. Giovannetti, S. Lloyd, L. Maccone, and P. W. Shor. Broadband channel capacities. *Physical Review A*, 68:062323, 2003. doi: 10.1103/PhysRevA.68.062323. URL https://link.aps.org/doi/10.1103/PhysRevA.68.062323.

[28] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi. Fundamental limits of repeaterless quantum communications. *Nature Communications*, 8(1):15043, 2017. doi: 10.1038/ncomms15043.

[29] M. Rosati, A. Mari, and V. Giovannetti. Narrow bounds for the quantum capacity of thermal attenuators. *Nature Communications*, 9(1):4339, 2018. doi: 10.1038/s41467-018-06848-0.

[30] K. Sharma, M. M. Wilde, S. Adhikari, and M. Takeoka. Bounding the energy-constrained quantum and private capacities of phase-insensitive bosonic Gaussian channels. *New Journal of Physics*, 20(6):063025, 2018. doi: 10.1088/1367-2630/aac11a.

[31] K. Noh, V. V. Albert, and L. Jiang. Quantum capacity bounds of Gaussian thermal loss channels and achievable rates with Gottesman-Kitaev-Preskill codes. *IEEE Transactions on Information Theory*, 65(4):2563–2582, 2019. ISSN 1557-9654. doi: 10.1109/TIT.2018.2873764.

[32] K. Noh, S. Pirandola, and L. Jiang. Enhanced energy-constrained quantum communication over bosonic Gaussian channels. *Nature Communications*, 11(1):457, 2020. doi: 10.1038/s41467-020-14329-6.

[33] M. Fanizza, F. Kianvash, and V. Giovannetti. Estimating quantum and private capacities of gaussian channels via degradable extensions. *Physical Review Letters*, 127:210501, 2021. doi: 10.1103/PhysRevLett.127.210501.

[34] S. Pirandola, R. García-Patrón, S. L. Braunstein, and S. Lloyd. Direct and reverse secret-key capacities of a quantum channel. *Physical Review Letters*, 102:050503, 2009. doi: 10.1103/PhysRevLett.102.050503.

[35] C. Ottaviani et al. Secret key capacity of the thermal-loss channel: improving the lower bound. In Mark T. Gruneisen, Miloslav Dusek, and John G. Rarity, editors, *Quantum Information Science and Technology II*, volume 9996, page 999609. International Society for Optics and Photonics, SPIE, 2016. doi: 10.1117/12.2244899.

[36] Y. Tamura, H. Sakuma, K. Morita, M. Suzuki, Y. Yamamoto, K. Shimada, Y. Honma, K. Sohma, T. Fujii, and T. Hasegawa. The first 0.14-db/km loss optical fiber and its impact on submarine transmission. *J. Lightwave Technol.*, 36(1):44–49, 2018. URL http://www.osapublishing.org/jlt/abstract.cfm?URI=jlt-36-1-44.

[37] M.-J. Li and T. Hayashi. Chapter 1 - advances in low-loss, large-area, and multicore fibers. In A. E. Willner, editor, *Optical Fiber Telecommunications VII*, pages 3–50. Academic Press, 2020. ISBN 978-0-12-816502-7. doi: https://doi.org/10.1016/B978-0-12-816502-7.00001-4.

[38] V. Giovannetti and G. M. Palma. Master equations for correlated quantum channels. *Phys. Rev. Lett.*, 108:040401, Jan 2012. doi: 10.1103/PhysRevLett.108.040401. URL https://link.aps.org/doi/10.1103/PhysRevLett.108.040401.

[39] F. Avram. On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probab. Theory Related Fields*, 79(1):37–45, 1988. ISSN 0178-8051. doi: 10.1007/BF00319101. URL https://doi.org/10.1007/BF00319101.

[40] S. V. Parter. On the distribution of the singular values of Toeplitz matrices. *Linear Algebra Appl.*, 80:115–130, 1986. ISSN 0024-3795. doi: 10.1016/0024-3795(86)90280-6. URL https://doi.org/10.1016/0024-3795(86)90280-6.

[41] M. M. Wolf, D. Pérez-García, and G. Giedke. Quantum capacities of bosonic channels. *Physical Review Letters*, 98:130501, 2007. doi: 10.1103/PhysRevLett.98.130501.

[42] L. Lami, K. K. Sabapathy, and A. Winter. All phase-space linear bosonic channels are approximately Gaussian dilatable. *New Journal of Physics*, 20(11):113012, 2018.

[43] C. D. Cushen and R. L. Hudson. A quantum-mechanical central limit theorem. *Journal of Applied Probability*, 8(3):454–469, 1971. doi: 10.2307/3212170.

[44] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, 1993.

[45] N. Davis, M. E. Shirokov, and M. M. Wilde. Energy-constrained two-way assisted private and quantum capacities of quantum channels. *Physical Review A*, 97:062310, 2018. doi: 10.1103/PhysRevA.97.062310.

[46] K. Goodenough, D. Elkouss, and S. Wehner. Assessing the performance of quantum repeaters for all phase-insensitive Gaussian bosonic channels. *New Journal of Physics*, 18(6):063005, 2016.

[47] M. Takeoka, S. Guha, and M. M. Wilde. Fundamental rate-loss tradeoff for optical quantum key distribution. *Nature Communications*, 5(1):5235, 2014.

[48] M. M. Wilde, M. Tomamichel, and M. Berta. Converse bounds for private communication over quantum channels. *IEEE Transactions on Information Theory*, 63(3):1792–1817, 2017. ISSN 1557-9654. doi: 10.1109/TIT.2017.2648825.

[49] M. Takeoka, S. Guha, and M. M. Wilde. The squashed entanglement of a quantum channel. *IEEE Transactions on Information Theory*, 60(8):4987–4998, 2014. doi: 10.1109/TIT.2014.2330313.

[50] S. Pirandola, S. L. Braunstein, R. Laurenza, C. Ottaviani, T. P. W. Cope, G. Spedalieri, and L. Banchi. Theory of channel simulation and bounds for private communication. *Quantum Science and Technology*, 3(3):035009, 2018.

[51] G. Wang, C. Ottaviani, H. Guo, and S. Pirandola. Improving the lower bound to the secret-key capacity of the thermal amplifier channel. *The European Physical Journal D*, 73(1):17, 2019. doi: 10.1140/epjd/e2018-90351-0. URL https://doi.org/10.1140/epjd/e2018-90351-0.

[52] L. Lami, S. Khatri, G. Adesso, and M. M. Wilde. Extendibility of bosonic Gaussian states. *Physical Review Letters*, 123:050501, 2019. doi: 10.1103/PhysRevLett.123.050501.

[53] F. Caruso, V. Giovannetti, and A. S. Holevo. One-mode bosonic Gaussian channels: a full weak-degradability classification. *New Journal of Physics*, 8(12):310–310, 2006. doi: 10.1088/1367-2630/8/12/310.

[54] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. Publications of the Scuola Normale Superiore. Scuola Normale Superiore, Pisa, Italy, 2011.

[55] G. Szegő. Beiträge zur Theorie der Toeplitzschen Formen. *Math. Z.*, 6(3):167–202, 1920. doi: 10.1007/BF01199955. URL https://doi.org/10.1007/BF01199955.

[56] U. Grenander and G. Szegő. *Toeplitz forms and their applications*. Chelsea Publishing Co., New York, second edition, 1984. ISBN 0-8284-0321-X.

[57] Robert M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory*, 2(3):155–239, 2006. doi: 10.1561/0100000006. URL http://dx.doi.org/10.1561/0100000006.

[58] A. Böttcher and S. M. Grudsky. *Toeplitz matrices, asymptotic linear algebra, and functional analysis*. Birkhäuser Verlag, Basel, 2000. ISBN 3-7643-6290-1. doi: 10.1007/978-3-0348-8395-5.

[59] A. Böttcher and S. M. Grudsky. *Spectral properties of banded Toeplitz matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005. ISBN 0-89871-599-7. doi: 10.1137/1.9780898717853. URL https://doi.org/10.1137/1.9780898717853.

[60] C. Garoni and S. Serra-Capizzano. *Generalized locally Toeplitz sequences: theory and applications. Vol. I*. Springer, Cham, 2018. ISBN 978-3-030-02232-7; 978-3-030-02233-4. doi: 10.1007/978-3-030-02233-4. URL https://doi.org/10.1007/978-3-030-02233-4.

[61] S. M. Grudsky. Eigenvalues of larger Toeplitz matrices: the asymptotic approach. Lecture notes, 2010.

[62] Ludovico Lami and Mark M. Wilde. Exact solution for the quantum and private capacities of bosonic dephasing channels. *Nature Photonics*, 17:525–530, June 2023. doi: 10.1038/s41566-023-01190-4. URL https://doi.org/10.1038/s41566-023-01190-4.

[63] Y.S. Chow and H. Teicher. Probability theory: Independence interchangeability martingales. 2012.

Part III

Quantum Shannon Theory with finite-dimensional systems

Chapter 8

Computable entanglement cost under PPT operations

Quantum information theory is plagued by the problem of regularisations, which require the evaluation of formidable asymptotic quantities. This makes it computationally intractable to gain a precise quantitative understanding of the ultimate efficiency of key operational tasks such as entanglement manipulation. Here we consider the problem of computing the asymptotic entanglement cost of preparing noisy quantum states under quantum operations with positive partial transpose (PPT). Building on a previous characterisation of the PPT entanglement cost in terms of a regularised formula, we construct a hierarchy of semi-definite programs that bypasses the issue of regularisation altogether, and converges to the true asymptotic value of the entanglement cost. Our main result establishes that this convergence happens exponentially fast, thus yielding an efficient algorithm that approximates the cost up to an additive error ε in time $\mathrm{poly}\big(D, \log(1/\varepsilon)\big)$, where D is the underlying Hilbert space dimension. To our knowledge, this is the first time that an asymptotic entanglement measure is shown to be efficiently computable despite no closed-form formula being available.

[This chapter is based on L. Lami, F. A. Mele, B. Regula, Computable Entanglement Cost under Positive Partial Transpose Operations, Physical Review Letters 134, 090202 (2025)]

8.1 Introduction

Quantum Shannon theory studies the fundamental limitations on the manipulation of quantum information in the presence of external noise. Calculating those limits often involves computing certain functions that encapsulate the ultimate capabilities of information carriers. Paradigmatic examples include the various capacities of quantum channels, such as the classical [1, 2], quantum [3–5], private [5, 6], and entanglement-assisted [7, 8] capacities, but also the operational entanglement measures that tell us how much entanglement can be extracted from a given bipartite quantum state, i.e. the distillable entanglement [9–13], and vice versa, how much entanglement must be invested to *create* that state [14–17]. This latter quantity, the entanglement cost, is the main focus of this work.

With the sole exception of the entanglement-assisted capacity, all of the above functions are expressed by *regularised formulas*, i.e. formulas that involve an explicit limit $n \to \infty$ over the number of uses of the channel or the available copies of the state. For example, by the Lloyd–Shor–Devetak theorem [3–5] the quantum capacity of a channel $\mathcal{N}$ equals $Q(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} I_c\big(\mathcal{N}^{\otimes n}\big)$, where $I_c(\mathcal{N})$ is the 'coherent information' of $\mathcal{N}$, and $\mathcal{N}^{\otimes n}$ represents n parallel uses of $\mathcal{N}$. In stark contrast with classical information theory, for quantum channels it holds in general that $I_c(\mathcal{N}^{\otimes n}) \neq n I_c(\mathcal{N})$, meaning that evaluating the limit cannot be avoided. Such non-additivity is a fundamental feature of most settings encountered in quantum information [18–23]. Analogously, the entanglement cost of preparing a state $\rho = \rho_{AB}$ using local operations and classical communication

(LOCC) is given by

$$E_{c,\,\mathrm{LOCC}}(\rho) = \lim_{n\to\infty} \frac{1}{n} E_f\left(\rho^{\otimes n}\right), \tag{8.1}$$

where E_f is the 'entanglement of formation' [11, 24]. The precise nature of these formulas is not so important here; what is important, however, is that the regularisation $n \to \infty$ makes them analytically hard to control and computationally intractable. Indeed, on the one hand the dimension of the quantum system on which $\mathcal{N}^{\otimes n}$ acts, or to which $\rho^{\otimes n}$ pertains, is exponential in n, quickly rendering numerical calculations infeasible as n grows; on the other, there is no guarantee on the quality of the approximation obtained by stopping at the nth level in any of these formulas — for instance, an unbounded n may be required to even check that the quantum capacity is non-zero [25]. The regularisation thus appears to be an omnipresent curse that stifles almost every attempt to quantitatively understand the ultimate limitations of quantum information manipulation.

But is that really so? In this work we show how to overcome this fundamental obstacle in a specific case, by efficiently calculating a type of entanglement cost — expressed as a regularised quantity — on all quantum states. To this end, we look at a problem that has been studied by many authors [26–30], but for which a full solution had not been found prior to our work: namely, zero-error asymptotic entanglement cost under 'positive partial transpose' (PPT) operations. The roots of PPT transformations lie in the fundamental connection between entanglement and partial transposition identified by Peres [31] and the Horodeckis [32, 33]. In the context of entanglement manipulation, PPT operations were introduced in the pioneering works of Rains [34, 35] as a mathematically natural relaxation of the much more complicated set of LOCC operations, providing a convenient way to gain some insights into the latter and, importantly, giving hope for an easier understanding of some of the fundamental limits of entanglement transformations. Because of this, they attracted significant attention in the operational study of quantum entanglement, but even in this technically simpler setting the fundamental questions in entanglement manipulation remained unsolved.

By introducing an efficient algorithm for the evaluation of the asymptotic PPT entanglement cost, we solve this long-standing problem and exhibit an entanglement measure that is both *computable* and *operationally meaningful* for general — pure and mixed — quantum states. This can be contrasted with other entanglement measures, which are either defined in terms of operational tasks and thus suffer from the problem of regularisation, making efficient evaluation impossible, or they are simply abstract mathematical constructions with no precise operational meaning. Prior to our work, the PPT entanglement cost was only known to be computable for a specific class of quantum states [26] (including pure states, Gaussian states, Werner states, and two-qubit states [36]), with its value given by a celebrated entanglement measure known as the *logarithmic negativity* [26, 37, 38]. The logarithmic negativity is one of the most widely employed entanglement measures, having found numerous applications in all corners of quantum physics [39–46], from quantum field theory [39] to condensed matter [42]. The success of the logarithmic negativity is due to two main reasons: first, its efficient computability [37], which makes it highly suitable for numerical applications, and second, its *partial* operational interpretation, which guarantees that it coincides with the zero-error PPT entanglement cost for this specific class of quantum states [26]. While encouraging, this partial operational interpretation is arguably not satisfactory if this measure is to be applied widely in quantum physics. Indeed, beyond this restricted class of states the negativity is not known to be an operational quantity, and it is unclear whether e.g. the states considered in [40, 41, 46] genuinely belong to the class in question.

It was recently claimed that the PPT entanglement cost can be computed exactly for all states [27, 28], but — as we show below (see also [47]) — this claim relies on some erroneous assertions. In this work, building on the partial results of [27, 28], we find the correct generalisation of the logarithmic negativity that enjoys the operational interpretation of being the zero-error PPT entanglement cost for *all* quantum states. Crucially, our generalisation is also efficiently

computable, requiring almost the same computational time as the logarithmic negativity. This opens the door to numerous applications in quantum physics, serving as a fully operationally motivated alternative to the negativity as a computable entanglement quantifier.

sectionZero-error PPT entanglement cost The goal of zero-error PPT entanglement dilution is to prepare n copies of a given bipartite quantum state $\rho = \rho_{AB}$ by consuming as few singlets Φ_2 (i.e. two-qubit maximally entangled states) as possible and using PPT operations only. Here, a channel Λ is called PPT if its partial transposition $\Gamma \circ \Lambda \circ \Gamma$ is also a valid quantum channel, where Γ is the partial transpose operation [31] defined as $\Gamma(X_A \otimes Y_B) = X_A \otimes Y_B^{\intercal}$. We say that a number R is an *achievable rate* if for all sufficiently large n there exists a PPT operation Λ_n with the property that $\Lambda_n\left(\Phi_2^{\otimes \lfloor Rn \rfloor}\right) = \rho^{\otimes n}$. By definition, the zero-error PPT entanglement cost of ρ, denoted as $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$, is the infimum of all achievable rates R. In this work, we show how to compute $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$ for any quantum state ρ.

In a nutshell, the key reason why many authors [26–30] have been interested in the problem of entanglement dilution under PPT operations is that it provides a more tractable model for the fundamental problem of entanglement dilution under LOCC operations, whose underlying figure of merit, given by (8.1), is computationally inaccessible in most cases of interest. In general, PPT operations are a superset of LOCC, but any PPT channel can be implemented in a stochastic manner by LOCC together with the assistance of PPT states [35, 48]; since it is known that PPT states possess only a weak form of entanglement that cannot be distilled [49], such states can be considered as a 'cheap' resource, thus providing intuition for the PPT operations being a prudent relaxation of the power of LOCC. Such PPT-based approaches have also attracted significant attention in other parts of entanglement theory [34, 35, 50, 51].

Importantly, the optimal performance achievable under PPT operations also establishes bounds and no-go limits on what can be achieved under LOCC in practice. These bounds are often the tightest available, as evidenced in the contexts of channel capacities [52, 53] and entanglement distillation [35]. Let us now consider the task of entanglement dilution under LOCC operations, which is defined as above but with LOCC operations replacing PPT operations. We denote the corresponding figure of merit by $E_{c,\,\mathrm{LOCC}}^{\mathrm{exact}}(\rho)$. Due to PPT being an outer approximation to LOCC, it holds that

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) \leq E_{c,\,\mathrm{LOCC}}^{\mathrm{exact}}(\rho) \tag{8.2}$$

for all bipartite states $\rho = \rho_{AB}$. Our main result allows us to calculate the left-hand side, which is particularly significant because establishing computable lower bounds on $E_{c,\,\mathrm{LOCC}}^{\mathrm{exact}}(\rho)$ is a priori difficult, as to do that one needs to constrain *all* possible LOCC entanglement dilution protocols. (To prove an upper bound, on the contrary, it suffices to exhibit an explicit dilution protocol.)

We also remark that, in our definition, 'zero error' means that we require the transformation of $\Phi_2^{\otimes \lfloor Rn \rfloor}$ into $\rho^{\otimes n}$ to be realised exactly. This is a model of entanglement dilution that has been studied before in several different contexts [16, 26, 54]. The setting contrasts with definitions that allow for an asymptotically vanishing error in the transformation [11, 14, 16] However, in the Supplemental Material (SM) [**?**] we show that no substantial change occurs if we require that the error, instead of being exactly zero, decay to zero sufficiently fast — the relevant figure of merit is then still $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}$.

sectionPrior work In a pioneering paper by Audenaert, Plenio, and Eisert [26] it was shown that the PPT entanglement cost $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$ can be evaluated exactly for all bipartite states $\rho = \rho_{AB}$ that satisfy a condition known as 'zero bi-negativity' [26, 36, 55]. Specifically, if $|\rho^{\Gamma}|^{\Gamma} \geq 0$, where X^{Γ} denotes the partial transpose $\Gamma(X)$ of an operator X and $|X| := \sqrt{X^{\dagger}X}$ is its absolute value, then

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = \log_2 \left\|\rho^{\Gamma}\right\|_1 =: E_N(\rho)\,. \tag{8.3}$$

The expression on the right, E_N, is the *logarithmic negativity* [37, 38], obtained by simply evaluating the trace norm $\|\cdot\|_1 := \mathrm{Tr}\,|\cdot|$ of the partially transposed state. This framework thus provides a partial operational interpretation for E_N through its equality with the zero-error PPT cost $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}$

for some states. At the same time, Eq. (8.3) effectively solves the problem of computing $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ for states with zero bi-negativity, because E_N is efficiently computable via a *semi-definite program* (SDP) [56–58]:

$$E_N(\rho) = \log_2 \min\left\{\mathrm{Tr}\, S :\ -S \leq \rho^\Gamma \leq S\right\}. \tag{8.4}$$

But what to do for those states that have non-zero bi-negativity, i.e. satisfy $|\rho^\Gamma|^\Gamma \not\geq 0$? In their recent works [27, 28, 47], Wang and Wilde showed that $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ can be expressed as a regularisation of another quantity called E_κ. They then made an even stronger claim that E_κ is in fact *additive*, meaning that $E_\kappa\left(\rho^{\otimes n}\right) \stackrel{?}{=} n\, E_\kappa(\rho)$ holds true for all bipartite states ρ and all positive integers n, thus completely eliminating the issue of regularisation. This would imply a general computable solution for $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$: it would simply coincide with E_κ, which is computable via an SDP [27, 28, 47]. However, in Lemma 4 below we construct a simple counterexample that disproves the claimed additivity of E_κ. Its existence shows that in general $E_\kappa \neq E_{c,\mathrm{PPT}}^{\mathrm{exact}}$, thus invalidating the computable solution claimed in [27, 28] and reopening the question of whether the asymptotic cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ can be efficiently evaluated. Further details concerning the claims of [27, 28] can be found in [47].

The equivalence between the regularisation of the quantity E_κ and the PPT entanglement cost revealed in [27, 28] will still prove useful to us, albeit *a priori* it is not clear how it could lead to a computable formula for $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ — the daunting problem of regularisation persists.

Main results. — In this work, we completely solve the problem of computing the asymptotic zero-error PPT entanglement cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$. To do this, we construct a converging hierarchy of semi-definite programs that can be used to calculate $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ for any given state to any degree of precision *efficiently*, i.e. in time polynomial in the underlying Hilbert space dimension and in $\log(1/\varepsilon)$, with ε being the additive error. The key quantities in our approach are a family of PPT entanglement monotones indexed by an integer $p \in \mathbb{N}$ and given by

$$E_{\chi,p}(\rho) := \log_2 \chi_p(\rho)\,, \tag{8.5}$$

where

$$\chi_p(\rho) := \min\left\{\mathrm{Tr}\, S_p :\ -S_i \leq S_{i-1}^\Gamma \leq S_i,\ i=0,...,p,\ S_{-1}=\rho\right\} \tag{8.6}$$

is an SDP with variables $S_0, \ldots, S_p$. These quantities are increasing in p for every fixed ρ, and we refer to them as the $\boldsymbol{\chi}$*-hierarchy*. Note also that $E_{\chi,0} = E_N$, hence $E_{\chi,p}$ can be regarded as a generalisation of the logarithmic negativity [37, 38].

Importantly, we show that the χ-hierarchy approximates the entanglement cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ from below, in the sense that $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) \geq E_{\chi,p}(\rho)$ for all states ρ and all p. The proof of this fact, which can be found in the Supplemental Material [**?**], relies on the connection between the entanglement cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ and the regularised form of E_κ shown in [27, 28]. Our first main result, the forthcoming Theorem 1, establishes that this approximation becomes increasingly tight as p increases, and the χ-hierarchy gives the value of $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ *exactly* in the limit $p \to \infty$. This allows us to replace the limit in the number of copies n, which is what makes $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ difficult to compute, with a limit in the hierarchy level p. This already provides a 'single-letter' formula for the PPT entanglement cost that no longer suffers from the curse of regularisation. However, because of the limiting procedure $p \to \infty$, it is still unclear if the expression can be evaluated easily. Crucially, in Theorem 1 we also show that calculating the limit of the χ-hierarchy is indeed significantly easier than evaluating regularised expressions: the convergence to the true value of $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ is *exponentially fast* uniformly on all states, which opens the way to an accurate calculation of $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ in practice.

Theorem 1 (Exact expression of the cost)**.** *For all bipartite states $\rho = \rho_{AB}$ on a system of minimal local dimension $d := \min\{|A|, |B|\} \geq 2$, and all positive integers $p \in \mathbb{N}^+$, it holds that*

$$0 \leq E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) - E_{\chi,p}(\rho) \leq \log_2 \frac{1}{1-\left(1-\frac{2}{d}\right)^p}\,, \tag{8.7}$$

entailing that

$$E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = \lim_{p\to\infty} E_{\chi,p}(\rho)\,. \tag{8.8}$$

The proof of Theorem 1 is outlined in the End Matter, with full details provided in the SM for interested readers [**?**]. Note that for every fixed value of d and for large p, the approximation error on the right-hand side of (8.7) can be estimated as $\left(1-\frac{2}{d}\right)^p (\log_2 e)$. In other words, the speed of convergence in (8.8) is exponential in p and furthermore independent of ρ.

The single-letter formula (8.8) can be used to establish two notable properties of the zero-error PPT entanglement cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$, namely additivity and continuity [**?** , § 8.4.3–8.4.3]. Also, Theorem 1 yields immediately a simple solution in the qubit-qudit case ($d = 2$), generalising Ishizaka's result that $E_{c,\mathrm{PPT}}^{\mathrm{exact}} = E_N$ for all two-qubit states [36]:

Corollary 2 (Simple formula for the cost of qubit-qudit states)**.** *For all states $\rho = \rho_{AB}$ on a $2 \times n$ bipartite quantum system, it holds that*

$$\begin{aligned} E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) &= E_{\chi,1}(\rho) \\ &= \log_2 \min\left\{ \left\| S_0^\Gamma \right\|_1 : -S_0 \le \rho^\Gamma \le S_0 \right\}. \end{aligned} \tag{8.9}$$

But the most important implication of Theorem 1 is that it allows us to construct an efficient algorithm that calculates $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ to any desired accuracy. This algorithm takes as input a bipartite state $\rho = \rho_{AB}$ and an error tolerance $\varepsilon > 0$, and returns as output a number $\tilde{E}$ such that $\left| E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) - \tilde{E} \right| \le \varepsilon$. It works as follows:

1. Use (8.7) to find p large enough so that $E_{\chi,p}(\rho)$ approximates $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$ up to an error $\epsilon/2$. A value of $p = \mathcal{O}(d \log(d/\epsilon))$ is sufficient.

2. Solve the SDP in (8.6) up to an (additive) error $\epsilon \ln(2)/2$. Taking the logarithm yields an estimate $\tilde{E}$ of $E_{\chi,p}(\rho)$ up to an (additive) error $\epsilon/2$ (see (8.5)).

3. Return $\tilde{E}$.

The time complexity of the above algorithm is analysed in the forthcoming Theorem 3, which shows the core result of our work: the zero-error PPT entanglement cost can be efficiently computed. The key observation behind this result is that climbing the χ-hierarchy up to level p introduces only polynomially many more constraints in the SDP in (8.5) and is thus relatively inexpensive.

Theorem 3 (Time to compute the cost)**.** *Let ρ be a bipartite state on a system of total dimension D and minimal local dimension d. Then, the above algorithm computes $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$ up to an additive error ε in time*

$$\mathcal{O}\Big((dD)^{6+o(1)} \operatorname{polylog}(1/\varepsilon) \Big)\,. \tag{8.10}$$

The proof of Theorem 3 is sketched in the End Matter. Remarkably, Theorem 3 implies that the time complexity required to compute $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ is only marginally larger than that of computing the logarithmic negativity E_N, which is bounded by the time complexity required for the diagonalisation of a $D \times D$ matrix — also polynomial in D and $\log(1/\varepsilon)$. In short, while both $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ and E_N can be computed efficiently, $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ stands out because, by definition, it has an operational meaning for all quantum states, unlike E_N. Our result is the first of its kind for two distinct reasons:

(a) First, because it establishes the efficient computability of an operationally meaningful asymptotic entanglement measure (i.e. a distillable entanglement or an entanglement cost). There is no known algorithm to estimate any other such measure, not even under the simplifying zero-error assumption.

(b) Second, because efficient computability is shown without exhibiting a closed-form single-letter formula, but rather by describing a converging SDP hierarchy. To the extent of our knowledge, the only other case in quantum information theory where a similar situation arises is in [59, Theorem 5.1]. However, unlike ours, the algorithm described there is computationally extremely expensive, featuring an exponential dependence on d^3/ε. More generally, expressing difficult-to-compute quantities through converging SDP hierarchies is a technical tool that has found various uses in quantum information [53, 60–64], but such applications typically do not result in efficiently computable algorithms or do not yield exact operational results.

8.2 Discussion and conclusions

In this paper, we have provided a solution to the problem of efficiently calculating the zero-error PPT entanglement cost of arbitrary (finite-dimensional) quantum states. To the best of our knowledge, it is the first time that any operational asymptotic entanglement measure is shown to be efficiently computable. A particularly interesting feature of our construction is that it does not rely on a closed-form formula, but rather on a converging hierarchy of semi-definite programs that approximate the cost from above and below with controllable error.

Our solution identifies the correct generalisation of the celebrated logarithmic negativity, which became popular in quantum physics for its efficient computability, despite being operationally meaningful only for states with zero bi-negativity. In contrast, our entanglement measure, $\lim_{p\to\infty} E_{\chi,p}$, is not only efficiently computable — specifically, it can be efficiently computed via an algorithm with almost the same time complexity as the logarithmic negativity — but it also has operational meaning for *all* quantum states, through its equality with the zero-error PPT entanglement cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$. For states with zero bi-negativity, our entanglement measure coincides with the logarithmic negativity, but for other states it can be significantly smaller.

An open question in our analysis is whether the χ-hierarchy collapses at any finite level for some — or even all — states. We found examples of states ρ such that $E_N(\rho) = E_{\chi,0}(\rho) < E_{\chi,1}(\rho)$ and also $E_{\chi,1}(\rho) < E_{\chi,2}(\rho)$, but we were not able to ascertain whether there exists in general a gap between $E_{\chi,2}$ and $E_{\chi,3}$. If $E_{\chi,2} = E_{\chi,3}$ holds in general, then this would mean that $E_{c,\mathrm{PPT}}^{\mathrm{exact}} = E_{\chi,2}$, and thus the cost could be computed with a simple single-letter formula. While this would be a considerable simplification from the analytical standpoint, we stress that it will only entail a $\mathrm{poly}(d)$ improvement in the time complexity of evaluating it numerically.

Note. The issue with the original argument by Wang and Wilde that leads to the additivity violation for E_κ (Lemma 4) is discussed in detail in the erratum [47].

8.3 Methods

The task. We start by defining $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$ in rigorous terms. A (quantum) channel $\Lambda : X \to Y$ is a completely positive and trace preserving map taking as input states of a quantum system X and outputting states of Y. The set of completely positive maps (respectively, quantum channels) from X to Y will be denoted as $\mathrm{CP}(X \to Y)$ (respectively, $\mathrm{CPTP}(X \to Y)$). If $X = AB$ and $Y = A'B'$ are both bipartite systems and $\Lambda \in \mathrm{CP}(AB \to A'B')$, we say that Λ is *PPT* if $\Gamma_{B'} \circ \Lambda \circ \Gamma_B$ is still completely positive, where Γ_B denotes the partial transpose on B, and analogously for $\Gamma_{B'}$. Another way to understand this class of channels is to realise that they completely preserve the set of PPT states, in the sense that $\Gamma_{B_1B_2'}[\mathrm{id}\otimes\Lambda(\sigma_{A_1B_1A_2B_2})] \geq 0$ for any state σ with $\Gamma_{B_1B_2}(\sigma_{A_1B_1A_2B_2}) \geq 0$. We can then define the *zero-error PPT entanglement cost* of

any bipartite state $\rho = \rho_{AB}$ as

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) := \inf\Big\{R\colon \text{ for all sufficiently large } n \in \mathbb{N} \\ \exists\ \Lambda_n \in \mathrm{PPT} \cap \mathrm{CPTP}\colon\ \Lambda_n\Big(\Phi_2^{\otimes \lfloor Rn \rfloor}\Big) = \rho^{\otimes n}\Big\}. \tag{8.11}$$

Here, $\Phi_2 := |\Phi_2\rangle\langle\Phi_2|$, where $|\Phi_2\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is the two-qubit maximally entangled state, i.e. the *ebit*, and Λ_n is required to be a PPT channel.

The quantifier E_κ. Wang and Wilde [27, 28, 47] introduced and studied the SDP-computable quantity

$$E_\kappa(\rho) := \log_2 \min\left\{\mathrm{Tr}\, S :\ -S \leq \rho^\Gamma \leq S,\ S^\Gamma \geq 0\right\}. \tag{8.12}$$

Among other things, they showed that: (i) E_κ is monotonically non-increasing under PPT channels; (ii) $E_\kappa(\rho) \geq E_N(\rho)$ for all states ρ, with equality when ρ has zero bi-negativity; (iii) E_κ is sub-additive, meaning that

$$E_\kappa(\rho \otimes \rho') \leq E_\kappa(\rho) + E_\kappa(\rho') \tag{8.13}$$

for all pairs of states $\rho = \rho_{AB}$ and $\rho' = \rho'_{A'B'}$; and (iv) its regularisation yields the zero-error PPT entanglement cost, i.e.

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_\kappa^\infty(\rho) := \lim_{n\to\infty} \frac{1}{n} E_\kappa\left(\rho^{\otimes n}\right). \tag{8.14}$$

As said, it was claimed in [27] that E_κ is additive, meaning that equality holds in (8.13). However, this claim is incorrect (see also [47]). To disprove it, it is useful to first note that additivity indeed holds when both ρ and ρ' have zero bi-negativity, simply because in that case E_κ coincides with the logarithmic negativity by property (ii), and this latter measure *is* additive, as one sees immediately by looking at its definition in terms of the 1-norm of the partially transposed state. Hence, our search for a counterexample must start with the construction of states with non-zero bi-negativity. That such states do exist was reported already in [36, 55] based on numerical evidence. However, here we present a simpler, analytical construction.

Punch card states. Let $A \geq 0$ be a positive semi-definite $d \times d$ matrix, and let Q be another $d \times d$ symmetric matrix with only 0/1 entries and all 1's on the main diagonal (i.e. such that $Q_{ii} = 1$ for all i). The associated *punch card state* is the bipartite quantum state on $\mathbb{C}^d \otimes \mathbb{C}^d$ defined by

$$\pi_{A,Q} := \frac{1}{N_{A,Q}} \left(\sum_{i,j} A_{ij}\, |ii\rangle\langle jj| + \sum_{i\neq j} Q_{ij}|A_{ij}|\, |ij\rangle\langle ij| \right), \tag{8.15}$$

where $N_{A,Q}$ is chosen so that $\mathrm{Tr}\,\pi_{A,Q} = 1$. It can be verified that

$$\pi_{A,Q}^\Gamma \simeq \frac{1}{N_{A,Q}} \left[\left(\textstyle\sum_i A_{ii}\, |ii\rangle\langle ii|\right) \oplus \bigoplus_{i<j} \begin{pmatrix} Q_{ij}|A_{ij}| & A_{ij} \\ A_{ij}^* & Q_{ij}|A_{ij}| \end{pmatrix} \right], \tag{8.16}$$

from which, using the identity

$$\left| \begin{pmatrix} Q_{ij}|A_{ij}| & A_{ij} \\ A_{ij}^* & Q_{ij}|A_{ij}| \end{pmatrix} \right| = \begin{pmatrix} |A_{ij}| & Q_{ij}A_{ij} \\ Q_{ij}A_{ij}^* & |A_{ij}| \end{pmatrix}, \tag{8.17}$$

valid because $Q_{ij} \in \{0, 1\}$, one derives that

$$\left|\pi_{A,Q}^\Gamma\right|^\Gamma = \frac{1}{N_{A,Q}} \left(\sum_{i,j} Q_{ij}A_{ij}\, |ii\rangle\langle jj| + \sum_{i\neq j} |A_{ij}|\, |ij\rangle\langle ij| \right). \tag{8.18}$$

Therefore, if A and Q are chosen such that $Q \circ A \not\geq 0$, where $\circ$ denotes the Hadamard (i.e. entry-wise) product between matrices, $\pi_{A,Q}$ will have non-zero bi-negativity, i.e. $\left|\pi_{A,Q}^{\Gamma}\right|^{\Gamma} \not\geq 0$. It is easy to construct examples of A and Q that meet the above criteria, the simplest one being

$$A_0 := \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad Q_0 := \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{8.19}$$

Having constructed a state with non-zero bi-negativity, we can wonder whether two copies of it already violate the additivity of E_κ. And sure enough, they do:

Lemma 4. *The punch card state $\pi_0 := \pi_{A_0,Q_0}$ defined by (8.15) with the substitution (8.19) satisfies*

$$1.001 \approx E_\kappa\left(\pi_0^{\otimes 2}\right) < 2E_\kappa(\pi_0) \approx 1.029\,. \tag{8.20}$$

In particular, the sub-additivity of E_κ and Lemma 4 imply that $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\pi_0) \leq \frac{1}{2} E_\kappa\left(\pi_0^{\otimes 2}\right) < E_\kappa(\pi_0)$ and therefore that $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\pi_0) \neq E_\kappa(\pi_0)$, thus invalidating the main claim of the works [27, 28].

Two SDP hierarchies. Recall that in (8.5)–(8.6) we introduced the χ-hierarchy $E_{\chi,p}(\rho) = \log_2 \chi_p(\rho)$ with

$$\chi_p(\rho) := \min\left\{ \operatorname{Tr} S_p :\ -S_i \leq S_{i-1}^{\Gamma} \leq S_i,\ i = 0, ..., p,\ S_{-1} = \rho \right\}, \tag{8.21}$$

which constitutes a generalisation of the logarithmic negativity. We now introduce another complementary hierarchy of SDPs, the ***κ*-hierarchy**, defined for $q \in \mathbb{N}^+$ ($q \geq 1$) by

$$E_{\kappa,q}(\rho) := \log_2 \kappa_q(\rho)\,, \tag{8.22}$$

with

$$\begin{aligned} \kappa_q(\rho) := \min\Big\{ \operatorname{Tr} S_{q-1} :\ & -S_i \leq S_{i-1}^{\Gamma} \leq S_i,\ i = 0, ..., q-1, \\ & S_{-1} = \rho,\ S_{q-1}^{\Gamma} \geq 0 \Big\}. \end{aligned} \tag{8.23}$$

Observe the resemblance to the definition of χ_p in (8.21): the only difference between the two optimisations is the condition $S_{p-1}^{\Gamma} \geq 0$; adding that to (8.21) yields immediately (8.23) with $q \mapsto p$, and indeed in that case the optimal S_p would automatically be $S_p = S_{p-1}^{\Gamma}$. Furthermore, $E_{\kappa,1} = E_\kappa$ coincides with the quantity (8.12) introduced by Wang and Wilde, of which the κ-hierarchy thus constitutes a generalisation.

In the SM [**?**] we explore the properties of the quantities $E_{\chi,p}$ and $E_{\kappa,q}$, showing them to be legitimate entanglement measures. In particular, the functions are all suitably normalised, continuous, faithful on PPT states, and strongly monotonic under PPT operations. The pivotal property that distinguishes $E_{\chi,p}$ from $E_{\kappa,q}$ is that, while the quantities $E_{\kappa,q}$ are only sub-additive, the χ-quantities are fully additive under tensor products, meaning that regularisation can always be avoided.

Two key insights lead to the proof of our main results. First, that the two hierarchies provide complementary bounds on $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$, the fundamental quantity we want to estimate: namely, the χ-hierarchy gives increasing lower bounds on $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$, while the κ-hierarchy gives decreasing upper bounds on it. In other words, on any fixed state ρ

$$E_{\chi,0} \leq E_{\chi,1} \leq ... \leq E_{c,\mathrm{PPT}}^{\mathrm{exact}} = E_\kappa^\infty \leq ... \leq E_{\kappa,2} \leq E_{\kappa,1}. \tag{8.24}$$

In particular, $E_{\chi,p}$ is increasing in p, while $E_{\kappa,q}$ is decreasing in q. Eq. (8.24) immediately shows the remarkable connection between two very different limits: one in the number of copies n, which is needed to compute E_κ^∞ (see Eq. (8.14)), and one in the hierarchy levels p and q.

The second insight is that there is a connection between the χ- and κ-hierarchies, as expressed by the following key technical result, proven in the SM [**?**].

Proposition 5. *For all states $\rho = \rho_{AB}$ on a system of minimal local dimension $d := \min\{|A|, |B|\} \geq 2$, and all $p \in \mathbb{N}^+$,*

$$\kappa_p(\rho) \leq \frac{d}{2}\,\chi_p(\rho) - \left(\frac{d}{2} - 1\right)\chi_{p-1}(\rho)\,. \tag{8.25}$$

With Proposition 5 at hand, we can now see how it implies our two main results, Theorems 1 and 3.

Proof sketch of Theorem 1. Combining (8.24) and (8.25) shows that

$$2^{E^{\text{exact}}_{c,\,\text{PPT}}(\rho)} \leq \kappa_p(\rho) \leq \frac{d}{2}\,\chi_p(\rho) - \left(\frac{d}{2} - 1\right)\chi_{p-1}(\rho)\,. \tag{8.26}$$

The quantity that we are really interested in, however, is the normalised difference between $\chi_p(\rho)$ and its claimed limiting value $2^{E^{\text{exact}}_{c,\,\text{PPT}}(\rho)}$. To see what the above inequality tells us in this respect, we can define the quantity

$$\varepsilon_p(\rho) := 1 - \frac{\chi_p(\rho)}{2^{E^{\text{exact}}_{c,\,\text{PPT}}(\rho)}}\,, \tag{8.27}$$

by means of which (8.26) can be cast as $\varepsilon_p(\rho) \leq \left(1 - \frac{2}{d}\right)\varepsilon_{p-1}(\rho)$. Iterating the above relation gives

$$\varepsilon_p(\rho) \leq \left(1 - \frac{2}{d}\right)^p \varepsilon_0(\rho) \leq \left(1 - \frac{2}{d}\right)^p\,, \tag{8.28}$$

which entails (8.7) after elementary algebraic manipulations. Taking the limit $p \to \infty$ in (8.7) proves also (8.8). □

Proof sketch of Theorem 3. It suffices to formalise the qualitative argument provided below the statement of the theorem. For $d > 2$, we first choose

$$p_d := \left\lceil \frac{\log_2(2d/\varepsilon)}{-\log_2\left(1 - \frac{2}{d}\right)} \right\rceil = \mathcal{O}(d \log(d/\varepsilon))\,, \tag{8.29}$$

so that, with the notation in (8.27), $\varepsilon_{p_d}(\rho) \leq \frac{\varepsilon}{2d}$. Using $\ln(2)\,|a - b| \leq \left|2^a - 2^b\right|$, valid for $a, b \geq 0$, we get

$$0 \leq E^{\text{exact}}_{c,\,\text{PPT}}(\rho) - E_{\chi,p_d}(\rho) \leq \frac{2^{E^{\text{exact}}_{c,\,\text{PPT}}(\rho)}\varepsilon}{2d \ln 2} \leq \frac{\varepsilon}{2 \ln 2}\,, \tag{8.30}$$

where the last inequality is a consequence of the fact that every state can be created via a quantum teleportation protocol — and hence with PPT operations — from a maximally entangled state, which entails that $E^{\text{exact}}_{c,\,\text{PPT}}(\rho) \leq \log_2 d$ for all ρ. We then solve the SDP for $\chi_{p_d}(\rho)$ up to an additive error $(\ln 2 - 1/2)\,\varepsilon$ by running an optimised SDP solver [65, 66]. Doing so yields an approximation of $E_{\chi,p_d}(\rho)$ up to an additive error $\left(1 - \frac{1}{2\ln 2}\right)\varepsilon$. Adding this up with the error in (8.30) yields a total error of ε. The time complexity in (8.10) can be calculated using known theoretical bounds on the complexity of SDPs, e.g. those found in [65]. □

8.4 Supplementary Material

8.4.1 Introduction

Quantum states and channels

The quantum systems we consider in this work are all represented by finite-dimensional Hilbert spaces $\mathcal{H}$. We will denote the set of linear operators on $\mathcal{H}$ by $\mathcal{L}(\mathcal{H})$. We occasionally refer to these operators also as matrices, to underline the fact that they are understood to be finite dimensional.

Quantum states pertaining to a quantum system with Hilbert space $\mathcal{H}$ are represented by *density operators*, i.e. positive semi-definite operators on $\mathcal{H}$ with unit trace. We will denote the set of positive semi-definite operators on $\mathcal{H}$ by $\mathcal{L}_+(\mathcal{H})$. Given two Hermitian operators $X, Y \in \mathcal{L}_+(\mathcal{H})$ we write $X \leq Y$ if $Y - X \in \mathcal{L}_+(\mathcal{H})$, and $X < Y$ if $Y - X$ is strictly positive definite. The relations $\geq$ and $>$ are defined similarly.

The *trace norm* of an arbitrary operator $X \in \mathcal{L}(\mathcal{H})$ is given by

$$\|X\|_1 \coloneqq \operatorname{Tr} \sqrt{X^\dagger X}\,, \tag{8.31}$$

where for a positive semi-definite operator A its square root is constructed as the unique semi-definite operator $\sqrt{A}$ such that $\sqrt{A}^2 = A$. If $X = X^\dagger$ is Hermitian, then it is possible to give the following simple characterisation of its trace norm. We include a proof for the sake of completeness, as the lemma below will be used a few times throughout this paper.

Lemma 6 (Variational characterisation of the trace norm)**.** *Let $X = X^\dagger \in \mathcal{L}(\mathcal{H})$ be a Hermitian matrix. Then*

$$\|X\|_1 = \min\left\{\operatorname{Tr} Y :\ -Y \leq X \leq Y\right\}. \tag{8.32}$$

Proof. Let $X = \sum_i x_i \,|i\rangle\langle i|$ be the spectral decomposition of X. Then by inspection we find $\|X\|_1 = \sum_i |x_i|$. On the one hand, setting $Y = |X| \coloneqq \sum_i |x_i| \,|i\rangle\langle i|$ we find that $\min\left\{\operatorname{Tr} Y :\ -Y \leq X \leq Y\right\} \leq \operatorname{Tr} \sum_i |x_i| \,|i\rangle\langle i| = \|X\|_1$. On the other, for an arbitrary Y such that $-Y \leq X \leq Y$ we have that $-\langle i|Y|i\rangle \leq \langle i|X|i\rangle = x_i \leq \langle i|Y|i\rangle$, implying that $\langle i|Y|i\rangle \geq |x_i|$. Therefore, $\operatorname{Tr} Y = \sum_i \langle i|Y|i\rangle \geq \sum_i |x_i| = \|X\|_1$; since Y was arbitrary, this concludes the proof. □

Given two quantum systems with Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, a linear map $\Lambda : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ is called *positive* if $\Lambda\Big(\mathcal{L}_+(\mathcal{H})\Big) \subseteq \mathcal{L}_+(\mathcal{H}')$, and *completely positive* if $I_k \otimes \Lambda$ is positive for all $k \in \mathbb{N}$, where I_k is the identity map on $\mathcal{L}\Big(\mathbb{C}^k\Big)$. It is well known that complete positivity can be formulated as a positive semi-definite constraint using the formalism of Choi states. Namely, introducing the maximally entangled state

$$\Phi_d \coloneqq |\Phi_d\rangle\langle\Phi_d|\ , \qquad |\Phi_d\rangle \coloneqq \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |ii\rangle\ , \tag{8.33}$$

a linear map $\Lambda : \mathcal{L}\Big(\mathbb{C}^d\Big) \to \mathcal{L}\Big(\mathbb{C}^{d'}\Big)$ is completely positive if and only if

$$\Big(I_d \otimes \Lambda\Big)(\Phi_d) \geq 0 \tag{8.34}$$

is a positive semi-definite operator.

A completely positive map and trace-preserving map $\Lambda : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ is called a *quantum channel*. In what follows, $\mathrm{CP}(X \to Y)$ and $\mathrm{CPTP}(X \to Y)$ will denote respectively the set of completely positive maps and the set of quantum channels between two quantum systems X and Y.

Positive partial transpose

Composite quantum systems are represented by the tensor product of the local Hilbert spaces; for the simplest case of a bipartite system, this is expressed in formula as $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. In what follows, we will often need to consider the minimal local dimension of a bipartite quantum system AB, i.e. the number $d = \min\{|A|, |B|\}$, where we denote by $|A|$ (respectively, $|B|$) the dimension of the local Hilbert space $\mathcal{H}_A$ (respectively, $\mathcal{H}_B$).

Given an arbitrary bipartite quantum system AB, the *partial transpose* on B is the linear map $\Gamma_B : \mathcal{L}\Big(\mathcal{H}_{AB}\Big) \to \Big(\mathcal{H}_{AB}\Big)$ uniquely defined by

$$\Gamma_B(X_A \otimes Y_B) = (X_A \otimes Y_B)^\Gamma \coloneqq X_A \otimes Y_B^\intercal \qquad \forall\ X_A \in \mathcal{L}(\mathcal{H}_A)\,,\ Y_B \in \mathcal{L}(\mathcal{H}_B)\,. \tag{8.35}$$

Here, $\intercal$ denotes the transposition with respect to a fixed basis of $\mathcal{H}_B$. Note that

$$\operatorname{Tr} Z_{AB} W_{AB} = \operatorname{Tr} Z_{AB}^{\Gamma} W_{AB}^{\Gamma} \qquad \forall\, Z, W \in \mathcal{L}(\mathcal{H}_{AB})\,, \tag{8.36}$$

and in particular

$$\operatorname{Tr} Z_{AB}^{\Gamma} = \operatorname{Tr} Z_{AB} \qquad \forall\, Z \in \mathcal{L}(\mathcal{H}_{AB})\,. \tag{8.37}$$

The definition of Γ_B does depend on which basis we choose, but all choices are unitarily equivalent to each other, in the sense that for any two different partial transpositions Γ, $\widetilde{\Gamma}$ there exists a unitary U_B acting on B such that

$$\widetilde{\Gamma}_B(Z_{AB}) = U_B\, \Gamma_B(Z_{AB})\, U_B^{\dagger}. \tag{8.38}$$

A state σ_{AB} on a bipartite quantum system AB is said to be a *PPT state* if $\sigma_{AB}^{\Gamma} = \Gamma_B(\sigma_{AB}) \geq 0$. The PPT condition is important in quantum information because it provides a relaxation of separability: any *separable* (i.e. unentangled) state σ_{AB}, which can thus be decomposed as $\sigma_{AB} = \sum_x p_x\, \alpha_x^A \otimes \beta_x^B$ for some probability distribution p and some sets of local states α_x^A and β_x^B [67], is PPT, but the converse is famously not true [49, 68].

A completely positive map $\Lambda \in \mathrm{CP}(AB \to A'B')$ with bipartite quantum systems AB and $A'B'$ as input and output is called *PPT* if

$$\Gamma_{B'} \circ \Lambda \circ \Gamma_B \in \mathrm{CP}(AB \to A'B')\,. \tag{8.39}$$

If Λ is in addition also trace preserving, then we call it a *PPT channel* (or a *PPT operation*). Such maps are sometimes also referred to as 'completely PPT preserving'. The importance of the set of PPT channels is that it provides an outer approximation to the set of quantum channels that can be implemented with local operations and classical communication (LOCC) on a bipartite system AB. The usefulness of the PPT approximation rests on the key observation that condition (8.39) amounts to a positive semi-definite constraint, as can be seen by employing (8.34).

Zero-error PPT entanglement cost

The *zero-error PPT entanglement cost* of a bipartite state $\rho = \rho_{AB}$ is defined as the minimum rate of *ebits* (or singlets)

$$\Phi_2 := |\Phi_2\rangle\langle\Phi_2|\,, \qquad |\Phi_2\rangle := \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right), \tag{8.40}$$

that need to be consumed in order to create copies of ρ with PPT operations and zero error. In formula,

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) := \inf\Big\{R:\ \text{for all sufficiently large } n \in \mathbb{N}\ \exists\ \Lambda_n \in \mathrm{PPT} \cap \mathrm{CPTP}:\ \Lambda_n\Big(\Phi_2^{\otimes \lfloor Rn \rfloor}\Big) = \rho^{\otimes n}\Big\}. \tag{8.41}$$

Wang and Wilde proved [27, 47] that $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}$ can be alternatively computed as[1]

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_{\kappa}^{\infty}(\rho) := \lim_{n\to\infty} \frac{1}{n} E_{\kappa}\left(\rho^{\otimes n}\right), \tag{8.42}$$

$$E_{\kappa}(\rho) := \log_2 \min\Big\{\operatorname{Tr} S:\ -S \leq \rho^{\Gamma} \leq S,\ S^{\Gamma} \geq 0\Big\}. \tag{8.43}$$

The first equality in (8.42) is an asymptotic consequence of the more general one-shot statement that the PPT entanglement cost of generating a single copy of a state ρ with PPT operations,

[1] Note that E_{κ} is clearly finite for all finite-dimensional states, as one deduces by making the ansatz $S = \mathbb{1}$ in the optimisation (8.43). In fact, using the convexity of the semi-definite program inside the logarithm in (8.43) together with the Schdmit decomposition for pure states on a bipartite system, it is not difficult to show that E_{κ} is finite provided that $\min\{|A|, |B|\} < \infty$.

denoted by $E^{(1)}_{c,\mathrm{PPT}}(\rho)$, can be bounded as $\log_2\left(2^{E_\kappa(\rho)}-1\right) \leq E^{(1)}_{c,\mathrm{PPT}}(\rho) \leq \log_2\left(2^{E_\kappa(\rho)}+2\right)$ [27, Proposition 2]. It is easy to see that the function E_κ is sub-additive, so that by Fekete's lemma [69] we can alternatively write the limit in (8.42) as an infimum over n, i.e.

$$E^\infty_\kappa(\rho) = \inf_{n\in\mathbb{N}^+} \frac{1}{n} E_\kappa\left(\rho^{\otimes n}\right). \tag{8.44}$$

Unfortunately, due to an error in the derivation of the super-additivity inequality, which can be traced back to [27, Eq. (S59)], E_κ is no longer known to be additive. Hence, the regularisation in the above identity is needed. And in fact, we will show in the forthcoming Section 8.4.2 that it *must* be included, because E_κ is indeed *not* additive in general.

We refer the interested reader to the erratum [47] for a detailed explanation of this issue and of its consequences. There, Wang and Wilde also present other examples of additivity violations for E_κ. In the next section, we will construct our own examples of additivity violation.

8.4.2 Violations of additivity of E_κ

To construct an example of additivity violation, it is useful to start by asking the opposite question: when is it that we know $E_\kappa(\rho)$ to be additive? There is a simple answer to this question that is already discussed in [27] based on the pioneering work [26]: E_κ is easily seen to be additive on many copies of ρ whenever ρ has zero bi-negativity, i.e. whenever $\left|\rho^\Gamma\right|^\Gamma \geq 0$. When this happens, then

$$E_\kappa\left(\rho^{\otimes n}\right) = \log_2 \left\|\left(\rho^{\otimes n}\right)^\Gamma\right\|_1 = n \log_2 \left\|\rho^\Gamma\right\|_1 = nE_N(\rho) = nE_\kappa(\rho)\,, \tag{8.45}$$

for all n. This observation tells us that to look for additivity violations of E_κ we need to construct states with non-zero bi-negativity. After a bit of trial and error, we homed in on the following construction.

Definition 7 (Punch card states)**.** *Let $A \geq 0$ be a $d\times d$ positive semi-definite matrix. Also, let Q be a $d\times d$ symmetric matrix with $Q_{ij}\in\{0,1\}$ for all $i,j=1,\ldots,d$, and $Q_{ii}=1$ for all $i=1,\ldots,d$. Then the associated punch card state is a bipartite quantum state acting on $\mathbb{C}^d\otimes\mathbb{C}^d$ and defined by*

$$\pi_{A,Q} \propto \sum_{i,j} A_{ij}\,|ii\rangle\langle jj| + \sum_{i\neq j} Q_{ij}|A_{ij}|\,|ij\rangle\langle ij|\,, \tag{8.46}$$

with the normalisation $\operatorname{Tr}\pi_{A,Q}=1$.

Note that

$$\begin{aligned}\pi^\Gamma_{A,Q} &\propto \sum_i A_{ii}\,|ii\rangle\langle ii| + \sum_{i<j}\left(Q_{ij}|A_{ij}|\,(|ij\rangle\langle ij| + |ji\rangle\langle ji|) + A_{ij}\,|ij\rangle\langle ji| + A^*_{ij}\,|ji\rangle\langle ij|\right)\\ &\simeq \left(\textstyle\sum_i A_{ii}\,|ii\rangle\langle ii|\right) \oplus \bigoplus_{i<j}\begin{pmatrix} Q_{ij}|A_{ij}| & A_{ij}\\ A^*_{ij} & Q_{ij}|A_{ij}|\end{pmatrix},\end{aligned} \tag{8.47}$$

from which, using the elementary observation that

$$\left|\begin{pmatrix} Q_{ij}|A_{ij}| & A_{ij}\\ A^*_{ij} & Q_{ij}|A_{ij}|\end{pmatrix}\right| = \begin{pmatrix} |A_{ij}| & Q_{ij}A_{ij}\\ Q_{ij}A^*_{ij} & |A_{ij}|\end{pmatrix}, \tag{8.48}$$

valid because $Q_{ij}\in\{0,1\}$, we obtain that

$$\left|\pi^\Gamma_{A,Q}\right|^\Gamma \propto \sum_{i,j} Q_{ij}A_{ij}\,|ii\rangle\langle jj| + \sum_{i\neq j}|A_{ij}|\,|ij\rangle\langle ij|\,, \tag{8.49}$$

where $\propto$ hides a positive constant. We therefore record the following observation:

Lemma 8. *Let $A \geq 0$ be a $d \times d$ positive semi-definite matrix, and let Q be a $d \times d$ symmetric matrix with $Q_{ij} \in \{0, 1\}$ for all $i, j = 1, \ldots, d$, and $Q_{ii} = 1$ for all $i = 1, \ldots, d$. If $Q \circ A \not\geq 0$, where $\circ$ denotes the Hadamard product, then the punch card state $\pi_{A,Q}$ does not have zero bi-negativity, meaning that*

$$\left|\pi_{A,Q}^{\Gamma}\right|^{\Gamma} \not\geq 0\,. \tag{8.50}$$

In particular, the state

$$\pi_0 := \pi_{A_0,Q_0}\,, \qquad A_0 := \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad Q_0 := \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \not\geq 0 \tag{8.51}$$

does not have zero bi-negativity.

And sure enough, the above punch card state is already enough to exhibit a small but meaningful violation of additivity for E_κ, already at the 2-copy level:

Proposition 9. *For the state π_0 defined by (8.51), it holds that*

$$2E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\pi_0) \leq E_\kappa\left(\pi_0^{\otimes 2}\right) < 2E_\kappa(\pi_0)\,. \tag{8.52}$$

In particular, $E_\kappa\left(\pi_0^{\otimes 2}\right) \approx 1.001$ and $2E_\kappa(\pi_0) \approx 1.029$.

Proof. The first inequality follows by combining (8.42) and (8.44). The numerical calculations of $E_\kappa\left(\pi_0^{\otimes 2}\right)$ and $E_\kappa(\pi_0)$ can be verified with an SDP solver. □

8.4.3 New entanglement monotones: $E_{\chi,p}$ and $E_{\kappa,q}$

Rationale of our construction

Let us investigate in greater depth the fundamental reason why E_κ fails to be additive. To this end, it is useful to look at the dual SDP program [57, 58] for E_κ, which takes the form [27, Eq. (S25)]

$$E_\kappa(\rho) = \log_2 \max\left\{\operatorname{Tr}\rho\,(V - W)^{\Gamma} :\ V, W \geq 0,\ V^{\Gamma} + W^{\Gamma} \leq \mathbb{1}\right\}. \tag{8.53}$$

Since E_κ is sub-additive, to prove additivity we would only need to prove super-additivity. This is most naturally done with the dual program, which involves a maximisation. The proof of $E_\kappa(\rho \otimes \rho') \overset{?}{\geq} E_\kappa(\rho) + E_\kappa(\rho')$ would proceed as follows: first we consider optimal feasible points V, W and V', W' that achieve the maxima in the dual SDPs (8.53) for ρ and ρ'; then we attempt to construct a feasible point V'', W'' for the dual SDP for $\rho \otimes \rho'$; such an ansatz would need to yield a value of the objective function equal to $E_\kappa(\rho) + E_\kappa(\rho')$ (upon taking the logarithm). A quick inspection reveals that this requires that

$$\operatorname{Tr}\rho\,(V'' - W'')^{\Gamma} \overset{?}{=} \left(\operatorname{Tr}\rho\,(V - W)^{\Gamma}\right)\left(\operatorname{Tr}\rho\,(V' - W')^{\Gamma}\right). \tag{8.54}$$

A natural ansatz at this point is the one found by Wang and Wilde [27], namely,

$$V'' = V \otimes V' + W \otimes W'\,, \qquad W'' = V \otimes W' + W \otimes V'\,. \tag{8.55}$$

Note that $V'', W'' \geq 0$; furthermore, this ansatz clearly satisfies (8.54). The problem is that

$$(V'' + W'')^{\Gamma} = (V + W)^{\Gamma} \otimes (V' + W')^{\Gamma} \not\leq \mathbb{1} \tag{8.56}$$

in general. Indeed, while $(V + W)^{\Gamma} \leq \mathbb{1}$ and $(V' + W')^{\Gamma} \leq \mathbb{1}$ holds by hypothesis, from this it does not follow that the tensor product of these operators is also upper bounded by the identity.

The reason is that both of these operators could have large negative eigenvalues, which would become large and positive when multiplied.

The first idea we have is to modify the definition of E_κ so as to make sure that this does not happen. To this end, it suffices to add the further constraint $(V+W)^\Gamma \geq -\mathbb{1}$. We can therefore define

$$\begin{aligned} E_\chi(\rho) &:= \log_2 \max\left\{\operatorname{Tr}\rho\,(V-W)^\Gamma :\ V,W\geq 0,\ -\mathbb{1}\leq V^\Gamma+W^\Gamma\leq\mathbb{1}\right\} \\ &= \log_2 \min\left\{\operatorname{Tr} T :\ -S\leq\rho^\Gamma\leq S,\ -T\leq S^\Gamma\leq T\right\} \\ &= \log_2 \min\left\{\left\|S^\Gamma\right\|_1 :\ -S\leq\rho^\Gamma\leq S\right\}, \end{aligned} \tag{8.57}$$

where in the second line we derived the primal program corresponding to the dual program in the first line, and in the third line we used Lemma 6.

We will see shortly that our approach fixes the additivity problem, in the sense that the new function E_χ we have just constructed is indeed additive, unlike E_κ. Of course, what we are really interested in is the regularisation of E_κ; in general, E_χ can only provide a lower bound to that quantity. Indeed, for all states it holds that $E_\chi \leq E_\kappa$, because taking the optimisation on the third line of (8.57) and restricting to operators S that satisfy $S^\Gamma \geq 0$ yields precisely E_κ, as can be seen by observing that in those cases $\left\|S^\Gamma\right\|_1 = \operatorname{Tr} S$. Assuming the additivity of E_χ, which we will prove shortly, we immediately deduce that

$$E_\kappa \geq E_\kappa^\infty \geq E_\chi\,. \tag{8.58}$$

While this is already interesting in itself, we can do better. In fact, the definition of E_χ in the second line of (8.57) lends itself naturally to some generalisations.

Definition 10 (χ-hierarchy)**.** *Let $\rho=\rho_{AB}$ be any bipartite quantum state. For $p\in\mathbb{N}$, we define the corresponding $\boldsymbol{\chi}$-quantity $E_{\chi,p}$ as*

$$E_{\chi,p}(\rho) := \log_2\chi_p(\rho)\,, \tag{8.59}$$

$$\chi_p(\rho) := \min\left\{\operatorname{Tr} S_p :\ -S_i\leq S_{i-1}^\Gamma\leq S_i,\ i=0,...,p,\ S_{-1}=\rho\right\} \tag{8.60}$$

$$= \min\left\{\operatorname{Tr} S_p :\ -S_0\leq\rho^\Gamma\leq S_0,\ -S_1\leq S_0^\Gamma\leq S_1,\ \ldots\ ,\ -S_p\leq S_{p-1}^\Gamma\leq S_p\right\}. \tag{8.61}$$

We will refer to the sequence of functions $E_{\chi,p}$, $p\in\mathbb{N}$, as the $\boldsymbol{\chi}$-hierarchy.

Remark 11. The minima in (8.60)–(8.61) are always achieved, essentially because of the compactness of the set of positive semi-definite operators with trace bounded by a fixed number.

Remark 12. For $p>0$, due to Lemma 6 we can equivalently write

$$\chi_p(\rho) = \min\left\{\left\|S_{p-1}^\Gamma\right\| :\ -S_0\leq\rho^\Gamma\leq S_0,\ -S_1\leq S_0^\Gamma\leq S_1,\ \ldots\ ,\ -S_{p-1}\leq S_{p-2}^\Gamma\leq S_{p-1}\right\}. \tag{8.62}$$

When $p=0$, clearly we have that

$$E_{\chi,0}(\rho) = \log_2\left\|\rho^\Gamma\right\|_1 = E_N(\rho) \tag{8.63}$$

yields the celebrated logarithmic negativity [37, 38]. For $p=1$, instead,

$$E_{\chi,1}(\rho) = E_\chi(\rho) \tag{8.64}$$

reproduces (8.57).

While this is quite interesting, we can now observe that the above functions $E_{\chi,p}$ are of a somewhat different nature than the E_κ monotone introduced by Wang and Wilde [27], whose definition is reported in (8.43). We can however try to generalise that construction as well. This is done as follows: for $q\geq 1$, one replaces p with q in (8.60), restricts the minimisation there to operators S_{q-1} satisfying $S_{q-1}^\Gamma\geq 0$, and sets $S_q = S_{q-1}^\Gamma$ accordingly; this gives a 'κ-hierarchy' defined as follows.

Definition 13 (κ-hierarchy)**.** *Let $\rho = \rho_{AB}$ be any bipartite quantum state. For an arbitrary positive integer $q \in \mathbb{N}^+$, we define the corresponding* $\boldsymbol{\kappa}$-quantity $E_{\kappa,q}$ *as*

$$E_{\kappa,q}(\rho) := \log_2 \kappa_q(\rho)\,, \tag{8.65}$$

$$\kappa_q(\rho) := \min\left\{\operatorname{Tr} S_{q-1} :\ -S_i \leq S_{i-1}^\Gamma \leq S_i,\ i = 0, ..., q-1,\ S_{-1} = \rho,\ S_{q-1}^\Gamma \geq 0\right\} \tag{8.66}$$

$$= \min\left\{\operatorname{Tr} S_{q-1} :\ -S_0 \leq \rho^\Gamma \leq S_0,\ \ldots\ ,\ -S_{q-1} \leq S_{q-2}^\Gamma \leq S_{q-1},\ S_{q-1}^\Gamma \geq 0\right\}. \tag{8.67}$$

We will refer to the sequence of functions $E_{\chi,q}$, $q \in \mathbb{N}^+$, as the $\boldsymbol{\kappa}$-hierarchy*.*

For reasons analogous to those explained in Remark 11, the minima in (8.66)–(8.67) are always achieved. Furthermore, note that

$$E_{\kappa,1}(\rho) = E_\kappa(\rho) \tag{8.68}$$

reproduces the function in (8.43). We shall now explore the properties of the χ- and κ-hierarchies in detail.

Relation between the χ- and κ-hierarchies

We now begin our investigation of the properties of the χ- and κ-hierarchies. First, we will study the relation between the two. A very simple yet very useful observation that we record as a preliminary lemma is as follows.

Lemma 14 (Increasing trace)**.** *Let the operators $S_0, \ldots, S_p$ be a feasible point for the SDP (8.60) that defines $\chi_p(\rho)$, for some bipartite state $\rho = \rho_{AB}$ and some $p \in \mathbb{N}$. Then $S_0, \ldots, S_p \geq 0$ are positive semi-definite, and moreover*

$$1 \leq \operatorname{Tr} S_0 \leq \operatorname{Tr} S_1 \leq \ldots \leq \operatorname{Tr} S_p\,. \tag{8.69}$$

Similarly, if $S'_0, \ldots, S'_{q-1}$ constitute a feasible point for the SDP (8.66) that defines $\kappa_q(\rho)$ ($q \in \mathbb{N}^+$), then $S'_0, \ldots, S'_{q-1} \geq 0$, and

$$1 \leq \operatorname{Tr} S'_0 \leq \operatorname{Tr} S'_1 \leq \ldots \leq \operatorname{Tr} S'_{q-1}\,. \tag{8.70}$$

Proof. From the inequalities $-S_i \leq S_{i-1}^\Gamma \leq S_i$ ($i = 0, \ldots, p$, with $S_{-1} = \rho$) it follows immediately that $-S_i \leq S_i$, i.e. $S_i \geq 0$. To prove (8.69), it suffices to take the trace of the inequalities $S_0^\Gamma \leq S_1$, $S_1^\Gamma \leq S_2$, ..., and $S_{p-1}^\Gamma \leq S_p$. The proof of $S'_0, \ldots, S'_{q-1} \geq 0$ and of (8.70) is entirely analogous. □

Remark 15. In particular, for all $p \in \mathbb{N}$, $q \in \mathbb{N}^+$, and bipartite states ρ, it holds that $\chi_p(\rho) \geq 1$ and $\kappa_q(\rho) \geq 1$.

We are now ready to prove the following result.

Proposition 16. *For all non-negative integers $p \in \mathbb{N}$, positive integers $q \in \mathbb{N}^+$, and for all bipartite states $\rho = \rho_{AB}$, it holds that*

$$E_N = E_{\chi,0} \leq \ldots \leq E_{\chi,p} \leq E_{\chi,p+1} \leq \ldots \leq E_{\kappa,q+1} \leq E_{\kappa,q} \leq \ldots \leq E_\kappa\,, \tag{8.71}$$

where we forwent the dependence on ρ. In other words, the function $\mathbb{N} \ni p \mapsto E_{\chi,p}(\rho)$ is increasing,[2] *while $\mathbb{N}^+ \ni q \mapsto E_{\kappa,q}(\rho)$ is decreasing; moreover, $E_{\chi,p}(\rho) \leq E_{\kappa,q}(\rho)$ holds for all $p \in \mathbb{N}$ and $q \in \mathbb{N}^+$.*

Proof. We prove the claims one by one.

- $\mathbb{N} \ni p \mapsto E_{\chi,p}$ *is increasing (for fixed ρ).* Let $S_0, \ldots, S_{p+1}$ be optimal feasible points for the SDP that defines χ_{p+1}. Then the operators $S_0, \ldots, S_p$ clearly constitute a feasible point for the SDP that defines χ_p, because all the inequalities that have to be obeyed are also present in the SDP that defines χ_{p+1}. Hence, $\chi_p \leq \operatorname{Tr} S_p \leq \operatorname{Tr} S_{p+1} = \chi_{p+1}$, where for the second inequality we leveraged Lemma 14 (and, more precisely, Eq. (8.69)). Taking the logarithm completes the argument.

[2] Not necessarily *strictly* increasing. The meaning of 'decreasing' should similarly be intended as 'non-increasing'.

- $\mathbb{N}^+ \ni q \mapsto E_{\kappa,q}$ *is decreasing (for fixed* ρ*).* The proof is somewhat similar, but reversed. If $S_0, \ldots, S_{q-1}$ represent the optimal feasible point for the SDP (8.66) that defines κ_q, then we can construct a corresponding feasible point $S_0, \ldots, S_{q-1}, S_q \coloneqq S_{q-1}^\Gamma$ for the SDP that defines κ_{q+1}. Indeed, the inequalities $-S_i \leq S_{i-1}^\Gamma \leq S_i$ are automatically satisfied for $i = 0, \ldots, q-1$ (here, $S_{-1} = \rho$). For $i = q$, we obviously have that $-S_q = -S_{q-1}^\Gamma \leq S_{q-1}^\Gamma = S_q$ due to the fact that $S_{q-1}^\Gamma \geq 0$. Finally, the last constraint in the SDP for κ_{q+1} is $S_q^\Gamma \geq 0$. This is also obeyed, because $S_q^\Gamma = S_{q-1}$, which is positive definite due to Lemma 14. Using (8.37) we thus obtain that $\kappa_{q+1} \leq \operatorname{Tr} S_q = \operatorname{Tr} S_{q-1}^\Gamma = \operatorname{Tr} S_{q-1} = \kappa_q$. Taking the logarithm completes the proof of the claim.

- $E_{\chi,p} \leq E_{\kappa,q}$ *for all* $p \in \mathbb{N}$ *and* $q \in \mathbb{N}^+$. Due to the fact that $\mathbb{N} \ni p \mapsto E_{\chi,p}$ is increasing and $\mathbb{N} \ni q \mapsto E_{\kappa,q}$ is decreasing,[3] the claim is equivalent to the statement that

$$\sup_{p \in \mathbb{N}} E_{\chi,p} = \lim_{p \to \infty} E_{\chi,p} \overset{?}{\leq} \lim_{q \to \infty} E_{\kappa,q} = \inf_{q \in \mathbb{N}^+} E_{\kappa,q} \,. \tag{8.72}$$

  Now, let $S_0, \ldots, S_{q-1}$ be optimal feasible points for the SDP (8.66) that represents κ_q, for some $q \in \mathbb{N}^+$. By direct inspection of (8.60), we see that these are also feasible points for the SDP that represents χ_{q-1}. Therefore,

$$\chi_{q-1} \leq \operatorname{Tr} S_{q-1} = \kappa_q \,. \tag{8.73}$$

  Since this holds for an arbitrary $q \in \mathbb{N}^+$, we can now take the limit $q \to \infty$ and subsequently the logarithm of both sides, thus concluding the proof.

□

To conclude this subsection, we observe that both of our hierarchies collapse into one single quantity in the simple case where the underlying state ρ has zero bi-negativity. This includes, in particular, the case where ρ is pure [55, Lemma 5].

Lemma 17. *If* $\rho = \rho_{AB}$ *has zero bi-negativity, in the sense that* $\left|\rho^\Gamma\right|^\Gamma \geq 0$*, then*

$$E_{\chi,p}(\rho) = E_{\kappa,q}(\rho) = E_N(\rho) = \log_2 \left\|\rho^\Gamma\right\|_1 \tag{8.74}$$

for all $p \in \mathbb{N}$ *and* $q \in \mathbb{N}^+$*. In particular, for all pure states* $\Psi = |\Psi\rangle\langle\Psi|$ *with Schmidt decomposition* $|\Psi\rangle = |\Psi\rangle_{AB} = \sum_i \sqrt{\lambda_i}\, |e_i\rangle_A \otimes |f_i\rangle_B$*, we have that*

$$E_{\chi,p}(\Psi) = E_{\kappa,q}(\Psi) = 2 \log_2 \left(\sum\nolimits_i \sqrt{\lambda_i} \right) \leq \log_2 d \,, \tag{8.75}$$

where $d \coloneqq \min\{|A|, |B|\}$*, for all* $p \in \mathbb{N}$ *and* $q \in \mathbb{N}^+$*, with equality if* $|\Psi\rangle = |\Phi_d\rangle$ *is the maximally entangled state defined by* (8.33)*.*

Proof. Wang and Wilde [27, Proposition 8] have already established that $E_\kappa(\rho) = E_N(\rho)$ if ρ has zero bi-negativity. Due to Proposition 16, this implies a complete collapse of both hierarchies. Pure states are known to have zero bi-negativity [55, Lemma 5], hence $E_{\chi,p}(\Psi) = E_{\kappa,q}(\Psi) = E_N(\Psi)$; the explicit expression in terms of the Schmidt coefficients reported in (8.75) is a consequence of the observation that $\left\|\Psi^\Gamma\right\|_1 = \left(\sum_i \sqrt{\lambda_i}\right)^2$ due to [37, Proposition 8]. In the case of a maximally entangled state of local dimension d, we have that $E_N(\Phi_d) = \log_2 d$ because $\lambda_i = 1/d$ for all $i = 1, \ldots, d$. □

[3] The function $\mathbb{N} \ni q \mapsto E_{\kappa,q}$ is also bounded, due to the inequality $E_{\kappa,q} \leq E_\kappa$ together with the fact that E_κ is finite for all (finite-dimensional) states.

$E_{\chi,p}$ and $E_{\kappa,q}$ as PPT monotones

We will now prove that our functions χ_p, κ_q, as well as their logarithmic versions $E_{\chi,p}$ and $E_{\kappa,q}$, represent novel entanglement monotones under PPT operations.

Proposition 18 (Strong monotonicity)**.** *For two bipartite quantum systems AB and $A'B'$, let $\left(\mathcal{N}_x\right)_x$ be a collection of PPT maps with input system AB and output system $A'B'$ such that $\sum_x \mathcal{N}_x$ is trace preserving. Then for any bipartite quantum state ρ on AB it holds that*

$$F(\rho) \geq \sum_{x:\, \mathrm{Tr}[\mathcal{N}_x(\rho)]>0} \mathrm{Tr}\left[\mathcal{N}_x(\rho)\right]\, F\left(\frac{\mathcal{N}_x(\rho)}{\mathrm{Tr}\left[\mathcal{N}_x(\rho)\right]}\right), \qquad F = \chi_p,\, \kappa_q,\, E_{\chi,p},\, E_{\kappa,q}\,, \qquad \forall\, p \in \mathbb{N},\ \forall\, q \in \mathbb{N}^+\,. \tag{8.76}$$

In particular, for any PPT channel $\mathcal{N}$ it holds that

$$\begin{aligned} E_{\chi,p}(\rho) &\geq E_{\chi,p}(\mathcal{N}(\rho))\,, \\ E_{\kappa,q}(\rho) &\geq E_{\kappa,q}(\mathcal{N}(\rho))\,. \end{aligned} \tag{8.77}$$

Proof. The cases $F = E_{\chi,p}$ and $F = E_{\kappa,q}$ of (8.76) follow by taking the logarithm of the same inequality written for the case $F = \chi_p$ and $F = \kappa_q$, while exploiting the concavity of the function $x \mapsto \log_2(x)$ together with the observation that $\left(\mathrm{Tr}\left[\mathcal{N}_x(\rho)\right]\right)_x$ is a probability distribution as $\sum_x \mathcal{N}_x$ is trace preserving.

We thus focus on the cases $F = \chi_p, \kappa_q$, starting from the former. Let $S_{-1}, S_0, \ldots, S_p$ be optimisers of the minimisation problem (8.60) that defines $\chi_p(\rho)$. In particular,

$$S_{-1} = \rho, \qquad -S_i \leq S_{i-1}^{\Gamma} \leq S_i \qquad \forall\, i = 0, 1, \ldots, p\,. \tag{8.78}$$

Since $\mathcal{N}_x$ is PPT, the map $\Gamma \circ \mathcal{N}_x \circ \Gamma$ is (completely) positive. Eq. (8.78) then implies that

$$-\left[\mathcal{N}_x(S_i^{\Gamma})\right]^{\Gamma} \leq \left[\mathcal{N}_x(S_{i-1})\right]^{\Gamma} \leq \left[\mathcal{N}_x(S_i^{\Gamma})\right]^{\Gamma} \quad \forall\, i = 0, 1, \ldots, p\,. \tag{8.79}$$

Moreover, Eq. (8.78) and the fact that $\mathcal{N}_x$ is completely positive imply that

$$-\mathcal{N}_x(S_i) \leq \mathcal{N}_x(S_{i-1}^{\Gamma}) \leq \mathcal{N}_x(S_i) \qquad \forall\, i = 0, 1, \ldots, p-1\,. \tag{8.80}$$

Consequently, for any $i = -1, 0, \ldots, p$ we can introduce the operators

$$S_i^{(x)} := \begin{cases} \left[\mathcal{N}_x(S_i^{\Gamma})\right]^{\Gamma} & \text{if } i \text{ is even}\,, \\ \mathcal{N}_x(S_i) & \text{if } i \text{ is odd} \end{cases} \tag{8.81}$$

(here, -1 is considered odd), in terms of which we can rewrite (8.79) as

$$-S_i^{(x)} \leq \left(S_{i-1}^{(x)}\right)^{\Gamma} \leq S_i^{(x)} \qquad (i \text{ even}), \tag{8.82}$$

and (8.80) as

$$-S_i^{(x)} \leq \left(S_{i-1}^{(x)}\right)^{\Gamma} \leq S_i^{(x)} \qquad (i \text{ odd}). \tag{8.83}$$

It follows that for any x such that $\mathrm{Tr}[\mathcal{N}_x(\rho)] > 0$ the operators $S_i^{(x)} \big/ \mathrm{Tr}\,\mathcal{N}_x(\rho)$, $i = -1, 0, \ldots, p$ are valid ansatzes in the minimisation problem (8.60) that defines $\chi_p\left(\frac{\mathcal{N}_x(\rho)}{\mathrm{Tr}[\mathcal{N}_x(\rho)]}\right)$. Therefore,

$$\begin{aligned} \chi_p(\rho) &= \mathrm{Tr}[S_p] \\ &\overset{\text{(i)}}{=} \sum_x \mathrm{Tr}\left[S_p^{(x)}\right] \\ &\overset{\text{(ii)}}{\geq} \sum_{x:\, \mathrm{Tr}[\mathcal{N}_x(\rho)]>0} \mathrm{Tr}\left[\mathcal{N}_x(\rho)\right]\, \mathrm{Tr}\left[\frac{S_p^{(x)}}{\mathrm{Tr}\left[\mathcal{N}_x(\rho)\right]}\right] \\ &\overset{\text{(iii)}}{\geq} \sum_{x:\, \mathrm{Tr}[\mathcal{N}_x(\rho)]>0} \mathrm{Tr}\left[\mathcal{N}_x(\rho)\right]\, \chi_p\left(\frac{\mathcal{N}_x(\rho)}{\mathrm{Tr}\left[\mathcal{N}_x(\rho)\right]}\right). \end{aligned} \tag{8.84}$$

Here, in (i) we used the fact that the maps $\sum_x \mathscr{N}_x$ and $\sum_x \Gamma \circ \mathscr{N}_x \circ \Gamma$ are trace preserving; in (ii) we observed that $S_p^{(x)} \geq 0$, which follows from (8.79) and (8.80); finally, in (iii) we leveraged the fact that the operators $S_i^{(x)} \big/ \operatorname{Tr} \mathscr{N}_x(\rho)$, $i = -1, 0, \ldots, p$ are valid ansatzes in the minimisation problem (8.60) that defines $\chi_p\Big(\frac{\mathscr{N}_x(\rho)}{\operatorname{Tr}[\mathscr{N}_x(\rho)]}\Big)$, as argued above. This proves (8.76) for $F = \chi_p$.

The proof of the strong monotonicity property (8.76) for the case where $F = \kappa_q$ is completely analogous. The only additional observation we need is that $\Big(S_{q-1}^{(x)}\Big)^\Gamma \geq 0$ if $S_{q-1}^\Gamma \geq 0$. Indeed, if q is odd then $\Big(S_{q-1}^{(x)}\Big)^\Gamma = \mathscr{N}_x\Big(S_{q-1}^\Gamma\Big)$, which is positive because $\mathscr{N}_x$ is a positive map; if q is even then $\Big(S_{q-1}^{(x)}\Big)^\Gamma = \mathscr{N}_x(S_{q-1})^\Gamma = (\Gamma \circ \mathscr{N}_x \circ \Gamma)\Big(S_{q-1}^\Gamma\Big)$, which is positive definite because $\Gamma \circ \mathscr{N}_x \circ \Gamma$ is a positive map. This concludes the proof. □

As an immediate consequence of Lemma 17 and Proposition 18, we deduce the following.

Corollary 19. *Let $\rho = \rho_{AB}$ be an arbitrary quantum state of a bipartite system AB with minimal local dimension $d \coloneqq \min\{|A|, |B|\}$. Then for all $p \in \mathbb{N}$ and $q \in \mathbb{N}^+$ it holds that*

$$\chi_p(\rho),\, \kappa_q(\rho) \leq d\,, \qquad E_{\chi,p}(\rho),\, E_{\kappa,q}(\rho) \leq \log_2 d\,. \tag{8.85}$$

Proof. An application of Nielsen's theorem [70] guarantees that there exists an LOCC transformation Λ such that $\Lambda(\Phi_d) = \rho_{AB}$, where Φ_d is the maximally entangled state with local dimension d.[4] Since LOCC is a subset of PPT, the result then follows by combining Lemma 17 and Proposition 18. Incidentally, an alternative, direct way of constructing an LOCC with the above property is to let the party with the maximum local dimension prepare ρ_{AB} locally and then apply the quantum teleportation protocol [71] to teleport the smaller system to the other party. □

Dual expressions

In what follows, we derive the SDPs dual to those in (8.60) and (8.66). We refer the reader to [56–58] for an introduction to the general theory of semi-definite programs. We start with the χ-quantities, whose dual expressions will play an important role in establishing their additivity.

Proposition 20 (Dual SDP for the χ-quantity)**.** *For an arbitrary bipartite state $\rho = \rho_{AB}$ and some $p \in \mathbb{N}$, the associated χ-quantity $\chi_p(\rho)$ defined by (8.60) can be equivalently expressed as*

$$\begin{aligned} \chi_p(\rho) \;=\; \text{max.}\quad & \operatorname{Tr}\Big[\rho\Big(V_0^\Gamma - W_0^\Gamma\Big)\Big] \\ & V_0, \ldots, V_p, W_0, \ldots, W_p \geq 0\,, \\ \text{s.t.}\quad & V_i + W_i = V_{i+1}^\Gamma - W_{i+1}^\Gamma \quad \forall\, i = 0, \ldots, p-1\,, \quad V_p + W_p = \mathbb{1}\,, \end{aligned} \tag{8.86}$$

or else, if $p \geq 1$, as

$$\begin{aligned} \chi_p(\rho) \;=\; \text{max.}\quad & \operatorname{Tr}\Big[\rho\Big(V_0^\Gamma - W_0^\Gamma\Big)\Big] \\ & V_0, \ldots, V_{p-1}, W_0, \ldots, W_{p-1} \geq 0\,, \\ \text{s.t.}\quad & V_i + W_i = V_{i+1}^\Gamma - W_{i+1}^\Gamma \quad \forall\, i = 0, \ldots, p-2\,, \quad -\mathbb{1} \leq \Big(V_{p-1} + W_{p-1}\Big)^\Gamma \leq \mathbb{1}\,. \end{aligned} \tag{8.87}$$

Remark 21. If $p = 0$, then the only constraints in (8.86) are $V_0, W_0 \geq 0$ and $V_0 + W_0 = \mathbb{1}$. Similarly, if $p = 1$ then the only constraints in (8.87) are $V_0, W_0 \geq 0$ and $-\mathbb{1} \leq \Big(V_0 + W_0\Big)^\Gamma \leq \mathbb{1}$. Note that for $p = 1$ the SDP in (8.87) reproduces (up to the logarithm) the expression on the first line of (8.57), as it should. In general, (8.87) involves one fewer pair of variables compared to (8.86), as it is deduced by eliminating V_p and W_p from (8.86) (see the proof below).

[4] Indeed, writing a spectral decomposition $\rho = \sum_i p_i \Psi_i$ of ρ, a direct consequence of Nielsen's theorem is that for all i there exists an LOCC Λ_i that satisfies $\Lambda_i(\Phi_d) = \Psi_i$. Setting $\Lambda(\cdot) \coloneqq \sum_i p_i \Lambda_i(\cdot)$ proves the claim.

Proof of Proposition 20. The $p+1$ (effectively free) variables $S_0, \dots, S_p$ in (8.60) obey $2(p+1)$ semi-definite constraints, namely $S_i \pm S_{i-1}^\Gamma \geq 0$ for $i = 0, \dots, p$, with the convention that $S_{-1} = \rho$. We therefore introduce $p+1$ positive semi-definite dual variables $V_0, \dots, V_p \geq 0$ for the inequalities with the minus and another $p+1$, namely $W_0, \dots, W_p \geq 0$, for the inequalities with the plus. The resulting Lagrangian is

$$\begin{aligned}\mathcal{L} &= \operatorname{Tr} S_p - \sum_{i=0}^{p} \operatorname{Tr} V_i \left(S_i - S_{i-1}^\Gamma\right) - \sum_{i=0}^{p} \operatorname{Tr} W_i \left(S_i + S_{i-1}^\Gamma\right) \\ &= \operatorname{Tr} \rho \left(V_0^\Gamma - W_0^\Gamma\right) - \sum_{i=0}^{p-1} \operatorname{Tr} S_i \left(V_i + W_i - V_{i+1}^\Gamma + W_{i+1}^\Gamma\right) + \operatorname{Tr} S_p \left(\mathbb{1} - V_p - W_p\right),\end{aligned} \tag{8.88}$$

where in the second line we employed (8.36) repeatedly. Minimising the above expression over $S_0, \dots, S_p$ yields $-\infty$ unless $V_i + W_i - V_{i+1}^\Gamma + W_{i+1}^\Gamma = 0$ for all $i = 0, \dots, p-1$, and moreover $V_p + W_p = \mathbb{1}$. This proves that the SDP dual to that in (8.60) is the one given by (8.86).

To see that strong duality holds (and the dual is achieved), it suffices to check that Slater's condition is obeyed, and that both the primal and the dual program are in fact strictly feasible. This simply means that all inequalities can be satisfied strictly. In the primal program (8.60), one can take $S_i = (i+1)\mathbb{1}$ $(i = 0, \dots, p)$; in the dual program (8.86), instead, it suffices to set $V_i = 2 \cdot 3^{-(p-i+1)}\mathbb{1}$, $W_i = 3^{-(p-i+1)}\mathbb{1}$ $(i = 0, \dots, p)$.

Finally, it remains to see that (8.87) is equivalent to (8.86) if $p \geq 1$. This can be done as follows. In (8.86), the only constraints on $V_p, W_p \geq 0$ are that $V_p - W_p = \left(V_{p-1} - W_{p-1}\right)^\Gamma$ and $V_p + W_p = \mathbb{1}$. It is possible to satisfy these constraints if and only if $-\mathbb{1} \leq \left(V_{p-1} - W_{p-1}\right)^\Gamma \leq \mathbb{1}$. Indeed, this condition is necessary, because on the one hand $\left(V_{p-1} - W_{p-1}\right)^\Gamma \leq V_p \leq V_p + W_p = \mathbb{1}$, and on the other $\left(V_{p-1} - W_{p-1}\right)^\Gamma \geq -W_p \geq -(V_p + W_p) = -\mathbb{1}$. It is also sufficient, since if it is obeyed we can set $V_p = \frac{1}{2}\left(\mathbb{1} + \left(V_{p-1} - W_{p-1}\right)^\Gamma\right)$ and $W_p = \frac{1}{2}\left(\mathbb{1} - \left(V_{p-1} - W_{p-1}\right)^\Gamma\right)$. □

Although it is not strictly needed in what follows, for completeness we derive also the dual program for the κ-quantities.

Proposition 22 (Dual SDP for the κ-quantity)**.** *For an arbitrary bipartite state $\rho = \rho_{AB}$ and some $q \in \mathbb{N}^+$, the κ-quantity $\kappa_q(\rho)$ defined by (8.66) can be equivalently expressed as*

$$\begin{aligned}\kappa_q(\rho) \;=\; \max.\quad & \operatorname{Tr}\left[\rho \left(V_0^\Gamma - W_0^\Gamma\right)\right] \\ & V_0, \dots, V_{q-1}, W_0, \dots, W_{q-1} \geq 0\,, \\ \text{s.t.}\quad & V_i + W_i = V_{i+1}^\Gamma - W_{i+1}^\Gamma \quad \forall\; i = 0, \dots, q-2\,, \quad \left(V_{q-1} + W_{q-1}\right)^\Gamma \leq \mathbb{1}\,.\end{aligned} \tag{8.89}$$

Remark 23. If $q = 1$, then the only constraints in (8.89) are $V_0, W_0 \geq 0$ and $\left(V_0 + W_0\right)^\Gamma \leq \mathbb{1}$. Note that for $q = 1$ the SDP in (8.89) reproduces (up to the logarithm) the expression in (8.53).

Proof of Proposition 22. The argument is very similar to that employed to prove Proposition 20. This time around we have $2q+1$ constraints, namely $S_i \pm S_{i-1}^\Gamma \geq 0$ for $i = 0, \dots, q-1$ (with the convention that $S_{-1} = \rho$), as well as $S_{q-1}^\Gamma \geq 0$. We thus introduce q positive semi-definite dual variables $V_0, \dots, V_{q-1} \geq 0$ for the first inequalities with the minus, another q, namely $W_0, \dots, W_{q-1} \geq 0$, for the first inequalities with the plus, and finally one more variable $Z \geq 0$ for the last inequality. The resulting Lagrangian is

$$\begin{aligned}\mathcal{L} &= \operatorname{Tr} S_{q-1} - \sum_{i=0}^{q-1} \operatorname{Tr} V_i \left(S_i - S_{i-1}^\Gamma\right) - \sum_{i=0}^{q-1} \operatorname{Tr} W_i \left(S_i + S_{i-1}^\Gamma\right) - \operatorname{Tr} Z S_{q-1}^\Gamma \\ &= \operatorname{Tr} \rho \left(V_0^\Gamma - W_0^\Gamma\right) - \sum_{i=0}^{q-2} \operatorname{Tr} S_i \left(V_i + W_i - V_{i+1}^\Gamma + W_{i+1}^\Gamma\right) + \operatorname{Tr} S_{q-1} \left(\mathbb{1} - Z^\Gamma - V_{q-1} - W_{q-1}\right).\end{aligned} \tag{8.90}$$

Minimising the above expression over $S_0, \ldots, S_{q-1}$ gives $-\infty$ unless $V_i + W_i - V_{i+1}^\Gamma + W_{i+1}^\Gamma = 0$ for all $i = 0, \ldots, q-2$, and moreover $V_{q-1} - W_{q-1} = \mathbb{1} - Z^\Gamma$. The maximisation over Z can be eliminated by observing that $V_{q-1} - W_{q-1} = \mathbb{1} - Z^\Gamma$ for some $Z \geq 0$ if and only if $\left(V_{q-1} - W_{q-1}\right)^\Gamma \leq \mathbb{1}$. This proves that the SDP in (8.89) is indeed the dual of that in (8.66). To see that they yield the same value, i.e. that the duality gap is zero, it suffices to check Slater's condition. This can be done with a reasoning entirely analogous to that presented in the proof of Proposition 20. □

Additivity properties

We will now look at the additivity properties of our new monotones $E_{\chi,p}$ and $E_{\kappa,q}$ constructed in Definitions 10 and 13, respectively. We will prove that the former are fully additive (Proposition 25) while the latter are only sub-additive (Proposition 26). Before we begin, we need a little preliminary result that is well known in the quantum information literature (see e.g. [72, Lemma 12.35]). We include a self-contained proof for completeness.

Lemma 24. *For $i = 1, 2$, let A_i and B_i be linear operators on the Hilbert space $\mathcal{H}_i$. Assume that*

$$-A_i \leq B_i \leq A_i \qquad \forall\ i = 1, 2\,. \tag{8.91}$$

Then

$$-A_1 \otimes A_2 \leq B_1 \otimes B_2 \leq A_1 \otimes A_2\,. \tag{8.92}$$

In particular, if $-A \leq B \leq A$ then for all $n \in \mathbb{N}$ it holds that $-A^{\otimes n} \leq B^{\otimes n} \leq A^{\otimes n}$.

Proof. By adding the two operator inequalities

$$\begin{aligned} 0 &\leq (A_1 + B_1) \otimes (A_2 + B_2) = A_1 \otimes A_2 + B_1 \otimes B_2 + A_1 \otimes B_2 + B_1 \otimes A_2\,, \\ 0 &\leq (A_1 - B_1) \otimes (A_2 - B_2) = A_1 \otimes A_2 + B_1 \otimes B_2 - A_1 \otimes B_2 - B_1 \otimes A_2\,, \end{aligned} \tag{8.93}$$

we infer that $B_1 \otimes B_2 \geq -A_1 \otimes A_2$. Analogously, by adding the two operator inequalities

$$\begin{aligned} 0 &\leq (A_1 - B_1) \otimes (A_2 + B_2) = A_1 \otimes A_2 - B_1 \otimes B_2 + A_1 \otimes B_2 - B_1 \otimes A_2\,, \\ 0 &\leq (A_1 + B_1) \otimes (A_2 - B_2) = A_1 \otimes A_2 - B_1 \otimes B_2 - A_1 \otimes B_2 + B_1 \otimes A_2\,, \end{aligned} \tag{8.94}$$

one obtains that $B_1 \otimes B_2 \leq A_1 \otimes A_2$. The last claim is applying (8.92) iteratively $n-1$ times, indexed $j = 1, \ldots, n-1$, with $A_1 = A^{\otimes j}$, $B_1 = B^{\otimes j}$, $A_2 = A$, and $B_2 = B$. □

We are now ready to state and prove the additivity of $E_{\chi,p}$.

Proposition 25 (Additivity of $E_{\chi,p}$)**.** *For all pairs of bipartite states ρ_{AB}, $\omega_{A'B'}$ and for all $p \in \mathbb{N}$, we have that*

$$\chi_p\left(\rho_{AB} \otimes \omega_{A'B'}\right) = \chi_p(\rho_{AB})\, \chi_p(\omega_{A'B'})\,, \tag{8.95}$$

$$E_{\chi,p}\left(\rho_{AB} \otimes \omega_{A'B'}\right) = E_{\chi,p}\left(\rho_{AB}\right) + E_{\chi,p}\left(\omega_{A'B'}\right). \tag{8.96}$$

Proof. Clearly, Eq. (8.96) follows from (8.95) upon taking the logarithm. We therefore focus on (8.95). To prove it, we will show that χ_p is both sub-multiplicative and super-multiplicative. Let us start by verifying the former property, i.e. the inequality

$$\chi_p(\rho \otimes \omega) \overset{?}{\leq} \chi_p(\rho)\, \chi_p(\omega)\,, \tag{8.97}$$

where we omitted the system labels for simplicity. Let $S_0, \ldots, S_p$ be optimisers of the minimisation problem (8.60) that defines $\chi_p(\rho)$. Analogously, let $S'_0, \ldots, S'_p$ be optimisers of the minimisation

problem that defines $\chi_p(\omega)$. By exploiting Lemma 24, one can easily check that $S_0 \otimes S'_0, \dots, S_p \otimes S'_p$ are valid ansatzes in the minimisation problem that defines $\chi_p(\rho \otimes \omega)$. Consequently,

$$\chi_p(\rho \otimes \omega) \leq \operatorname{Tr}\left[S_p \otimes S'_p\right] = \chi_p(\rho)\, \chi_p(\omega)\,. \tag{8.98}$$

Now, let us prove the super-multiplicativity of χ_p. We thus need to verify that

$$\chi_p(\rho \otimes \omega) \overset{?}{\geq} \chi_p(\rho)\, \chi_p(\omega)\,. \tag{8.99}$$

To this end, we need to employ the dual SDP for χ_p that we derived in Proposition 20. Let $V_0, \dots, V_p, W_0, \dots, W_p \geq 0$ be optimisers of the SDP in (8.86) for $\chi_p(\rho)$. Analogously, let $V'_0, \dots, V'_p, W'_0, \dots, W'_p \geq 0$ be optimisers of the SDP as in (8.86) but for the state ω. For any $i = 0, \dots, p$, define

$$\begin{aligned} V''_i &:= V_i \otimes V'_i + W_i \otimes W'_i\,, \\ W''_i &:= V_i \otimes W'_i + W_i \otimes V'_i\,. \end{aligned} \tag{8.100}$$

Note that

$$\begin{aligned} V''_i + W''_i &= (V_i + W_i) \otimes (V'_i + W'_i)\,, \\ V''_i - W''_i &= (V_i - W_i) \otimes (V'_i - W'_i)\,, \end{aligned} \tag{8.101}$$

one can easily check that $V''_0, \dots, V''_p, W''_0, \dots, W''_p$ are valid ansatzes for the optimisation as in (8.86) but for the state $\rho \otimes \omega$. Indeed, for $i = 0, \dots, p$ we have that $V''_i, W''_i \geq 0$; moreover, if $i \leq p-1$ it also holds that

$$\begin{aligned} V''_i + W''_i &= (V_i + W_i) \otimes (V'_i + W'_i) \\ &= (V_{i+1} - W_{i+1})^{\Gamma} \otimes \left(V'_{i+1} - W'_{i+1}\right)^{\Gamma} \\ &= \left((V_{i+1} - W_{i+1}) \otimes \left(V'_{i+1} - W'_{i+1}\right)\right)^{\Gamma} \\ &= \left(V''_{i+1} - W''_{i+1}\right)^{\Gamma}\,; \end{aligned} \tag{8.102}$$

finally, for $i = p$ we see that

$$V''_p + W''_p = (V_p + W_p) \otimes \left(V'_p + W'_p\right) = \mathbb{1} \otimes \mathbb{1} = \mathbb{1}\,. \tag{8.103}$$

Consequently, it holds that

$$\begin{aligned} \chi_p(\rho \otimes \omega) &\geq \operatorname{Tr}\left[(\rho \otimes \omega)\, (V''_0 - W''_0)^{\Gamma}\right] \\ &= \operatorname{Tr}\left[\rho \left(V_0 - W_0\right)^{\Gamma}\right] \operatorname{Tr}\left[\omega \left(V'_0 - W'_0\right)^{\Gamma}\right] \\ &= \chi_p(\rho)\, \chi_p(\omega)\,. \end{aligned} \tag{8.104}$$

This concludes the proof. □

Unlike the χ-quantities, the corresponding κ-quantities are only sub-additive but in general not fully additive (as we have seen in Section 8.4.2). Although we will not make use of this property in what follows, we state and prove it for completeness.

Proposition 26 (Sub-additivity of $E_{\kappa,q}$)**.** *For all pairs of bipartite states ρ_{AB}, $\omega_{A'B'}$ and for all $q \in \mathbb{N}^+$, we have that*

$$\kappa_q\left(\rho_{AB} \otimes \omega_{A'B'}\right) \leq \kappa_q(\rho_{AB})\, \kappa_q(\omega_{A'B'})\,, \tag{8.105}$$

$$E_{\kappa,q}\left(\rho_{AB} \otimes \omega_{A'B'}\right) \leq E_{\kappa,q}\left(\rho_{AB}\right) + E_{\kappa,q}\left(\omega_{A'B'}\right). \tag{8.106}$$

Proof. The argument is totally analogous to that in the first part of the proof of the above Proposition 25. Let $S_0, \dots, S_{q-1}$ and $S'_0, \dots, S'_{q-1}$, respectively, be optimisers of the SDPs that define $\kappa_q(\rho)$ and $\kappa_q(\omega)$ as in (8.66). Due once more to Lemma 24, one sees that $S_0 \otimes S'_0, \dots, S_{q-1} \otimes S'_{q-1}$ constitute a feasible point for the SDP problem that defines $\kappa_q(\rho \otimes \omega)$. Consequently,

$$\kappa_q(\rho \otimes \omega) \leq \operatorname{Tr}\left[S_{q-1} \otimes S'_{q-1}\right] = \kappa_q(\rho)\, \kappa_q(\omega)\,. \tag{8.107}$$

This concludes the proof. □

Continuity

To close off this section, we explore the continuity properties of our new monotones. We need a couple of simple lemmas first. To simplify the exposition, it is useful to note that the functions χ_p and κ_q ($p \in \mathbb{N}$, $q \in \mathbb{N}^+$) given by (8.60) and (8.66) are perfectly well defined and non-negative not only for states, but also for arbitrary Hermitian operators. In fact, looking at the SDPs in (8.60) and (8.66), one sees that $\chi_p(X) \geq 0$ and $\kappa_q(X) \geq 0$ must hold for all Hermitian X, because in both cases any feasible point must satisfy $S_i \geq 0$ for all i.

It is easy to verify that

$$F(X+Y) \leq F(X) + F(Y)\,, \qquad F = \chi_p,\, \kappa_q, \quad p \in \mathbb{N},\ q \in \mathbb{N}^+, \tag{8.108}$$

and moreover

$$F(\lambda X) = |\lambda|\, F(X)\,, \qquad F = \chi_p,\, \kappa_q, \quad p \in \mathbb{N},\ q \in \mathbb{N}^+, \quad \lambda \in \mathbb{R}\,. \tag{8.109}$$

Here, the absolute value in (8.109) is a consequence of the fact that X appears with both signs in the definition of both $\chi_p(X)$ and $\kappa_q(X)$. We can now extend Lemma 17 to the non-positive case.

Lemma 27. *For all Hermitian operators $X = X_{AB}$ on a bipartite quantum system of local dimension $d \coloneqq \min\{|A|, |B|\}$, it holds that*

$$F(X) \leq d\, \|X\|_1\,, \qquad F = \chi_p,\, \kappa_q, \quad p \in \mathbb{N},\ q \in \mathbb{N}^+\,. \tag{8.110}$$

Proof. Due to the properties of F, for all Hermitian operators X we have that

$$F(X) \overset{\text{(i)}}{=} F\left(\textstyle\sum_i x_i \Psi_i\right) \overset{\text{(ii)}}{\leq} \displaystyle\sum_i F(x_i \Psi_i) \overset{\text{(iii)}}{=} \sum_i |x_i|\, F(\Psi_i) \overset{\text{(iv)}}{\leq} d \sum_i |x_i| = d\, \|X\|_1\,. \tag{8.111}$$

Here, in (i) we introduced a spectral decomposition $X = \sum_i x_i \Psi_i$ of X, (ii) follows from (8.108), (iii) descends from (8.109), and finally in (iv) we leveraged Lemma 17, and in particular (8.75). □

We are now ready to state the continuity properties of our monotones.

Proposition 28 (Continuity)**.** *Let ρ, ρ' be two quantum states on a bipartite system AB with minimal local dimension $d \coloneqq \min\{|A|, |B|\}$. Setting $\varepsilon \coloneqq \frac{1}{2}\, \|\rho - \rho'\|_1$, for all $p \in \mathbb{N}$ and $q \in \mathbb{N}^+$ we have that*

$$|F(\rho) - F(\rho')| \leq 2d\varepsilon\,, \qquad F = \chi_p,\, \kappa_q\,. \tag{8.112}$$

Consequently,

$$|f(\rho) - f(\rho')| \leq \log_2(1 + 2d\varepsilon) \leq 2\,(\log_2 e)\, d\varepsilon, \qquad f = E_{\chi,p},\, E_{\kappa,q}\,. \tag{8.113}$$

Proof. Let $X \coloneqq \rho' - \rho$. Since $\|X\|_1 = 2\varepsilon$, we can write

$$F(\rho') = F(\rho + X) \leq F(\rho) + F(X) \leq F(\rho) + 2d\varepsilon\,, \tag{8.114}$$

where we have used (8.108) and Lemma 27. Applying this reasoning with ρ and ρ' exchanged yields also the reverse inequality and thus completes the proof of (8.112). As for (8.113), it suffices to note that

$$F(\rho') \leq F(\rho) + 2d\varepsilon \leq F(\rho)\,(1 + 2d\varepsilon)\,, \tag{8.115}$$

where in the last line we used that $F(\rho) \geq 1$ by Remark 15. The claim follows by taking logarithms in (8.115), and subsequently exchanging the roles of ρ and ρ'.

□

Remark 29. The type of continuity stated in Proposition 28 is substantially weaker than *asymptotic continuity*, a key property of entanglement monotones [73, 74]. The main point is that the Lipschitz constant appearing on the right hand side of (8.112)–(8.113) is proportional to d instead of $\log_2 d$, as would be needed to have asymptotic continuity. This is not surprising, because the logarithmic negativity itself is known to be *not* asymptotically continuous [37], and our monotones can be thought of generalisations of the negativity. It is worth remarking, however, that even the weaker form of continuity established in Proposition 28 can be extremely useful. The same property for the negativity, for example, which can be re-derived by setting $F = \chi_0$ in (8.112), is key to proving that it upper bounds the distillable entanglement [37, Section IV]. This observation also underpins the whole approach of [75].

8.4.4 Main results

Enter the regularisation

In the previous section we have investigated two families of new PPT monotones, the χ- and the κ-hierarchies. In Proposition 16 we have established a precise hierarchical relation between the two hierarchies, with the increasing functions $E_{\chi,p}(\rho)$ ($p \in \mathbb{N}$) lying below the decreasing functions $E_{\kappa,q}(\rho)$ ($q \in \mathbb{N}^+$). However, our ultimate goal is to understand the zero-error PPT entanglement cost of an arbitrary state ρ, which, by the results of Wang and Wilde (Eq. (8.42)), coincides with the regularised E_κ, in formula $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_\kappa^\infty(\rho)$. The purpose of this subsection, therefore, is to understand how E_κ^∞ fits in the hierarchy delineated by Proposition 16. A first insight can be deduced by leveraging the results we have obtained in the previous section.

Lemma 30. *Let $\rho = \rho_{AB}$ be an arbitrary bipartite state. Then for all $p \in \mathbb{N}$ it holds that*

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_\kappa^\infty(\rho) \geq E_{\chi,p}(\rho)\,. \tag{8.116}$$

In particular,

$$E_\kappa^\infty(\rho) \geq \lim_{p\to\infty} E_{\chi,p}(\rho) = \sup_{p\in\mathbb{N}} E_{\chi,p}(\rho)\,. \tag{8.117}$$

Proof. For all positive integers $n \in \mathbb{N}^+$ we can write

$$\frac{1}{n} E_\kappa\Big(\rho^{\otimes n}\Big) \overset{\text{(i)}}{\geq} \frac{1}{n} E_{\chi,p}\Big(\rho^{\otimes n}\Big) \overset{\text{(ii)}}{=} E_{\chi,p}(\rho)\,, \tag{8.118}$$

where (i) holds due to Proposition 16, and in (ii) we exploited the additivity of $E_{\chi,p}$ (Proposition 25). Taking the limit $n \to \infty$ and leveraging (8.42) concludes the proof of (8.116). Upon taking the limit $p \to \infty$, we obtain also (8.117). □

We have thus established that the function E_κ^∞ lies above the entire χ-hierarchy. We shall now investigate its relationship with the κ-hierarchy. The main result of this subsection is the following.

Theorem 31. *Let $q \in \mathbb{N}^+$ be a positive integer. For any bipartite state $\rho = \rho_{AB}$, it holds that*

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_\kappa^\infty(\rho) \leq E_{\kappa,q}(\rho)\,. \tag{8.119}$$

In particular, for every ρ we have that

$$E_N = E_{\chi,0} \leq E_{\chi,1} \leq \ldots \leq \lim_{p\to\infty} E_{\chi,p} \leq E_{c,\,\mathrm{PPT}}^{\mathrm{exact}} = E_\kappa^\infty \leq \lim_{q\to\infty} E_{\kappa,q} \leq \ldots \leq E_{\kappa,2} \leq E_{\kappa,1} = E_\kappa\,. \tag{8.120}$$

Proof. Clearly, it suffices to prove (8.119), as (8.120) would then follow by combining that with the results of Lemma 30 and Proposition 16. In light of these considerations, we thus set out to prove (8.119).

Let $S_{-1}, S_0, \ldots, S_{q-1}$ be optimisers for the SDP in (8.66) for $\kappa_q(\rho)$. Therefore,

$$-S_i \leq S_{i-1}^\Gamma \leq S_i \qquad \forall\, i = 0, \ldots, q-1\,, \qquad S_{-1} = \rho\,, \qquad S_{q-1}^\Gamma \geq 0\,. \tag{8.121}$$

For an arbitrary positive integer $n \in \mathbb{N}^+$, Lemma 24 implies that

$$-S_i^{\otimes n} \leq \left(S_{i-1}^{\otimes n}\right)^\Gamma \leq S_i^{\otimes n} \qquad \forall\, i = 0, \ldots, q-1\,, \qquad \left(S_{q-1}^\Gamma\right)^{\otimes n} \geq 0 \tag{8.122}$$

Let us define

$$\tilde{S}_n := \sum_{i=0}^{q-1} \left(S_i^{\otimes n} + \left(S_i^\Gamma\right)^{\otimes n}\right). \tag{8.123}$$

We will now show that $\tilde{S}_n$ is in fact a feasible point of the minimisation problem that defines $\kappa_1\left(\rho^{\otimes n}\right) = 2^{E_\kappa(\rho^{\otimes n})}$. To this end, we need to verify that

$$-\tilde{S}_n \overset{?}{\leq} \left(\rho^{\otimes n}\right)^\Gamma \overset{?}{\leq} \tilde{S}_n\,, \qquad \tilde{S}_n^\Gamma \overset{?}{\geq} 0\,. \tag{8.124}$$

To prove the first two inequalities, note that

$$\begin{aligned}
\tilde{S}_n \pm \left(\rho^{\otimes n}\right)^\Gamma &= S_0^{\otimes n} + \sum_{i=0}^{q-2} \left(S_{i+1}^{\otimes n} + \left(S_i^\Gamma\right)^{\otimes n}\right) + \left(S_{q-1}^\Gamma\right)^{\otimes n} \pm \left(\rho^{\otimes n}\right)^\Gamma \\
&\geq S_0^{\otimes n} + \left(S_{q-1}^\Gamma\right)^{\otimes n} \pm \left(\rho^{\otimes n}\right)^\Gamma \\
&\geq S_0^{\otimes n} \pm \left(\rho^{\otimes n}\right)^\Gamma \\
&\geq 0\,,
\end{aligned} \tag{8.125}$$

where all three inequalities come from (8.122). Consequently, we have proved that $-\tilde{S}_n \leq \left(\rho^{\otimes n}\right)^\Gamma \leq \tilde{S}_n$. Note that these two inequalities together imply that $\tilde{S}_n \geq 0$ (because $\tilde{S}_n \geq -\tilde{S}_n$). In addition, by exploiting the definition of $\tilde{S}_n$ in (8.123), we also have that $\tilde{S}_n^\Gamma = \tilde{S}_n$, and hence $\tilde{S}_n^\Gamma \geq 0$. We have thus proved that $\tilde{S}_n$ is a feasible point of the minimisation problem that defines (up to a logarithm) $E_\kappa\left(\rho^{\otimes n}\right)$. Hence, it follows that

$$\begin{aligned}
E_\kappa\left(\rho^{\otimes n}\right) &\leq \log_2 \operatorname{Tr} \tilde{S}_n \\
&= \log_2 \operatorname{Tr} \left[\sum_{i=0}^{q-1} \left(S_i^{\otimes n} + \left(S_i^\Gamma\right)^{\otimes n}\right)\right] \\
&\overset{\text{(i)}}{=} \log_2 \left(2 \sum_{i=0}^{q-1} (\operatorname{Tr} S_i)^n\right) \\
&\overset{\text{(ii)}}{\leq} \log_2 \left(2q \left(\operatorname{Tr} S_{q-1}\right)^n\right) \\
&= \log_2(2q) + n \log_2(\operatorname{Tr} S_{q-1}) \\
&\overset{\text{(iii)}}{=} \log_2(2q) + n\, E_{\kappa,q}(\rho)\,.
\end{aligned} \tag{8.126}$$

Here, in (i) we exploited (8.37), (ii) follows from the inequalities $\operatorname{Tr} S_0 \leq \operatorname{Tr} S_1 \leq \ldots \leq \operatorname{Tr} S_{q-1}$ (Lemma 14), and finally (iii) holds by the optimality of $S_{-1}, S_0, \ldots, S_{q-1}$. As a consequence of (8.126), we have that for all fixed $q \in \mathbb{N}^+$

$$E_\kappa^\infty(\rho) = \lim_{n\to\infty} \frac{1}{n} E_\kappa\left(\rho^{\otimes n}\right) \leq E_{\kappa,q}(\rho)\,, \tag{8.127}$$

which establishes (8.119) and thus concludes the proof. □

Remark 32. Theorem 31 also shows that

$$E^{\infty}_{\kappa,q}(\rho) := \lim_{n\to\infty} \frac{1}{n} E_{\kappa,q}\left(\rho^{\otimes n}\right) = E^{\infty}_{\kappa}(\rho) = E^{\text{exact}}_{c,\,\text{PPT}}(\rho) \qquad \forall\ \rho, \quad \forall\ q \in \mathbb{N}^{+}\,. \tag{8.128}$$

A key technical result: proof of Proposition 5

This subsection is devoted to the detailed proof of the crucial Proposition 5, which links the χ- and the κ-hierarchies in a profound way. This connection will be the keystone on which all of our main results are build.

Proposition 5. *For all bipartite states $\rho = \rho_{AB}$ on a system of minimal local dimension $d := \min\{|A|, |B|\} \geq 2$, and all positive integers $p \in \mathbb{N}^{+}$, it holds that*

$$\kappa_p(\rho) \leq \chi_p(\rho) + \left(\frac{d}{2} - 1\right)\left(\chi_p(\rho) - \chi_{p-1}(\rho)\right) \tag{8.129}$$

In particular,

$$E_{\chi,p}(\rho) \geq \frac{2}{d} E_{\kappa,p}(\rho) + \left(1 - \frac{2}{d}\right) E_{\chi,p-1}(\rho)\,. \tag{8.130}$$

To obtain the above result we first need a key technical lemma, proved below. Before stating it, let us fix some terminology. Given an arbitrary positive semi-definite bipartite operator $T = T_{AB}$, we set

$$d_{\max}\left(T \| \text{PPT}\right) := \min\left\{\operatorname{Tr} L :\ T \leq L\,,\quad L \in \text{PPT}\right\} = \min\left\{\operatorname{Tr} L :\ T \leq L,\ L \geq 0,\ L^{\Gamma} \geq 0\right\}. \tag{8.131}$$

It is well known that for all positive semi-definite T it holds that [76]

$$d_{\max}\left(T \| \text{PPT}\right) \leq d \operatorname{Tr} T\,. \tag{8.132}$$

One way to see this is, similarly to Corollary 19, by first noticing that for the maximally entangled state one has $d_{\max}(\Phi_d \| \text{PPT}) = d$ and then using the fact that any state $T/\operatorname{Tr} T$ can be obtained from Φ_d by an LOCC transformation, which can never increase $d_{\max}(\cdot \| \text{PPT})$. For completeness, let us give here a more self-contained proof of Eq. (8.132).

Up to taking positive linear combinations, it suffices to prove (8.132) when $T = |\psi\rangle\langle\psi|$ is the rank-one projector onto the pure state $|\psi\rangle = |\psi\rangle_{AB}$ with Schmidt decomposition $|\psi\rangle_{AB} = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |e_i\rangle_A |f_i\rangle_B$. Setting $L := |\psi\rangle\langle\psi|_{AB} + \sum_{i\neq j} \sqrt{\lambda_i \lambda_j}\, |e_i\rangle\langle e_i|_A \otimes |f_j\rangle\langle f_j|_B$ one sees that on the one hand $L \geq T = |\psi\rangle\langle\psi|$, while on the other

$$L^{\Gamma} \simeq \bigoplus_i \lambda_i \oplus \bigoplus_{i<j} \sqrt{\lambda_i \lambda_j} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \geq 0\,. \tag{8.133}$$

Then, using Cauchy–Schwarz and noting that $\sum_i \lambda_i = 1$ one finds that

$$\operatorname{Tr} L = 1 + \sum_{i\neq j} \sqrt{\lambda_i \lambda_j} = \sum_i \lambda_i + \sum_{i\neq j} \sqrt{\lambda_i \lambda_j} = \left(\sum\nolimits_i \sqrt{\lambda_i}\right)^2 \leq d\,, \tag{8.134}$$

which concludes the proof of (8.132).

The key lemma that is needed to prove Proposition 5 is the following.

Lemma 33. *For $p \geq 1$, let $S_{-1}, S_0, \ldots, S_{p-1}, S_p$ be optimisers for the SDP (8.60) that defines $\chi_p(\rho)$. Then there exists a PPT operator $M = M_{AB} \in \text{PPT}$ such that $S_{-1}, S_0, \ldots, S_{p-2}, S_{p-1} + M$ are valid ansatzes for the SDP that defines $\kappa_p(\rho)$ as in (8.66), and moreover $\operatorname{Tr} M = d_{\max}\left(\left(S^{\Gamma}_{p-1}\right)_{-} \middle\| \text{PPT}\right)$.*

Proof. Since $-S_i \leq S_{i-1}^\Gamma \leq S_i$ holds for all $i = 0, \ldots, p-1$ by assumption, and moreover $M \geq 0$, the only inequality that is left to verify is $(S_{p-1} + M)^\Gamma \geq 0$. By definition of $d_{\max}(\cdot \| \mathrm{PPT})$, there exists a PPT operator $L = L_{AB}$ such that $\left(S_{p-1}^\Gamma\right)_- \leq L$ and $\operatorname{Tr} L = d_{\max}\left(\left(S_{p-1}^\Gamma\right)_- \middle\| \mathrm{PPT}\right)$. We can then define $M := L^\Gamma$, so that

$$(S_{p-1} + M)^\Gamma = S_{p-1}^\Gamma + L = \left(S_{p-1}^\Gamma\right)_+ - \left(S_{p-1}^\Gamma\right)_- + L \geq \left(S_{p-1}^\Gamma\right)_+ \geq 0\,. \tag{8.135}$$

Since $\operatorname{Tr} M = \operatorname{Tr} L = d_{\max}\left(\left(S_{p-1}^\Gamma\right)_- \middle\| \mathrm{PPT}\right)$, this concludes the proof of the lemma. □

We are now ready to give the complete proof of Proposition 5.

Proof of Proposition 5. Let $S_{-1}, S_0, \ldots, S_{p-1}, S_p$ be optimisers for the SDP (8.60) that defines $\chi_p(\rho)$. Then clearly $S_{-1}, S_0, \ldots, S_{p-1}$ are valid ansatzes for the SDP that defines $\chi_{p-1}(\rho)$. Therefore,

$$\operatorname{Tr} S_{p-1} \geq \chi_{p-1}(\rho) = \chi_p(\rho) - \Delta_p = \operatorname{Tr} S_p - \Delta_p\,, \tag{8.136}$$

where we defined $\Delta_p := \chi_p(\rho) - \chi_{p-1}(\rho)$. We deduce that

$$0 \leq \delta_p := \operatorname{Tr}\left[S_p - S_{p-1}\right] \leq \Delta_p\,. \tag{8.137}$$

The first inequality is a consequence of Lemma 16 — basically, of the fact that $\delta_p = \operatorname{Tr}\left[S_p - S_{p-1}^\Gamma\right]$, and the operator inside the trace is positive semi-definite. Now, by Lemma 33 there exists a PPT operator M such that $S_{-1}, S_0, \ldots, S_{p-2}, S_{p-1} + M$ are legitimate ansatzes for the SDP that defines $\kappa_p(\rho)$ as in (8.66), and moreover

$$\begin{aligned}
\operatorname{Tr} M &= d_{\max}\left(\left(S_{p-1}^\Gamma\right)_- \middle\| \mathrm{PPT}\right) \\
&\overset{\text{(i)}}{\leq} d \operatorname{Tr}\left[\left(S_{p-1}^\Gamma\right)_-\right] \\
&= d\, \frac{\left\|S_{p-1}^\Gamma\right\|_1 - \operatorname{Tr} S_{p-1}}{2} \\
&\overset{\text{(ii)}}{\leq} d\, \frac{\operatorname{Tr} S_p - \operatorname{Tr} S_{p-1}}{2} \\
&= \frac{d\delta_p}{2}\,,
\end{aligned} \tag{8.138}$$

where the bound in (i) follows from (8.132), and that in (ii) comes from Lemma 6 applied to the operator inequalities $-S_p \leq S_{p-1}^\Gamma \leq S_p$. (The latter is, in fact, an equality if the operators S_i are optimisers for $\chi_p(\rho)$; however, we shall not make use of this observation.)

Putting all together, we obtain that

$$\begin{aligned}
\kappa_p(\rho) &\leq \operatorname{Tr}\left[S_{p-1} + M\right] \\
&= \operatorname{Tr} S_p - \delta_p + \operatorname{Tr} M \\
&\overset{\text{(iii)}}{=} \chi_p(\rho) - \delta_p + \operatorname{Tr} M \\
&\overset{\text{(iv)}}{\leq} \chi_p(\rho) + \left(\frac{d}{2} - 1\right)\delta_p \\
&\overset{\text{(v)}}{\leq} \chi_p(\rho) + \left(\frac{d}{2} - 1\right)\Delta_p \\
&= \chi_p(\rho) + \left(\frac{d}{2} - 1\right)\left(\chi_p(\rho) - \chi_{p-1}(\rho)\right),
\end{aligned} \tag{8.139}$$

which reproduces (8.129). Here, (iii) follows by the optimality of $S_{-1}, S_0, \ldots, S_p$ for $\chi_p(\rho)$, in (iv) we employed (8.138), and finally (v) follows from (8.137).

Rearranging, taking the logarithms, and using the concavity of the $\log_2$ function, one obtains also (8.130). □

What have we achieved so far?

Convergence

Proposition 5 is the core technical finding on which all of our main results rest. Now that we have proved it, most of the remaining proofs are comparatively straightforward. To show the importance of Proposition 5, it is instructive to pause for a second and explore some of its immediate consequences. In this subsection we do precisely that.

We first show that Proposition 5 already implies (8.8), i.e. that the limits of the χ- and κ-hierarchies coincide and yield the true zero-error PPT entanglement cost. In fact, taking the limit $p \to \infty$ on both sides of (8.129) and using the fact that both $\lim_{p\to\infty} \chi_p(\rho)$ and $\lim_{p\to\infty} \kappa_p(\rho)$ exist by monotonicity (Proposition 16) shows that

$$\lim_{p\to\infty} \kappa_p(\rho) \leq \lim_{p\to\infty} \chi_p(\rho)\,. \tag{8.140}$$

Combined with the reverse inequality, which is an obvious consequence of (8.71), one deduces that indeed $\lim_{p\to\infty} \kappa_p(\rho) = \lim_{p\to\infty} \chi_p(\rho)$. Now, going back to Theorem 31, and in particular to (8.120), we see that this implies that in fact

$$\lim_{p\to\infty} E_{\kappa,p}(\rho) = \lim_{p\to\infty} E_{\chi,p}(\rho) = E_\kappa^\infty(\rho) = E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)\,, \tag{8.141}$$

i.e. the four central quantities appearing in (8.120) coincide for all states. We will re-derive this equalities with some additional guarantees on the speed of convergence while proving Theorem 1 below.

The identity in (8.141) is remarkable because it provides a precise connection between the zero-error PPT cost, given by the regularisation of E_κ and thus by a limit $n \to \infty$ over the number of copies n, and the limits *of the hierarchies*, intended as limits $p \to \infty$ on the hierarchy level. The fact that there would be any sort of connection between the asymptotic limit in the number of copies and that in the hierarchy level was a priori totally unclear, and it should be regarded as one of the neater results of our approach. However, Eq. (8.141) is still not completely satisfactory from a computational standpoint, because it does not provide any rigorous guarantee on the speed of convergence to the limit. Having such guarantees (either on the limit in n or on that in p) would turn (8.141) into an *algorithm*, because it would tell us at which n or p we would need to stop in order to achieve a certain approximation. If that could be done, then, due to the fact that $E_\kappa\left(\rho^{\otimes n}\right)$, $E_{\chi,p}(\rho)$, and $E_{\kappa,p}(\rho)$ are all computable via SDPs, we would know how to approximate $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$ up to an arbitrary accuracy. In short, whether the algorithm we have described works or not depends on whether or not we can say something about the speed of convergence to the limit in either n or p. Furthermore, even if the algorithm does exist, then whether it is efficient or not depends on the actual speed of convergence to the limit, either in n or in p.

The key observation underpinning our entire approach is that while there seems to be no easy way of achieving either of the above two goals when dealing with the limit in the number of copies n, the situation changes dramatically for the better when considering the limit in the hierarchy level p. Not only will we be able to give universal, uniform bounds on the convergence speed, but the resulting convergence will also turn out to be exponentially fast (Theorem 1), resulting in an *efficient* algorithm to calculate the zero-error PPT entanglement cost on all states (Theorem 3).

The local qubit case

Proposition 5 also implies an immediate simple solution to the problem of computing $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho_{AB})$ when either A or B is a single-qubit system. In this case, it turns out that both hierarchies collapse at the first level ($p = 1$), entailing that $E_\chi = E_\kappa = E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}$ already gives the true zero-error PPT cost.

Corollary 2. *Let AB be a bipartite quantum system in which either A or B is a single qubit, i.e. $d = \min\{|A|, |B|\} = 2$. Then for all states $\rho = \rho_{AB}$ it holds that*

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_\chi(\rho) = E_\kappa(\rho)\,, \tag{8.142}$$

where E_χ and E_κ are defined by (8.57) and (8.43), respectively.

Note. Remember that we identify $E_\chi = E_{\chi,1}$ (cf. (8.57) and (8.59)) and $E_\kappa = E_{\kappa,1}$ (cf. (8.43) and (8.65)).

Proof. Writing down (8.129) for $d = 2$ and $p = 1$ shows that $\kappa_1(\rho) \leq \chi_1(\rho)$. Combining this with Theorem 31 then implies the claim. □

Remark 34. The above Corollary 2 does not tell us what happens to the zeroth level of the χ-hierarchy, namely, the logarithmic negativity. An older result by Ishizaka [36] states that all *two-qubit* states have zero bi-negativity, implying, by Lemma 17, that

$$|A| = |B| = 2 \qquad \Longrightarrow \qquad E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_N(\rho) = \log_2 \left\|\rho^\Gamma\right\|_1 \qquad \forall\ \rho = \rho_{AB}\,. \tag{8.143}$$

When e.g. $|A| = 2$ but $|B| > 2$, Corollary 2 guarantees that $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}} = E_\chi = E_\kappa$, but it would be even better if one could establish a closed-form expression for this quantity. One way to achieve this, for instance, would be by generalising Ishizaka's result so as to encompass all qubit-qudit systems. If that could be done, then again we would find that $E_{c,\,\mathrm{PPT}}^{\mathrm{exact}} = E_N$, which would be our sought closed-form expression. We leave a full understanding of the role of the logarithmic negativity in the qubit-qudit case as an open problem.

Exponential convergence: proof of Theorem 1

This subsection is devoted to the proof of our first main result, Theorem 1, which will be seen to be a relatively straightforward consequence of the key Proposition 5.

Theorem 1. *For all bipartite states $\rho = \rho_{AB}$ on a system of minimal local dimension $d := \min\{|A|, |B|\} \geq 2$, and all positive integers $p \in \mathbb{N}^+$, it holds that*

$$E_{\chi,p}(\rho) \leq E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) \leq E_{\kappa,p}(\rho) \leq E_{\chi,p}(\rho) + \log_2 \frac{1}{1 - \left(1 - \frac{2}{d}\right)^p}\,. \tag{8.144}$$

In particular,

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_\kappa^\infty(\rho) = \lim_{p\to\infty} E_{\chi,p}(\rho) = \lim_{p\to\infty} E_{\kappa,p}(\rho)\,, \tag{8.145}$$

with the convergence in p being exponentially fast uniformly on ρ.

Proof. We give here a slightly modified and extended version of the argument presented in the main text. The reason for doing so is to incorporate also the convergence of the κ-hierarchy in a single statement. To deal with the χ- and κ-hierarchies simultaneously, it is useful to set by convention $\kappa_0(\rho) := d$. Now, define the numbers

$$\varepsilon_p(\rho) := 1 - \frac{\chi_p(\rho)}{\kappa_p(\rho)}\,, \tag{8.146}$$

which satisfy $\varepsilon_p \in [0, 1]$ for all $p \in \mathbb{N}$ due to Proposition 16. Note that this definition differs from that employed in the main text (Eq. (8.27)) — in particular, the quantity in (8.146) is larger than that in (8.27). We can now write

$$\begin{aligned} 1 &\overset{\text{(i)}}{\leq} \frac{d}{2}\frac{\chi_p(\rho)}{\kappa_p(\rho)} - \left(\frac{d}{2} - 1\right)\frac{\chi_{p-1}(\rho)}{\kappa_p(\rho)} \\ &\overset{\text{(ii)}}{\leq} \frac{d}{2}\frac{\chi_p(\rho)}{\kappa_p(\rho)} - \left(\frac{d}{2} - 1\right)\frac{\chi_{p-1}(\rho)}{\kappa_{p-1}(\rho)} \\ &\overset{\text{(iii)}}{=} 1 - \frac{d}{2}\,\varepsilon_p + \left(\frac{d}{2} - 1\right)\varepsilon_{p-1}\,. \end{aligned} \tag{8.147}$$

The justification of the above steps is as follows: the inequality in (i) is just a rephrasing of that established by Proposition 5, obtained by diving both sides by $\kappa_p(\rho)$ (remember that $\kappa_p(\rho) \geq 1$ by Remark 15); in (ii) we observed that $\kappa_p(\rho) \leq \kappa_{p-1}(\rho)$ due to Proposition 16 (this is also true for $p = 1$ due to Lemma 17) and $d \geq 2$; finally, in (iii) we employed the definition of ε_p in (8.146).

We now see that (8.147) can be readily massaged into

$$\varepsilon_p \leq \left(1 - \frac{2}{d}\right) \varepsilon_{p-1} \,. \tag{8.148}$$

Iterating this p times and using the fact that $\varepsilon_0 \leq 1$ and $\varepsilon_p \geq 0$ yields immediately

$$0 \leq \varepsilon_p \leq \left(1 - \frac{2}{d}\right)^p , \tag{8.149}$$

which in turn can be rephrased as

$$\kappa_p(\rho) \leq \frac{\chi_p(\rho)}{1 - \left(1 - \frac{2}{d}\right)^p} \,. \tag{8.150}$$

Taking the logarithms of both sides proves the last inequality in (8.144). The first three inequalities are already known from Theorem 8.120.

Taking the limit $p \to \infty$ establishes also the convergence of both hierarchies to the true zero-error PPT entanglement cost (Eq. (8.145)). Note that the speed of convergence depends on d but not on ρ. □

The above result is central in our approach, because it provides a quantitative guarantee on what level of the hierarchies we have to resort to in order to obtain a prescribed approximation of the true value of the zero-error PPT entanglement cost. Before we proceed with the description of an algorithm that builds on this observation, we take note of two remarkable consequences of Theorem 1, namely, the full addivity of the zero-error PPT entanglement cost and its continuity.

Intuitively, the former means that the cheapest way to generate many copies of a state of the form $\rho_{AB} \otimes \omega_{A'B'}$, where A and A' belong to Alice, and B and B' to Bob, is to manufacture many copies of ρ_{AB} and $\omega_{A'B'}$ *separately*: in other words, there is no advantage to be gained in considering joint protocols. The continuity of the zero-error PPT cost, instead, allows us to estimate the cost of a state that is close enough to one for which the cost is known. The proofs of both of these facts are paradigmatic examples of how the knowledge we have gathered so far allows us to say a lot about the problem of zero-error PPT entanglement dilution even without a closed-form expression for the corresponding cost.

Corollary 35. *The zero-error entanglement cost under PPT operations is fully tensor additive, i.e. for all pairs of bipartite states $\rho_{AB}, \omega_{A'B'}$ it holds that*

$$E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}\left(\rho_{AB} \otimes \omega_{A'B'}\right) = E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}\left(\rho_{AB}\right) + E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}\left(\omega_{A'B'}\right) . \tag{8.151}$$

Proof. It suffices to take the limit $p \to \infty$ of (8.96) using (8.145). □

Corollary 36 (Continuity of the zero-error PPT entanglement cost)**.** *Let ρ, ρ' be two quantum states on a bipartite system AB with minimal local dimension $d \coloneqq \min\{|A|, |B|\}$. Let $\varepsilon \coloneqq \frac{1}{2} \left\|\rho - \rho'\right\|_1$. Then*

$$\left|2^{E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)} - 2^{E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho')}\right| \leq 2d\varepsilon \,, \tag{8.152}$$

and therefore

$$\left|E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) - E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho')\right| \leq \log_2\left(1 + 2d\varepsilon\right) \leq 2\left(\log_2 e\right) d\varepsilon \,. \tag{8.153}$$

Proof. It suffices to take the limit $p \to \infty$ of (8.112) and (8.113) written for $F = \chi_p$ and $f = E_{\chi,p}$, using Theorem 1, and in particular (8.145), to simplify the left-hand sides. □

An explicit algorithm to compute the zero-error PPT cost: proof of Theorem 3

We will now exploit the above Theorem 1 to prove the last of our main result, Theorem 3. To this end, we will design an efficient algorithm that computes the zero-error PPT entanglement cost up to any desired accuracy.

Theorem 3. *[(Efficient algorithm to compute the cost)] Let AB be a bipartite quantum system of total dimension $D := |AB| = |A||B|$ and minimal local dimension $d := \min\{|A|, |B|\}$. Then there exists an algorithm that for an arbitrary state $\rho = \rho_{AB}$ computes $2^{E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)}$ up to a multiplicative error ε — and hence, a fortiori, $E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)$ up to an additive error ε — in time*

$$\mathcal{O}\left((dD)^6\,\mathrm{polylog}\big(p, D, d, 1/\varepsilon\big)\right) = \mathcal{O}\left((dD)^{6+o(1)}\,\mathrm{polylog}(1/\varepsilon)\right). \tag{8.154}$$

Proof. To start, choose

$$p := \left\lceil \frac{\ln(2d/\varepsilon)}{\ln \frac{d}{d-2}} \right\rceil, \tag{8.155}$$

so that

$$\left(1 - \frac{2}{d}\right)^p \leq \frac{\varepsilon}{2d}. \tag{8.156}$$

Due to (8.144), we then know that

$$1 - \frac{\varepsilon}{2d} \leq \frac{\chi_p(\rho)}{2^{E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)}} \leq 1, \tag{8.157}$$

which implies that

$$\left|\chi_p(\rho) - 2^{E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)}\right| \leq \frac{\varepsilon}{2d}\, 2^{E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)} \leq \frac{\varepsilon}{2}, \tag{8.158}$$

where the last equality is a consequence of the fact that $E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho) \leq \log_2 d$ for all ρ by Lemma 17.

Now, $\chi_p(\rho)$ is given by an SDP and hence it can be computed efficiently up to additive error $\varepsilon/2$. We will look at the time it takes to carry out this computation in a moment, but for the time being assume that it yields an estimator $\widehat{\chi}_p(\rho)$ with the property that

$$|\widehat{\chi}_p(\rho) - \chi_p(\rho)| \leq \frac{\varepsilon}{2}. \tag{8.159}$$

Then by combining (8.158) and (8.159) one obtains that

$$\left|\widehat{\chi}_p(\rho) - 2^{E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)}\right| \leq \varepsilon. \tag{8.160}$$

This yields the claimed computation of $E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)$ up to an additive error ε. To estimate the error one incurs when computing $E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)$, it suffices to observe that

$$\left| E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho) - \log_2 \widehat{\chi}_p(\rho) \right| \leq (\log_2 e) \left| \widehat{\chi}_p(\rho) - 2^{E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)} \right| \leq \varepsilon \log_2 e, \tag{8.161}$$

so the error on $E^{\mathrm{exact}}_{c,\,\mathrm{PPT}}(\rho)$ is also ε, up to a constant. In the above derivation, we used the estimate $|a - b| \leq (\log_2 e)\left|2^a - 2^b\right|$, valid for all $a, b \geq 0$.

We still need to estimate the running time of the SDP that computes the estimator $\widehat{\chi}_p(\rho)$. To do that, we use the best known SDP solvers, which have time complexity [65] (see also the quantum-friendly review in [66])

$$\mathcal{O}\left(m\left(m^2 + n^\omega + mns\right)\mathrm{polylog}(m, n, R, 1/\varepsilon)\right) \tag{8.162}$$

once the SDP is put in the standard form

$$\begin{aligned} \max.\quad & \operatorname{Tr} CX \\ \text{s.t.}\quad & X \geq 0, \\ & \operatorname{Tr} A_j X \leq b_j, \quad j = 1, \ldots, m, \end{aligned} \tag{8.163}$$

where $\|A_j\|_\infty, \|C\|_\infty \leq 1$ for $j = 1, \ldots, m$, all matrices are $n \times n$, R is any upper bound on the optimal value, $\omega \in [2, 2.373)$ is the matrix multiplication exponent, and s is the sparsity, i.e. maximum number of non-zero entries in any row of the input matrices A_j, C.

What we have to do now is to cast the SDP for $\chi_p(\rho)$ (Eq. (8.60)) in the standard form (8.163). This can be done by restricting the variable X to be of the form

$$X \longrightarrow \bigoplus_{i=0}^{p} \begin{pmatrix} S_i & S_{i-1}^\Gamma \\ S_{i-1}^\Gamma & S_i \end{pmatrix}, \tag{8.164}$$

where $S_{-1} = \rho$. To see why, note that for a matrix of the above block form positive semi-definiteness implies that

$$\begin{pmatrix} S_i & S_{i-1}^\Gamma \\ S_{i-1}^\Gamma & S_i \end{pmatrix} \geq 0 \qquad \forall\, i = 0, \ldots, p\,, \tag{8.165}$$

which by a simple unitary rotation can be rephrased as

$$\frac{1}{2}\begin{pmatrix} \mathbb{1} & \mathbb{1} \\ \mathbb{1} & -\mathbb{1} \end{pmatrix}\begin{pmatrix} S_i & S_{i-1}^\Gamma \\ S_{i-1}^\Gamma & S_i \end{pmatrix}\begin{pmatrix} \mathbb{1} & \mathbb{1} \\ \mathbb{1} & -\mathbb{1} \end{pmatrix}^\dagger = \begin{pmatrix} S_i + S_{i-1}^\Gamma & 0 \\ 0 & S_i - S_{i-1}^\Gamma \end{pmatrix} \geq 0 \qquad \forall\, i = 0, \ldots, p\,, \tag{8.166}$$

i.e. $S_i \pm S_{i-1}^\Gamma \geq 0$ for all $i = 0, \ldots, p$, matching the positive semi-definite constraints in (8.60).

The above reasoning tells us that it is a good idea to try to enforce the structure (8.164) on X. This fixes the size of the matrices to be

$$n = 2(p+1)D = \mathcal{O}(pD)\,, \tag{8.167}$$

where $D = |A||B|$ is the total dimension. The problem therefore becomes that of ensuring that X has the structure in (8.164). To this end, it is useful to think of X as a block matrix of size $(p+1) \times (p+1)$, where each block has size $2D \times 2D$. We need to make sure that:

(1) all the $p(p+1)$ off-diagonal blocks are zero;

(2) inside each diagonal block X_{ii} $(i = 0, \ldots, p)$, which can in turn be thought of as a 2×2 block matrix $X_{ii} = \left(\begin{smallmatrix} H_i & K_i \\ L_i & M_i \end{smallmatrix}\right)$ with each block having size D, we have that $H_i = M_i$ and $K_i = L_i$;

(3) for all $i = 1, \ldots, p$, $K_i = H_{i-1}^\Gamma$; and

(4) $K_0 = \rho^\Gamma$.

All of these constraints are equalities instead of inequalities, but any equality can be written as two inequalities, and we do not count factors of 2. How many linear equality constraints on X do we need?

(1): $p(p+1)D^2$ constraints;

(2): $2(p+1)D^2$ constraints;

(3): pD^2 constraints;

(4): D^2 constraints.

We are therefore dealing with a total of $(p+1)(p+3)D^2$ linear equality constraints. It is therefore clear that we can set

$$m = \mathcal{O}(p^2 D^2)\,. \tag{8.168}$$

Now, constraints (1)–(3) are of the form $\operatorname{Tr} AX = 0$, where A has only two non-zero elements on the same row or column, while the constraints (4) are of the form $\operatorname{Tr} AX = b$, where A has a single non-zero entry (equal to 1). Note also that since we want $\operatorname{Tr} CX = \operatorname{Tr} S_p$, C needs to be

block diagonal, with the first p diagonal blocks equal to zero and the last one equal to $\mathbb{1}/2$ (all blocks are of size $2D$). We can thus set

$$s = \mathcal{O}(1)\,. \tag{8.169}$$

As for R, by Lemma 17 we already know that

$$R = d \tag{8.170}$$

is an upper bound on $\chi_p(\rho)$. Plugging the values (8.167)–(8.170) into (8.162) yields a time complexity

$$\begin{aligned}\mathcal{O}\left((pD)^6\,\mathrm{polylog}\big(p, D, d, 1/\varepsilon\big)\right) &= \mathcal{O}\left((dD)^6\,\mathrm{polylog}\big(p, D, d, 1/\varepsilon\big)\right)\\ &= \mathcal{O}\left((dD)^{6+o(1)}\,\mathrm{polylog}(1/\varepsilon)\right),\end{aligned} \tag{8.171}$$

where we used the fact that $p = \mathcal{O}\left(d\log(d/\varepsilon)\right)$ due to (8.155). □

Remark 37. It can be verified that the time complexity of solving the SDP that defines the negativity E_N is $\mathcal{O}\big(D^{6+o(1)}\mathrm{polylog}(1/\varepsilon)\big)$. While this is smaller than the complexity for computing $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$, it is only marginally so — the difference being a mere factor $\mathrm{poly}(d)$.

Open problem: hierarchy collapse

Although we showed that evaluating the limit $\lim_{p\to\infty} E_{\chi,p}(\rho)$ to any desired precision is computationally not much more demanding than evaluating $E_{\chi,p}$ for a fixed level of the hierarchy p, one may still wonder — perhaps from an analytical or aesthetic standpoint — whether computing the limit is truly necessary, or whether the hierarchy collapses at a certain point and there exists a finite $p^\star$ such that $E_{\chi,p} = E_{\chi,p^\star}\ \forall p \geq p^\star$. An analogous question can be asked for the κ-hierarchy.

One immediate consequence of Proposition 5 is that, if $E_{\chi,p}(\rho) = E_{\chi,p-1}(\rho)$ for some p, then $E_{\kappa,p}(\rho) \leq E_{\chi,p}(\rho)$, which would imply a complete collapse of both hierarchies, that is,

$$E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_{\chi,p}(\rho) = E_{\kappa,p}(\rho)\,, \tag{8.172}$$

due to Proposition 16.

We were not able to confirm nor disprove that this happens. However, numerical evidence suggests that the hierarchies collapse, and indeed they do so already at the second level of the χ-hierarchy. This leads us to posit the following conjecture.

Conjecture 38. *For all states $\rho = \rho_{AB}$, the χ-quantities defined by (8.60) satisfy that $E_{\chi,3}(\rho) = E_{\chi,2}(\rho)$. As a consequence, the χ- and κ-hierarchies collapse and $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_{\chi,2}(\rho) = E_{\kappa,3}(\rho)$.*

In fact, we were not even able to find a gap between $E_{\chi,2}$ and $E_{\kappa,2}$, so one could even make the stronger conjecture that $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho) = E_{\chi,2}(\rho) = E_{\kappa,2}(\rho)$ holds for all states, which would entail that both hierarchies collapse at the second level. For the sake of obtaining a single-letter, limit-free formula for $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$, it would be very interesting to resolve this question in future work. However, as remarked in the main text, doing so would only amount to a $\mathrm{poly}(d)$ reduction of the computational complexity of calculating the zero-error PPT entanglement cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}$.

8.4.5 From zero error to very small error

Zero-error entanglement manipulation tasks may seem on the surface un-physical, because nothing in nature happens with zero error, and small, undetectable errors should therefore always be included into the picture. Here we will do precisely that, and we will find that *if those errors are assumed to be sufficiently small (but non-zero), the resulting entanglement dilution rates are the*

same as in the zero-error case. In this context, 'sufficiently small' could mean for example 'going to 0 super-exponentially in the number of copies of the given state'.

This is no different from what happens in classical information theory, where the capacity for communicating on a noisy classical channel with error probability going to 0 super-exponentially equals its zero-error capacity, defined by Shannon in his landmark 1956 paper [77]. The argument to prove this claim seems to be part of the folklore in information theory, and it was brought to our attention by Andreas Winter. It goes as follows. A classical channel $\mathcal{N}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ can be thought of as a transition matrix $N(y|x) \geq 0$, with $\sum_y N(y|x) = 1$ for all $x \in \mathcal{X}$. A code to communicate on $\mathcal{N}^{\otimes n}$ is a list of words $x_1^n, \ldots, x_M^n$, where $M = 2^{\lceil nR \rceil}$, and R is the rate of the code. Suppose that for a fixed n a word in $\{1, \ldots, M\}$ is drawn at random. Whatever the decoder is, the probability of making an error can be estimated from below as follows. Pick two distinct $i, j \in \{1, \ldots, M\}$. For a fixed $y^n \in \mathcal{Y}^n$, call $p_i := \mathcal{N}^{\otimes n}(y^n|x_i^n)$ and $p_j := \mathcal{N}^{\otimes n}(y^n|x_j^n)$ the probabilities that these two words get transformed into the same output word $y^n \in \mathcal{Y}^n$. When that happens, even if the decoding party knows that the message was either i or j all they can do is to guess i or j with a maximum likelihood rule, i.e. i with probability $\frac{p_i}{p_i+p_j}$ and j with probability $\frac{p_j}{p_i+p_j}$. The total probability of error is thus at least

$$\frac{1}{M}\, p_i \, \frac{p_j}{p_i + p_j} + \frac{1}{M}\, p_j \, \frac{p_i}{p_i + p_j} = \frac{1}{M}\, \frac{2 p_i p_j}{p_i + p_j} \geq \frac{1}{M} \min\{p_i, p_j\} \,. \tag{8.173}$$

Now, if the right-hand side is not zero for some y^n, then it must be at least t^n/M, where $t := \min_{x,y:\, N(y|x)>0} N(y|x)$. Therefore the probability of error of the whole process (encoding, transmission, and decoding) is at least $t^n 2^{-\lceil nR \rceil}$, which cannot go to 0 super-exponentially. This completes the summary of the state of affairs in classical communication over noisy channels.

To arrive at analogous conclusions in the case of PPT entanglement dilution, we first need to fix some terminology. Our first task is to design a way to control the errors incurred in an arbitrary entanglement manipulation protocol. This can be done by introducing a rate-error pair achievability region for entanglement manipulation under a given set of free operations $\mathscr{F}$, defined as follows.

Definition 39. *For two bipartite states $\rho = \rho_{AB}$ and $\omega = \omega_{AB}$ and a class of free operations $\mathscr{F}$, we say that the rate-error pair (r, s) is achievable for the transformation $\rho \to \omega$ with operations in $\mathscr{F}$ if there exists a sequence of protocols $(\Lambda_n)_{n \in \mathbb{N}^+}$, with $\Lambda_n \in \mathscr{F}\,(A^n B^n \to A'^n B'^n)$ for all $n \in \mathbb{N}^+$, such that*

$$\liminf_{n \to \infty} \left\{ -\frac{1}{n} \log_2 \frac{1}{2} \left\| \Lambda_n\left(\rho^{\otimes \lceil rn \rceil}\right) - \omega^{\otimes n} \right\|_1 \right\} \geq s \,. \tag{8.174}$$

In our case, we care about the case where $\mathscr{F} = \mathrm{PPT}$ is the set of PPT operations, the initial state is the ebit Φ_2, and the final state is an arbitrary ρ. In this case, a rate-error pair that is achievable for $\Phi_2 \to \rho$ under PPT operations is also called an achievable rate-error pair for PPT entanglement dilution to ρ. The main result of this section is as follows.

Theorem 40. *Let $\rho = \rho_{AB}$ be an arbitrary finite-dimensional bipartite state. If (r, s) is achievable for PPT entanglement dilution to ρ, then*

$$\max\left\{r,\, \log_2 d - s\right\} \geq E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho) \,. \tag{8.175}$$

In particular, for all $s > \log_2 d - E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$ the rate-error pair (r, s) is achievable if and only if $r \geq E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$.

Before proving the above result, let us discuss its implications. What this shows is that if the error is required to decay *sufficiently fast*, i.e. faster than $2^{-n\left(\log_2 d - E_{c,\,\mathrm{PPT}}^{\mathrm{exact}}(\rho)\right)}$ asymptotically, then the optimal rate of entanglement dilution coincides with its zero-error value. In other words, once

a certain threshold in s is passed, it does not make a difference for the rate r whether s increases further, even if it goes all the way to infinity. Note that Theorem 40 applies to two suggestive special cases, namely where (i) we require the error to decay faster than d^{-n}; and therefore also when (ii) we require a super-exponential decay law. In both of these cases, the relevant optimal rate of PPT entanglement dilution coincides with the zero-error PPT entanglement cost $E_{c,\mathrm{PPT}}^{\mathrm{exact}}(\rho)$.

Proof of Theorem 40. Consider a sequence of PPT operations $(\Lambda_n)_{n\in\mathbb{N}^+}$ such that

$$\varepsilon_n := \frac{1}{2}\left\|\Lambda_n\left(\Phi_2^{\otimes\lfloor rn\rfloor}\right) - \rho^{\otimes n}\right\|_1 \tag{8.176}$$

exhibits the asymptotic decay rate

$$\liminf_{n\to\infty}\left\{-\frac{1}{n}\log_2\varepsilon_n\right\} \geq s\,. \tag{8.177}$$

Then,

$$\begin{aligned} 2^{\lfloor rn\rfloor} &= \kappa_1\left(\Phi_2^{\otimes\lfloor rn\rfloor}\right) \\ &\overset{\text{(i)}}{\geq} \kappa_1\left(\Lambda_n\left(\Phi_2^{\otimes\lfloor rn\rfloor}\right)\right) \\ &\overset{\text{(ii)}}{\geq} \kappa_1\left(\rho^{\otimes n}\right) - 2d^n\varepsilon_n \end{aligned} \tag{8.178}$$

where (i) is a consequence of the monotonicity of $\kappa_1 = 2^{E_\kappa}$ under PPT operations (Proposition 18), while (ii) is an application of Proposition 28, and in particular of (8.112) with $F = \kappa_1$. Therefore,

$$\begin{aligned} 1 + \max\left\{\,\lfloor rn\rfloor\,,\, 1 + n\log_2 d + \log_2\varepsilon_n\right\} &= \log_2\left(2\max\left\{2^{\lfloor rn\rfloor},\, 2d^n\varepsilon_n\right\}\right) \\ &\geq \log_2\left(2^{\lfloor rn\rfloor} + 2d^n\varepsilon_n\right) \\ &\geq \log_2\kappa_1\left(\rho^{\otimes n}\right) \\ &= E_\kappa\left(\rho^{\otimes n}\right). \end{aligned} \tag{8.179}$$

Dividing by n and taking the lim sup of both sides as $n\to\infty$ yields precisely

$$\begin{aligned} \max\{r,\,\log_2 d - s\} &\geq \limsup_{n\to\infty}\frac{1}{n}\left(1 + \max\left\{\,\lfloor rn\rfloor\,,\, 1 + n\log_2 d + \log_2\varepsilon_n\right\}\right) \\ &\geq \limsup_{n\to\infty}\frac{1}{n}E_\kappa\left(\rho^{\otimes n}\right) \\ &= \lim_{n\to\infty}\frac{1}{n}E_\kappa\left(\rho^{\otimes n}\right) \\ &= E_\kappa^\infty(\rho)\,, \end{aligned} \tag{8.180}$$

concluding the proof. □

Bibliography

[1] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory*, 44(1):269–273, 1998. ISSN 1557-9654. doi: 10.1109/18.651037.

[2] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997. doi: 10.1103/PhysRevA.56.131.

[3] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, 55:1613–1622, 1997. doi: 10.1103/PhysRevA.55.1613.

[4] P. Shor. The quantum channel capacity and coherent information. Lecture notes, MSRI Workshop on Quantum Computation, 2002.

[5] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, 51(1):44–55, 2005. ISSN 1557-9654. doi: 10.1109/TIT.2004.839515.

[6] N. Cai, A. Winter, and R. W. Yeung. Quantum privacy and quantum wiretap channels. *Probl. Inf. Transm.*, 40(4):318–336, 2004. doi: 10.1007/s11122-005-0002-x. URL https://doi.org/10.1007/s11122-005-0002-x.

[7] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83(15):3081, 1999. doi: 10.1103/PhysRevLett.83.3081.

[8] 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. *IEEE Trans. Inf. Theory*, 48(10): 2637–2655, 2002. ISSN 0018-9448. doi: 10.1109/TIT.2002.802612.

[9] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, 1996. doi: 10.1103/PhysRevA.53.2046.

[10] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–725, 1996. doi: 10.1103/PhysRevLett.76.722.

[11] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, 1996. doi: 10.1103/PhysRevA.54.3824.

[12] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. Royal Soc. A*, 461(2053):207–235, 2005. doi: 10.1098/rspa.2004.1372.

[13] L. Lami and B. Regula. Distillable entanglement under dually non-entangling operations. *Nat. Commun.*, 15(1):10120, 2024. doi: 10.1038/s41467-024-54201-5. URL https://doi.org/10.1038/s41467-024-54201-5.

[14] P. M. Hayden, M. Horodecki, and B. M. Terhal. The asymptotic entanglement cost of preparing a quantum state. *J. Phys. A*, 34(35):6891–6898, 2001. doi: 10.1088/0305-4470/34/35/314.

[15] D. Yang, M. Horodecki, R. Horodecki, and B. Synak-Radtke. Irreversibility for all bound entangled states. *Phys. Rev. Lett.*, 95:190501, 2005. doi: 10.1103/PhysRevLett.95.190501.

[16] F. Buscemi and N. Datta. Entanglement cost in practical scenarios. *Phys. Rev. Lett.*, 106: 130503, 2011. doi: 10.1103/PhysRevLett.106.130503. URL https://link.aps.org/doi/10.1103/PhysRevLett.106.130503.

[17] H. Yamasaki, K. Kuroiwa, P. Hayden, and L. Lami. Entanglement cost for infinite-dimensional physical systems. *Preprint arXiv:2401.09554*, 2024. doi: 10.48550/arXiv.2401.09554.

[18] K. G. H. Vollbrecht and R. F. Werner. Entanglement measures under symmetry. *Phys. Rev. A*, 64:062307, 2001. doi: 10.1103/PhysRevA.64.062307.

[19] P. W. Shor. Equivalence of additivity questions in quantum information theory. *Commun. Math. Phys.*, 246(3):473–473, 2004. doi: 10.1007/s00220-004-1071-1.

[20] P. Hayden and A. Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$. *Commun. Math. Phys.*, 284(1):263–280, 2008. doi: 10.1007/s00220-008-0624-0.

[21] T. Cubitt, A. W. Harrow, D. Leung, A. Montanaro, and A. Winter. Counterexamples to additivity of minimum output p-Rényi entropy for p close to 0. *Commun. Math. Phys.*, 284 (1):281–290, 2008. doi: 10.1007/s00220-008-0625-z.

[22] M. B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nat. Phys.*, 5(4):255–257, 2009. doi: 10.1038/nphys1224.

[23] G. Smith and J. Yard. Quantum communication with zero-capacity channels. *Science*, 321 (5897):1812–1815, 2008. doi: 10.1126/science.1162242.

[24] W. K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.*, 80:2245–2248, 1998. doi: 10.1103/PhysRevLett.80.2245. URL https://link.aps.org/doi/10.1103/PhysRevLett.80.2245.

[25] T. Cubitt, D. Elkouss, W. Matthews, M. Ozols, D. Pérez-García, and S. Strelchuk. Unbounded number of channel uses may be required to detect quantum capacity. *Nat. Commun.*, 6(1): 6739, 2015. doi: 10.1038/ncomms7739. URL https://doi.org/10.1038/ncomms7739.

[26] K. Audenaert, M. B. Plenio, and J. Eisert. Entanglement cost under positive-partial-transpose-preserving operations. *Phys. Rev. Lett.*, 90:027901, 2003. doi: 10.1103/PhysRevLett.90.027901.

[27] X. Wang and M. M. Wilde. Cost of quantum entanglement simplified. *Phys. Rev. Lett.*, 125: 040502, 2020. doi: 10.1103/PhysRevLett.125.040502.

[28] X. Wang and M. M. Wilde. Exact entanglement cost of quantum states and channels under positive-partial-transpose-preserving operations. *Phys. Rev. A*, 107:012429, 2023. doi: 10.1103/PhysRevA.107.012429.

[29] G. Gour and C. M. Scandolo. Dynamical entanglement. *Phys. Rev. Lett.*, 125:180505, 2020. doi: 10.1103/PhysRevLett.125.180505.

[30] G. Gour and C. M. Scandolo. Entanglement of a bipartite channel. *Phys. Rev. A*, 103:062422, 2021. doi: 10.1103/PhysRevA.103.062422.

[31] A. Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, 1996. doi: 10.1103/PhysRevLett.77.1413.

[32] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A*, 223(1–2):1–8, 1996. doi: https://doi.org/10.1016/S0375-9601(96)00706-2.

[33] P. Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Phys. Lett. A*, 232(5):333–339, 1997. doi: https://doi.org/10.1016/S0375-9601(97)00416-7.

[34] E. M. Rains. Bound on distillable entanglement. *Phys. Rev. A*, 60:179–184, 1999. doi: 10.1103/PhysRevA.60.179.

[35] E. M. Rains. A semidefinite program for distillable entanglement. *IEEE Trans. Inf. Theory*, 47(7):2921–2933, 2001. doi: 10.1109/18.959270.

[36] S. Ishizaka. Binegativity and geometry of entangled states in two qubits. *Phys. Rev. A*, 69:020301, 2004. doi: 10.1103/PhysRevA.69.020301. URL https://link.aps.org/doi/10.1103/PhysRevA.69.020301.

[37] G. Vidal and R. F. Werner. Computable measure of entanglement. *Phys. Rev. A*, 65:032314, 2002. doi: 10.1103/PhysRevA.65.032314.

[38] M. B. Plenio. Logarithmic negativity: A full entanglement monotone that is not convex. *Phys. Rev. Lett.*, 95:090503, 2005. doi: 10.1103/PhysRevLett.95.090503.

[39] Pasquale Calabrese, John Cardy, and Erik Tonni. Entanglement negativity in quantum field theory. *Phys. Rev. Lett.*, 109:130502, Sep 2012. doi: 10.1103/PhysRevLett.109.130502. URL https://link.aps.org/doi/10.1103/PhysRevLett.109.130502.

[40] P. Calabrese, J. Cardy, and E. Tonni. Entanglement negativity in extended systems: a field theoretical approach. *J. Stat. Mech.: Theory Exp.*, 2013(02):P02008, February 2013. ISSN 1742-5468. doi: 10.1088/1742-5468/2013/02/p02008. URL http://dx.doi.org/10.1088/1742-5468/2013/02/P02008.

[41] Pasquale Calabrese, John Cardy, and Erik Tonni. Finite temperature entanglement negativity in conformal field theory. *J. Phys. A*, 48(1):015006, 2014. ISSN 1751-8121. doi: 10.1088/1751-8113/48/1/015006. URL http://dx.doi.org/10.1088/1751-8113/48/1/015006.

[42] Yirun Arthur Lee and Guifre Vidal. Entanglement negativity and topological order. *Phys. Rev. A*, 88:042318, 2013. doi: 10.1103/PhysRevA.88.042318. URL https://link.aps.org/doi/10.1103/PhysRevA.88.042318.

[43] Marianne Hoogeveen and Benjamin Doyon. Entanglement negativity and entropy in non-equilibrium conformal field theory. *Nucl. Phys. B*, 898:78–112, September 2015. ISSN 0550-3213. doi: 10.1016/j.nuclphysb.2015.06.021. URL http://dx.doi.org/10.1016/j.nuclphysb.2015.06.021.

[44] V. Eisler and Z. Zimboras. Entanglement negativity in the harmonic chain out of equilibrium. *New J. Phys.*, 16(12):123020, December 2014. ISSN 1367-2630. doi: 10.1088/1367-2630/16/12/123020. URL http://dx.doi.org/10.1088/1367-2630/16/12/123020.

[45] Paola Ruggiero, Vincenzo Alba, and Pasquale Calabrese. Entanglement negativity in random spin chains. *Phys. Rev. B*, 94:035152, Jul 2016. doi: 10.1103/PhysRevB.94.035152. URL https://link.aps.org/doi/10.1103/PhysRevB.94.035152.

[46] X. Dong, J. Kudler-Flam, and P. Rath. Entanglement negativity and replica symmetry breaking in general holographic states. *Preprint arXiv:2409.13009*, 2024. doi: 10.48550/arXiv.2409.13009.

[47] X. Wang and M. M. Wilde. Errata for "Cost of quantum entanglement simplified" and "Exact entanglement cost of quantum states and channels under PPT-preserving operations", 2024. preprint https://doi.org/10.5281/zenodo.11061607.

[48] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein. Entangling operations and their implementation using a small amount of entanglement. *Phys. Rev. Lett.*, 86:544–547, 2001. doi: 10.1103/PhysRevLett.86.544.

[49] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature? *Phys. Rev. Lett.*, 80:5239–5242, 1998. doi: 10.1103/PhysRevLett.80.5239.

[50] T. Eggeling, K. G. H. Vollbrecht, R. F. Werner, and M. M. Wolf. Distillability via protocols respecting the positivity of partial transpose. *Phys. Rev. Lett.*, 87:257902, 2001. doi: 10.1103/PhysRevLett.87.257902.

[51] X. Wang and R. Duan. Irreversibility of asymptotic entanglement manipulation under quantum operations completely preserving positivity of partial transpose. *Phys. Rev. Lett.*, 119:180506, 2017. doi: 10.1103/PhysRevLett.119.180506.

[52] X. Wang, W. Xie, and R. Duan. Semidefinite programming strong converse bounds for classical capacity. *IEEE Trans. Inf. Theory*, 64(1):640–653, jan 2018. ISSN 1557-9654. doi: 10.1109/TIT.2017.2741101. URL http://dx.doi.org/10.1109/TIT.2017.2741101.

[53] K. Fang and H. Fawzi. Geometric Rényi divergence and its applications in quantum channel capacities. *Commun. Math. Phys.*, 384(3):1615–1677, 2021. doi: 10.1007/s00220-021-04064-4. URL https://doi.org/10.1007/s00220-021-04064-4.

[54] M. Hayashi. *Quantum Information Theory: Mathematical Foundation*. Graduate Texts in Physics. Springer Berlin Heidelberg, 2nd edition, 2017.

[55] K. Audenaert, B. De Moor, K. G. H. Vollbrecht, and R. F. Werner. Asymptotic relative entropy of entanglement for orthogonally invariant states. *Phys. Rev. A*, 66:032310, 2002. doi: 10.1103/PhysRevA.66.032310.

[56] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. doi: 10.1017/9781316848142.

[57] Paul Skrzypczyk and Daniel Cavalcanti. *Semidefinite Programming in Quantum Information Science*. IOP Publishing, March 2023. ISBN 9780750333412. doi: 10.1088/978-0-7503-3343-6. URL http://dx.doi.org/10.1088/978-0-7503-3343-6.

[58] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38:49–95, 1996. ISSN 0036-1445. doi: 10.1137/1038003.

[59] H. Fawzi and O. Fawzi. Defining quantum divergences via convex optimization. *Quantum*, 5: 387, 2021. ISSN 2521-327X. doi: 10.22331/q-2021-01-26-387. URL https://doi.org/10.22331/q-2021-01-26-387.

[60] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Complete family of separability criteria. *Phys. Rev. A*, 69:022308, 2004.

[61] M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.*, 10:073013, 2008. ISSN 1367-2630. doi: 10.1088/1367-2630/10/7/073013.

[62] A. W. Harrow, A. Natarajan, and X. Wu. An improved semidefinite programming hierarchy for testing entanglement. *Commun. Math. Phys.*, 352:881–904, 2017. ISSN 1432-0916. doi: 10.1007/s00220-017-2859-0.

[63] O. Fawzi, A. Shayeghi, and H. Ta. A hierarchy of efficient bounds on quantum capacities exploiting symmetry. *IEEE Trans. Inf. Theory*, 68:7346–7360, 2022. ISSN 0018-9448, 1557-9654. doi: 10.1109/TIT.2022.3182101.

[64] M. Berta, F. Borderi, O. Fawzi, and V. B. Scholz. Semidefinite programming hierarchies for constrained bilinear optimization. *Math. Program.*, 194:781–829, 2022. ISSN 1436-4646. doi: 10.1007/s10107-021-01650-1.

[65] Y. T. Lee, A. Sidford, and S. C.-W. Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *Proc. 56th Annual Symp. Found. Computer Science*, pages 1049–1065, 2015. doi: 10.1109/FOCS.2015.68.

[66] J. van Apeldoorn, A. Gilyén, S. Gribling, and R. de Wolf. Quantum SDP-Solvers: Better upper and lower bounds. *Quantum*, 4:230, 2020. ISSN 2521-327X. doi: 10.22331/q-2020-02-14-230.

[67] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, 1989. doi: 10.1103/PhysRevA.40.4277.

[68] D. Bruß and A. Peres. Construction of quantum states with bound entanglement. *Phys. Rev. A*, 61:030301, 2000. doi: 10.1103/PhysRevA.61.030301.

[69] M. Fekete. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.*, 17(1):228–249, 1923. ISSN 1432-1823. doi: 10.1007/BF01504345.

[70] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83:436–439, 1999. doi: 10.1103/PhysRevLett.83.436.

[71] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993. doi: 10.1103/PhysRevLett.70.1895.

[72] S. Khatri and M. M. Wilde. Principles of quantum communication theory: A modern approach. *preprint arXiv:2011.04672*, 2020.

[73] M. Horodecki. Entanglement measures. *Quantum Inf. Comput.*, 1(1):3–26, 2001. ISSN 1533-7146. doi: 10.26421/QIC1.1-2.

[74] M. J. Donald, M. Horodecki, and O. Rudolph. The uniqueness theorem for entanglement measures. *J. Math. Phys.*, 43(9):4252–4272, 2002. doi: 10.1063/1.1495917.

[75] L. Lami and B. Regula. No second law of entanglement manipulation after all. *Nat. Phys.*, 19(2):184–189, 2023. doi: 10.1038/s41567-022-01873-9.

[76] G. Vidal and R. Tarrach. Robustness of entanglement. *Phys. Rev. A*, 59:141–155, 1999. doi: 10.1103/PhysRevA.59.141.

[77] C. Shannon. The zero error capacity of a noisy channel. *IRE Trans. Inf. Theory*, 2(3):8–19, 1956. doi: 10.1109/TIT.1956.1056798.

Part IV

Conclusion

Conclusion

This thesis investigated quantum information theory with continuous-variable (CV) systems, while also including selected results on finite-dimensional systems. Specifically, the thesis was organised into three parts, covering quantum learning theory and quantum Shannon theory in the CV setting, and quantum Shannon theory in the finite-dimensional setting.

In Part I, we investigated the intersection between quantum learning theory and CV systems. In particular, we addressed the question of the ultimate achievable limits of quantum tomography in the CV setting. Our results establish that: (i) tomography of non-Gaussian states is extremely inefficient; (ii) tomography of Gaussian states is efficient; and (iii) the sample complexity of CV tomography grows exponentially with the degree of non-Gaussianity of the unknown state. As a by-product of our analyses, we established several bounds on the trace distance between CV states in terms of their covariance matrices, which provides a useful toolbox for research in quantum learning theory with CV systems and potentially beyond.

In Part II, we turned to quantum Shannon theory in the CV setting. We began by determining the maximum level of noise compatible with quantum key distribution and entanglement distribution over phase-insensitive bosonic Gaussian channels, which constitute standard models for optical fibres and free-space communication links. To this end, we developed a novel protocol for entanglement—-and hence secret-key—distribution, yielding state-of-the-art lower bounds on both the secret-key capacity and the two-way quantum capacity in the high-noise regime. We then investigated quantum communication over the bosonic loss-dephasing channel, discovering novel regions of the loss–dephasing parameter space for which it is impossible to perform quantum communication and quantum error correction. Finally, we studied quantum optical communication in the presence of non-Markovian (memory) effects. We exactly computed several capacities (specifically, the quantum capacity Q, the two-way quantum capacity Q_2, and the secret-key capacity K) of the resulting quantum memory channels, showing that such memory effects can be leveraged to enhance communication performance, even in noisy regimes that would be prohibitive under memoryless noise assumptions.

In Part III, the thesis shifted from continuous-variable systems to finite-dimensional systems and addressed a central obstacle in quantum Shannon theory: the problem of regularisations, i.e. formulas that involves an explicit limit $n \to \infty$ over the number of uses of the channel or available copies of the state. Regularisations make it computationally intractable to gain a precise quantitative understanding of the ultimate efficiency of key operational tasks such as entanglement manipulation. Here we considered the problem of computing the asymptotic entanglement cost of preparing noisy quantum states exactly under PPT operations. We constructed two hierarchies of semidefinite programmes that converged to the true asymptotic value from above and from below, and we proved that the convergence was exponentially fast, yielding an efficient algorithm that approximated the cost up to an additive error ε in time $\text{poly}\big(D, \log(1/\varepsilon)\big)$. To our knowledge, this was the first time that an asymptotic entanglement measure was shown to be efficiently computable despite no closed-form single-letter formula being available.