UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Distillation and Simulation in Quantum Information

by

Kun Fang

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Supervisor: Prof. Runyao Duan
Co-supervisor: Prof. Mingsheng Ying

Sydney, Australia

2018

Certificate of Original Authorship

I certify that the work in this thesis has not been previously submitted for a degree nor has it been submitted as a part of the requirements for other degree except as fully acknowledged within the text.

I also certify that this thesis has been written by me. Any help that I have received in my research and in the preparation of the thesis itself has been fully acknowledged. In addition, I certify that all information sources and literature used are quoted in the thesis.

Production Note:
Signature removed prior to publication.

Signature:

Date: 17/09/2018

Acknowledgements

I am very fortunate to meet lots of brilliant people during my PhD studies. Little of the work in this thesis would have been possible without their continuous support.

First and foremost, I would like to thank my supervisor Runyao Duan for his patient guidance and supervision during my research. I am deeply grateful for his inspiring discussions and thank him for encouraging me to discuss and collaborate with others. I would like to show my greatest appreciation to my co-supervisor Mingsheng Ying for his willingness to provide me with some insightful suggestions and keep reminding me to broaden my research areas.

A special thanks goes to my mentor Marco Tomamichel, who generously set aside time to answer my questions and patiently guide me in the entropy zoo. I can always learn something new after discussing with him. I also would like to send my special thanks to my friend and frequent collaborator Xin Wang for our innumerable discussions. Thanks him for teaching me the powerful tool of semidefinite programming and some practical techniques for doing numerical experiments.

Most of my work has been collaborative, and I am grateful to all my collaborators for inspiring discussions and productive collaborations. I would like to thank all my co-authors: Gerardo Adesso, Mario Berta, Runyao Duan, Xiaoyu He, Ludovico Lami, Bartosz Regula, Xiaoming Sun, Marco Tomamichel, Xin Wang and Wei Xie for many stimulating discussions, helpful advice and well-appreciated support. Besides the colleagues I have mentioned so far, I want to further extend my gratitude to Charles Bennett, Matthias Christandl, Omar Fawzi, Gilad Gour, Ji Guan, Masahito Hayashi, Yinan Li, Laura Mančinska, Alexander Müller-Hermes, Dong Yang for many valuable discussions.

I am grateful to the Centre for Quantum Software and Information at the University of Technology Sydney for providing an excellent research environment including a number of travel opportunities for graduate students. I also thank all the centre members creating such a pleasant and productive atmosphere. In the course of my PhD studies, I always had a great time and benefited a lot from my academic travels. In particular, I want to thank Gerardo Adesso, Mario Berta, Matthias Christandl, Xiaoming Sun for their hospitable host during my visits.

I would like to express my sincere thanks to all my friends. Particular thanks goes to Yinan Li and Xin Wang for their accompany. Finally, I want to save my deepest thanks to my family for their unconditional love and support.

Abstract

We use the techniques of convex optimization, especially semidefinite programming, to study two kinds of fundamental tasks, i.e., distillation and simulation in quantum information theory. We investigate these tasks in a unified framework of resource theory and focus on their computation and characterization with finite resources. Particularly we study the tradeoff among relevant parameters such as the number of resource copies, resource transformation rate, error tolerance and success probability.

In the first part, we study the task of distillation for two different resources, maximally entangled state and maximally coherent state, representing nonlocal and local "quantumness" respectively. For entanglement distillation, we derive an efficiently computable second-order estimation of the distillation rate for general quantum states, which are tight for quantum states of practical interest. Our study overcomes the limitation of conventional research either focusing on the asymptotic rate or ignoring the computability. For the coherence distillation, we perform finite analysis for both deterministic and probabilistic scenarios. Our results unveil several new features of coherence from a resource theoretic viewpoint and contribute to an increased understanding of the fundamental properties of different sets of free operations.

In the second part, we investigate the resource cost of simulating a quantum channel via quantum coherence or another quantum channel. We introduce the channel's analogs of max-relative entropy, logarithmic robustness and max-information of quantum states, providing their operational interpretation with the channel simulation cost via different resources. Particularly, we establish the asymptotic equipartition property of the channel's max-information, that is, it converges to the quantum mutual information of the channel in the independent and identically distributed asymptotic limit. As applications, this asymptotic equipartition property implies the quantum reverse Shannon theorem in the presence of non-signalling correlations.

From the perspective of resource theory, the worth of a resource can usually be characterized by the minimum distance to a set of useless resources under a proper distance measure. We give such characterization for all the tasks studied in this thesis, and find that the distance measure for the distillation and simulation process naturally corresponds to the quantum hypothesis testing relative entropy and the max-relative entropy, respectively.

List of Publications

During the time of my PhD study, I am honored to collaborate with many excellent researchers. Parts of this thesis are based on material contained in the following papers.

- **K. Fang**, X. Wang, M. Tomamichel, and R. Duan, *Non-asymptotic entanglement distillation*, submitted, arXiv:1706.06221, 2017. [Chapter 3]
- B. Regula, **K. Fang**, X. Wang, and G. Adesso, *One-shot coherence distillation*, Physical Review Letters 121, 010401, 2018. [Chapter 4]
- **K. Fang**, X. Wang, L. Lami, B. Regula, G. Adesso, *Probabilistic distillation of quantum coherence*, Physical Review Letters 121, 070404, 2018. [Chapter 4]
- **K. Fang**, X. Wang, M. Tomamichel, and M. Berta, *Quantum channel simulation and the channel's max-information*, submitted, arXiv:1807.05354, 2018. [Chapter 5]
- M. Díaz, **K. Fang**, X. Wang, M. Rosati, M. Skotiniotis, J. Calsamiglia, A. Winter, *Using and reusing coherence to realize quantum processes*, submitted, arXiv:1805.04045, 2018. [Chapter 5]

Other work on which this manuscript does not focus:

- W. Xie, **K. Fang**, X. Wang, and R. Duan, *Approximate broadcasting of quantum correlations*, Physical Review A 96, 022302, 2017.
- X. Wang, **K. Fang**, and M. Tomamichel, *On converse bounds for classical communication over quantum channels*, submitted, arXiv:1709.05258, 2017.
- X. Wang, **K. Fang**, and R. Duan, *Semidefinite programming converse bounds for quantum communication*, submitted, arXiv:1709.00200, 2017.
- X. He, **K. Fang**, X. Sun, and R. Duan, *Quantum advantages in Hypercube game*, arXiv: 1806.02642, 2018.

Contents

List of Figures

Abbreviations and Notations

Abbreviation	Description
CPTP	Completely positive trace-preserving map
SDP	Semidefinite program
i.i.d.	Independent and identically distributed
1-LOCC	Local operations and one-way classical communication
LOCC	Local operations and classical communication
SEP	Separable
PPT	Positive partial transpose preserving
NS	Non-signalling
MIO	Maximally incoherent operation
DIO	Dephasing-covariant incoherent operation
IO	Incoherent operation
SIO	Strictly incoherent operation
PIO	Physical incoherent operation
AEP	Asymptotic equipartition property
QRST	Quantum reverse Shannon theorem
l.h.s.	Left-hand side of an equation
r.h.s.	Right-hand side of an equation
min	minimize
max	maximize
s.t.	subject to

Table 1 : A list of abbreviations.

Notation	Description
A, B', AB	Typical physical systems and joint systems
$\mathcal{H}_A$	Hilbert space over system A
$d_A, \|A\|$	Dimension of the system A
$\mathcal{S}_= (A)$	The set of normalized quantum states on $\mathcal{H}_A$
$\mathcal{S}_\leq (A)$	The set of sub-normalized quantum states on $\mathcal{H}_A$
SEP(A:B)	The set of separable states shared between A and B
PPT(A:B)	The set of PPT states shared between A and B
PPT′(A:B)	The Rains set
$\mathcal{I}$	The set of incoherent quantum states
$\mathcal{I}_H$	The set of diagonal Hermitian operators
ρ, σ, ω	Typical quantum states or subnormalized states
$\|\varphi\rangle, \varphi$	Typical pure states, i.e., rank-1 quantum states
$\Phi_{AA'}, \Phi_k$	Maximally entangled state (with local dimension k)
Ψ_k	Maximally coherent state (with dimension k)
Tr_A	Partial trace over system A
id_A	Quantum noiseless channel on the space $\mathcal{H}_A$
$F(\rho, \sigma)$	Generalized fidelity between ρ and σ
$P(\rho, \sigma)$	Purified distance between ρ and σ
$\log \equiv \log_2$	Logarithm in base 2
$\ln$	Natural logarithm
$\mathcal{L}(A), \mathcal{L}(A, B)$	The set of linear operators on $\mathcal{H}_A$ and from $\mathcal{H}_A$ to $\mathcal{H}_B$
$\mathcal{P}(A)$	The set of positive semi-definite operators on $\mathcal{H}_A$
$\mathcal{H}erm(A)$	The set of Hermitian operators on $\mathcal{H}_A$
$X \geq Y$	Equivalent to $X - Y \in \mathcal{P}(\mathcal{H})$
$X^\dagger$	The adjoint operator of X
$\overline{X}$	The complex conjugate of X
X^T, X^{T_A}	The transpose and partial transpose of X
$\mathbb{1}_A$	Identity operator on Hilbert space $\mathcal{H}_A$
$\mathcal{E}, \mathcal{F}$	Typical completely positive and trace-preserving maps
Π	Typical free quantum operations (codes)
Ω	Typical set of free operations (codes)
Δ	Completely dephasing channel (diagonalizing map)
$\mathcal{N}_{A\to B}, \mathcal{M}_{A\to B}$	Typical quantum channels from $\mathcal{L}(A)$ to $\mathcal{L}(B)$
$J_\mathcal{N}$	Choi-Jamiołkowski operator of the operation $\mathcal{N}$
$\|X\|_1$	Trace norm, i.e., the sum of all singular values of X
$\|X\|_\infty$	Operator norm, i.e., the largest singular value of X
$\|\mathcal{E}\|_\diamond$	Diamond norm (Completely bounded trace norm)

Table 2 : A list of basic notations.

Chapter 1

Introduction

1.1 Quantum information theory

The discipline of information theory was founded by Claude Shannon in a remarkable paper [Sha48] which laid down the foundations of this subject. The study of information theory concerns how information can be compressed, stored, transformed and processed. In retrospect, quantum information theory is largely motivated by its classical counterpart while adding the magic ingredient of *quantum mechanics*. This new theory contains quantum generalizations of classical components such as sources, channels, and codes, as well as some of its distinctive elements, like *quantum superposition* and *quantum entanglement*. It has become clear that an information theory based on quantum principles extends and completes the classical information theory, somewhat as complex numbers extend and complete the reals. Quantum information theory aims to explore the nature of information at the quantum level and reveals new information processing capabilities beyond what is possible in the traditional models.

One fundamental difference between quantum information theory and its classical counterpart is the basic unit of information. While classical information is characterized by the *bit* (binary digits) which would have to be in zero or one regardless of its physical representation, quantum information is, by contrast, measured by the *qubit* (quantum bit) — a two-state quantum-mechanical system, such as the polarization of a single photon, allowing for a superposition of both states at the same time. More generally, a quantum state of n qubits is represented by a complex unit vector in the 2^n-dimensional Hilbert space. The exponentially large dimensionality of this space distinguishes quantum information processing from its classical counterpart, whose state is described by a number of parameters that grows only linearly with the size of the system. Alternatively speaking, the classical counterpart can occupy any one of a complete set of orthogonal quantum states, whereas the quantum system can be not only in the orthogonal states, but also in any linear superposition of the orthogonal states. *"Hilbert*

space is a big place!", said Carlton M. Caves. Even simple quantum systems, having only a small Hilbert-space dimension, have the potential for considerable complexity due to quantum superposition.

Quantum information systems allow us to transmit data that is fundamentally secure and solve problems that are beyond the power of classical computers. For instance, *quantum cryptography* allows us to perform more secure protocols by exploiting the fact that observing data encoded in a quantum state changes the state [BB84]. Another well-known application is given by Shor's algorithm [Sho94] used to efficiently factorize large numbers which cannot be done in the classical world.

With the boosting investment from industry and academia, practical quantum information applications are just around the corner, with prototypes of quantum cryptographic setups, small-scale quantum computers already working in laboratories and the transnational quantum internet under construction. A deeper theoretical understanding of the information-theoretic aspects of the quantum theory would foster more exciting applications. Given the theoretical promises and the current rate of experimental progress, quantum information technologies are expected to bring transformative advances to our society in the coming decades.

In this thesis, we will study three fundamental components in quantum information theory, that is, quantum coherence, quantum entanglement and quantum channels. We introduce a bit more details about them as follows.

Quantum coherence

As a more general form of quantum superposition, *quantum coherence* represents one of the most fundamental features that set the difference of quantum mechanics from the classical realm. It plays a central role in physics as it enables applications that are impossible within classical mechanics. The rise of quantum mechanics as a unified picture of waves and particles further strengthened the prominent role of coherence in physics. Indeed, by combination of the energy quantization and the tensor product structure of state space, coherence underlies phenomena such as multi-particle interference and entanglement that play a central role in applications of quantum physics and quantum information science.

Quantum optical methods provide an important set of tools for the manipulation of coherence, and indeed, at its basis lies the formulation of the quantum theory of coherence. Here, coherence is studied in terms of phase space distributions and multi-point correlation functions to provide a framework that relates closely to classical electromagnetic phenomena [Gla63, Sud63]. However, quantum coherence is not only restricted to the optical fields. More importantly, as the key ingredient that powers quantum technologies, it would be highly

desirable to be able to precisely quantify the usefulness (worth) of coherence as a resource for such applications.

Following an early approach to quantifying superpositions of orthogonal quantum states by [Abe06], and the independent yet related resource theory of asymmetry [GMS09, GS08, MS16], a resource theory of quantum coherence has been primarily proposed in [BCP14, LM14] and further developed in [CG16b, CG16a, WY16]. Such a theory asks the question what can be achieved and at what resource cost when the devices that are available to us are essentially classical, that is, they cannot create coherence in a preferred basis. This analysis, currently still under development, attempts to provide a complete and rigorous framework to describe quantum coherence, in analogy with what has been done for quantum entanglement and other non-classical resources.

Within such a framework, recent progress has witnessed a growing number of applications certified to rely on various incarnations of quantum coherence as a primary ingredient. For example, the role of coherence in quantum algorithms was discussed in [Hil16], with particular focus on the Deutsch-Jozsa algorithm [DJ92]. This quantum algorithm decides whether a boolean function is constant or balanced by just one evaluation of the function, while in the classical case the number of evaluations grows exponentially in the number of input bits. As shown in [Hil16], quantum coherence is a resource in such protocol in the sense that a smaller amount of coherence in the protocol decreases the error of guessing whether the evaluated function was constant or balanced. Other applications of quantum coherence could be found in quantum key distribution [CML16], quantum state merging [SCR+16], state redistribution [AJS18], channel simulation [DFW+18], thermodynamics [LJR15] and quantum metrology [FD11].

Quantum entanglement

When an overall quantum system is made of parts, the superposition principle leads inevitably to the correlations called entanglement. Quantum entanglement is a striking phenomenon which occurs when pairs or groups of particles are generated or interact in ways such that the quantum state of each particle cannot be described independently of the state of the others, even when the particles are separated by a large distance — instead, a quantum state must be described for the system as a whole. Such phenomenon was the main subject of a paper by Einstein, Podolsky, and Rosen in 1935 [EPR35], and two papers by Schrödinger shortly thereafter [Sch35, Sch36], describing what came to be known as the EPR paradox. Einstein and others considered such behavior (also referred as "spooky action at a distance") to be impossible, as it violated the local realist view of *causality* and argued that the accepted formulation

of quantum mechanics must therefore be incomplete. Later, however, the counterintuitive predictions of quantum mechanics were verified experimentally (e.g. the latest one [RBG+17]) in tests where the polarization or spin of entangled particles were measured at separate locations, statistically violating *Bell's inequality* [Bel64], demonstrating that the classical conception of "local realism" cannot be correct.

It turns out that quantum entanglement is not only subject of philosophical debates but a new quantum resource useful in practice. One particular interesting example is given by the so-called *quantum teleportation* protocol [BBC+93]. With the help of classical communication and shared quantum entanglement, this protocol provides us a way of transporting a qubit from one location to another without having to move the physical particle along with it. Quantum teleportation was first realized in single photons [BPM+97], later being demonstrated in various material systems such as atoms, ions, electrons and superconducting circuits. Other applications of quantum entanglement in quantum information processing tasks include the efficient transmission of information via quantum dense coding [BW92, MWKZ96], the secure data transmission through entanglement-based quantum cryptography [Eke91], as well as both device-independent randomness generation [Col09, PAM+10] or amplification [CR12, GMDLT+13, BRG+16], quantum metrology [EdMFD11, GLM11] and the efficient solution of the factorization problem [Sho94]. All these protocols necessarily rely on entanglement resources, especially on the maximally entangled states.

Quantum channels

Any device, regardless of its physical implementation, taking states of classical or quantum systems of a certain type to other (possibly different) state of classical or quantum systems is a *channel*. Mathematically speaking, it is a formalism used to describe a broad class of transformations that a physical system can undergo. This definition contains any processing step in information theory, from preparations to free and controlled time evolution, as well as measurements. Channels are thus among the central concepts of both classical and quantum information science. *Classical channels* are those that can transmit or store only classical information, like electrical wires or the field-effect transistors while *quantum channels* can transmit both classical and quantum information. Physical realizations of quantum channels include everything from optical fibers or coupled spin chains for quantum communication, to charged atoms in ion traps for quantum storage.

As for communication, a quantum channel can be used in different ways: it can transmit classical information, private classical information, or quantum information. It can be used alone, with shared entanglement, or together with other channels. For each of these settings

there is a *capacity* that quantifies a channel's fundamental potential for communication. From a theoretical viewpoint, computing these various capacities becomes one of the central topics for quantum Shannon theory — a subfield of quantum information theory. However, from a practical point of view, it becomes essential to figure out how to implement a quantum channel as well as how much cost it takes for such implementation. This will correspond to another main topic called *channel simulation* [Sha48, BSST02].

1.2 Quantum resource theory

Quantum resource theory can be regarded as a theory of *interconversions* among different resources which can be classified as quantum or classical, noisy or noiseless, static or dynamic, and therefore enable the diversity of quantum information processing tasks. There are three elementary ingredients for the structure of any resource theory: *resources* (e.g. entanglement), *free* / *useless resources* (e.g. separable states) and restricted set of *free operations* (e.g. local operations and classical communication). To admit a compatible framework, these components are not independent of each other. It is required that we cannot generate resources from free resources via free operations.

Quantum resource theory provides a rather general framework to study various quantum information processing tasks. Indeed, much of quantum information theory is simply a theory of the interconversion between different resources [DW04, DHW08]. For instance, quantum teleportation is essentially a process of converting maximally entangled states to quantum noiseless channels, while the study of *quantum* (*classical*) *capacity* is, in fact, discussing the optimal rate of converting a given quantum channel to quantum (classical) noiseless channels. As shown in Figure 1.1 we summarize part of such resource trading processes discussed in this thesis. The maximal resource indicated in yellow is a particular resource that can be transformed to any others via free operations, playing the role of "currency" during the process of resource trading.

Depending on our tasks, there are usually two directions of concern. The process that transforms any given resource to a maximal resource is called *distillation* or *concentration* process, analogous to selling our resource to gain some "currency". Typical examples are entanglement distillation and coherence distillation. The distillation process can help us better utilize our resources for other purposes. On the other hand, the process of converting the maximal resource to the others are called *dilution* or *simulation*, such as entanglement dilution and coherence dilution. This is similar to spend our "currency" to purchase other resources. It provides us a different perspective to characterize the worth of a given resource. Notably, these two processes are not reversible in general, making the *reversibility problem* one of the main

topics in resource theory [BG15].

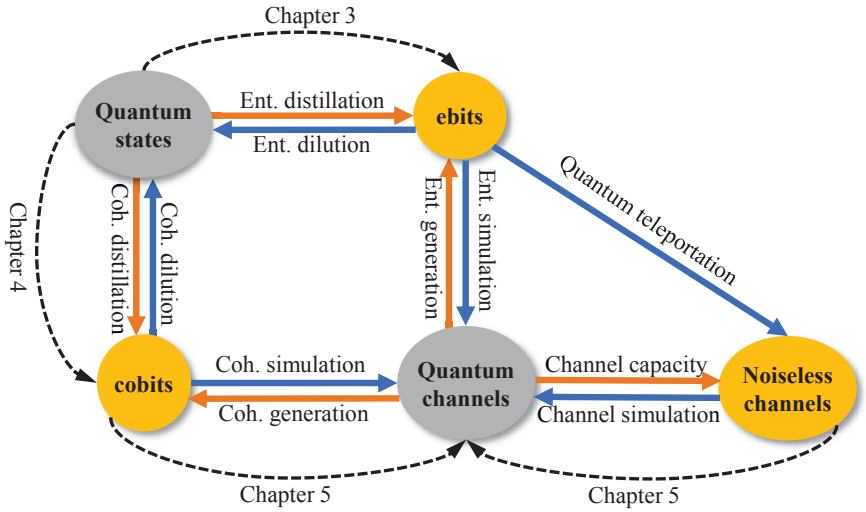
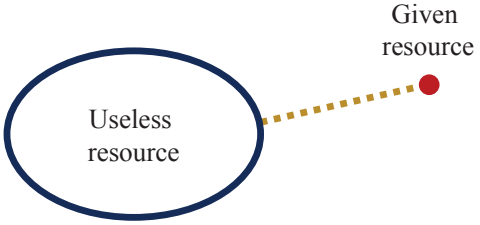


Figure 1.1 : Resources trading framework discussed in this thesis. “ebits” stands for the maximally entangled state while “cobit” stands for the maximally coherent state.

Quantum information is the study of the achievable limits to information processing possible within quantum mechanics. One important assumption regarding the information resource is that it is *independent and identically distributed* (i.i.d.). For example, we always suppose to have n copies of exactly the same quantum states which are not correlated with each other. Or we can use the same quantum channel n times while the channel itself has no quantum memory to store any information. Under such assumption, the conventional approach is to study the *asymptotic limit* of the rate that one resource can be transformed to another via free operations. However, from a practical point of view, the number of i.i.d. prepared resources is necessarily limited, and it remains a great challenge to physicists and engineers to develop techniques for performing large-scale quantum information processing in the near future. Thus it is important to characterize the interconversion process with *finite resources*, especially through small-sized or one-shot analysis.



In the spirit of resource theory, one can usually quantify the worth of a resource by the minimum "distance" to a set of well-defined useless resources with the proper distance measure. Such characterization is compatible with our intuition that the closer the given resource to the useless class, the less resource it contains. We summarize a few of such characterizations in Table 1.1, including the distillable entanglement under positive partial transpose preserving operations, distillable coherence under maximally incoherent operations and dephasing-covariant incoherent operations, coherence simulation cost under maximally incoherent operations and quantum simulation cost under non-signalling correlations. Particularly we find that the distance measure for the distillation and simulation process naturally corresponds to the quantum hypothesis testing relative entropy D_H^ε and the max-relative entropy $D_{\max}^\varepsilon$, respectively.

Quantifier	Useless resource	Distance measure	Reference
$E_{\mathrm{D,PPT}}^{(1),\varepsilon}$	Hermitian operators G s.t. $\|G^{T_B}\|_1 \leq 1$	D_H^ε	Section 3.2
$C_{\mathrm{c,M(D)IO}}^{(1),\varepsilon}$	Diagonal Hermitian operators with unit trace	D_H^ε	Section 4.2
$S_{c,\mathrm{MIO}}^{(1),\varepsilon}$	Maximally incoherent operations	$D_{\max}^\varepsilon$	Section 5.2
$S_{\mathrm{NS},\varepsilon}^{(1)}$	Constant channels	$D_{\max}^\varepsilon$	Section 5.3

Table 1.1 : Distance characterizations of resource theory discussed in this thesis. We can refer to Appendix A for a more complete summary of such formalism

1.3 Thesis organization

This thesis studies several information processing tasks with finite resource in details, including entanglement distillation, coherence distillation and quantum channel simulation. The remainder of this thesis is organized as follows.

Chapter 2 Preliminaries

Before we proceed to investigate these tasks, we introduce some commonly used tools and mathematical foundations in this chapter. We first review the basic components of quantum mechanics, including quantum states, quantum measurement and quantum operations. Then we give an overview of various quantum entropies and their basic properties. Finally we introduce the tool of semidefinite programming, its duality proof technique and a list of important semidefinite programs for later use.

Chapter 3 Entanglement distillation

Entanglement distillation, an essential quantum information processing task, refers to the conversion from multiple copies of noisy entangled states to a smaller number of highly entangled states. In this chapter we study the non-asymptotic fundamental limits for entanglement distillation. We investigate the optimal tradeoff between the distillation rate, the number of prepared states, and the error tolerance. First, we establish a precise connection between two different tasks in quantum information theory, i.e., entanglement distillation (under positive partial transpose preserving operations) and quantum hypothesis testing. Second, we provide efficiently computable estimations of the distillation rate for general quantum states, in both small and large-scale scenarios. In particular, we provide explicit as well as approximate evaluations for various quantum states of practical interest, including pure states, mixture of Bell states, maximally correlated states and isotropic states. These estimations can be used to benchmark the distillable entanglement for experimentalists.

Chapter 4 Coherence distillation

Coherence distillation is the task of converting general quantum states into maximally coherent states under a class of free operations. In this chapter we study coherence distillation in both deterministic and probabilistic scenarios. We investigate coherence distillation under two classes of free operations, highlighting differences in their capabilities and establishing their fundamental limitations in state transformations.

In section 4.2 we study the deterministic setting. We show that the one-shot distillable coherence under maximally incoherent operations (MIO) and dephasing-covariant incoherent operations (DIO) is efficiently computable with a semidefinite program, which then proved to relate with a quantum hypothesis testing problem. Notably, we find that MIO and DIO have the exactly same power in the task of one-shot coherence distillation.

In section 4.3 we develop a general framework of probabilistic distillation of quantum coherence, characterizing the maximal probability of success in the operational task of extracting maximally coherent states in a one-shot setting. We first provide a geometric interpretation for the maximal success probability. In a stark contrast with the deterministic case, we find that MIO can be much more powerful than DIO in general. We prove a fundamental no-go result that distilling perfect coherence from any full-rank state is impossible even probabilistically. We then present a phenomenon which prohibits any tradeoff between the maximal success probability and the distillation fidelity beyond a certain threshold. Finally, we consider probabilistic distillation assisted by a catalyst and demonstrate, with specific examples, its superiority to the deterministic case.

Chapter 5 Quantum channel simulation

As a fundamental problem in quantum information theory, quantum channel simulation considers simulating a given quantum channel via other quantum resources. In this chapter, we study the problem of channel simulation via quantum coherence or another quantum channel.

In section 5.2 we study the framework of quantum channel simulation via quantum coherence, discussing the simulation under maximally incoherent operations in details. First we show that the minimum error of coherence simulation and the one-shot coherence simulation cost can be both efficiently calculated via semidefinite programs. Second, we prove that the one-shot zero-error coherence simulation cost is additive and it is exactly equal to the maximal coherence generated from the channel, i.e., the cohering power of the channel. Finally, we introduce a channel version of max-relative entropy, building a distance characterization of the one-shot coherence simulation cost.

In section 5.3, we study the general framework of quantum channel simulation via another quantum channel. First, we show that the minimum error of simulation and the one-shot quantum simulation cost under non-signalling assisted codes are efficiently computable via semidefinite programming. Second, we introduce the channel's smooth max-information, which can be seen as a one-shot generalization of the mutual information of a quantum channel. We provide an exact operational interpretation of the channel's smooth max-information as the one-shot quantum simulation cost. We further introduce the channel's log-robustness and elaborate its relation with the channel's max-information. Third, we derive the asymptotic equipartition property (AEP) of the channel's smooth max-information, i.e., it converges to the quantum mutual information of the channel in the independent and identically distributed asymptotic limit. This implies the quantum reverse Shannon theorem (QRST) in the presence of non-signalling correlations. As applications, we explore finite blocklength simulation cost of fundamental quantum channels and provide both numerical and analytical solutions.

Chapter 2

Preliminaries

2.1 Basics of quantum mechanics

In this thesis we will focus on finite-dimensional Hilbert space and frequently use symbols such as $\mathcal{H}_A$ (or $\mathcal{H}_{A'}$) and $\mathcal{H}_B$ (or $\mathcal{H}_{B'}$) to denote Hilbert spaces associated with Alice and Bob, respectively. Let $|A|$ or d_A denote the dimension of $\mathcal{H}_A$. Let $\mathcal{L}(A,B)$ denote the set of linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$ and $\mathcal{L}(A) := \mathcal{L}(A,A)$. Let $\mathcal{H}erm(A)$ denote the set of all Hermitian operators on $\mathcal{H}_A$. Let $\mathcal{P}(A)$ denote the subset of positive semidefinite operators. We write $X \geq 0$ if $X \in \mathcal{P}(A)$. The identity operator on Hilbert space $\mathcal{H}_A$ is denoted as $\mathbb{1}_A = \sum_{i=0}^{d_A-1} |i_A\rangle\langle i_A|$. The identity map on $\mathcal{L}(A)$ is denoted as id_A. For a linear operator X, we define $|X| = \sqrt{X^\dagger X}$ where $X^\dagger$ is the Hermitian conjugate of X. The trace norm $\|X\|_1 = \mathrm{Tr}\,|X|$ is the sum of all singular values of X. The operator norm $\|X\|_\infty$ is defined as the maximum eigenvalue of $|X|$.

A full list of basic notations can be found in the part of Abbreviations and Notations. For more detailed introduction of quantum mechanics and quantum information theory, we can refer to the books [NC11, Wat17, Wil17].

2.1.1 Quantum states

A *quantum state* or *density operator* is a positive semidefinite operator with unit trace, usually denoted as ρ_A with subscript A indicating its corresponding Hilbert space. A *subnormalized quantum states* is a positive semidefinite operator with trace no greater than one. The set of quantum states and the set of subnormalized quantum states on Hilbert space $\mathcal{H}_A$ are respectively defined by

$$\mathcal{S}_=(A) := \{\, \rho \geq 0 \mid \mathrm{Tr}\,\rho = 1 \,\}, \tag{2.1}$$

$$\mathcal{S}_\leq(A) := \{\, \rho \geq 0 \mid 0 < \mathrm{Tr}\,\rho \leq 1 \,\}. \tag{2.2}$$

We call a quantum state *pure* if it has rank one. Otherwise, it is called *mixed state*. Pure states represent situations of minimal ignorance, in which, in principle, there is nothing more to be known about the system. Pure states are fundamental in that the quantum mechanics of a closed system can be completely described as a unitary evolution of pure states in an appropriately dimensioned Hilbert space, without the need of further notions.

For any bipartite quantum state ρ_{AB}, we usually denote $\rho_A := \mathrm{Tr}_B \rho_{AB}$ and $\rho_B := \mathrm{Tr}_A \rho_{AB}$ the marginal states of ρ_{AB}. For any quantum state ρ_A we can find a pure state $\phi_{AA'}$ such that $\rho_A = \mathrm{Tr}_{A'} \phi_{AA'}$. Then we call $\phi_{AA'}$ a *purification* of ρ_A. A canonical purification is given by [1]

$$\phi_{AA'} = \rho_A^{1/2} \widehat{\Phi}_{AA'} \rho_A^{1/2}, \tag{2.3}$$

where $\widehat{\Phi}_{AA'} = \sum_{i=0}^{d_A-1} |i_A i_{A'}\rangle\langle j_A j_{A'}|$ is the unnormalized maximally entangled state, $\{|i_A\rangle\}$ and $\{|i_{A'}\rangle\}$ are the standard, orthonormal bases on $\mathcal{H}_A$ and $\mathcal{H}_{A'}$ respectively. According to Uhlmann's theorem [Uhl76], any two purifications of the same state are local unitary equivalence. Suppose $\phi_{AA'}$ and $\widetilde{\phi}_{AA'}$ are two purifications of ρ_A, then there exists a unitary $U_{A'}$ such that

$$\phi_{AA'} = U_{A'} \widetilde{\phi}_{AA'} U_{A'}^\dagger. \tag{2.4}$$

Proposition 2.1 (Schmidt decomposition) *For any bipartite pure state $|\phi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ with $d = \min\{d_A, d_B\}$, there exist orthonormal bases $\{|e_j\rangle \in \mathcal{H}_A\}$ nad $\{|f_j\rangle \in \mathcal{H}_B\}$ such that*

$$|\phi_{AB}\rangle = \sum_{j=0}^{d-1} \sqrt{\lambda_j} |e_j\rangle \otimes |f_j\rangle, \tag{2.5}$$

with $\lambda_j \geq 0$ and $\sum_{j=0}^{d-1} \lambda_j = 1$. The coefficients $\sqrt{\lambda_j}$ are called Schmidt coefficients and the number of non-zero λ_j is the Schmidt rank of $|\phi\rangle$.

A bipartite quantum state ρ_{AB} is called *separable* (SEP) if it can be written as a convex combination of tensor product states, i.e.,

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i, \tag{2.6}$$

where $\rho_A^i \in \mathcal{S}_= (A)$, $\rho_B^i \in \mathcal{S}_= (B)$ and p_i is a probability distribution. The set of all separable states shared between A and B is denoted as $\mathrm{SEP}(A:B)$.

[1] In this thesis, we will write only X_A to indicate $X_A \otimes \mathbb{1}_{A'}$ when there is no risk of confusion.

A bipartite quantum state is *entangled* if it is not separable. One of the most important entangled states is the *maximally entangled state*, denoted as

$$\Phi_{AB} := \frac{1}{k} \sum_{i=0}^{k-1} |i_A i_B\rangle\langle j_A j_B|, \tag{2.7}$$

where k is the dimension of $\mathcal{H}_A$ and $\mathcal{H}_B$, $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ are the standard, orthonormal bases on $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. We sometimes denote Φ_{AB} as Φ_k to emphasize the dimension.

Proposition 2.2 (Transpose trick) *For any operator $M_A \in \mathcal{L}(A)$ and maximally entangled state Φ_{AB}, it holds*

$$(M_A \otimes \mathbb{1}_B) |\Phi_{AB}\rangle = \left(\mathbb{1}_A \otimes M_B^T\right) |\Phi_{AB}\rangle, \tag{2.8}$$

where M_B^T is the transpose of the operator M with respect to the basis $\{|i_B\rangle\}$.

A positive semidefinite operator $X \in \mathcal{P}(AB)$ is said to be *positive under partial transpose* (PPT) if its partial transpose remains positive. That is, $X^{T_B} \geq 0$, where T_B means the partial transpose over system B, i.e., $(|i_A j_B\rangle\langle k_A l_B|)^{T_B} = |i_A l_B\rangle\langle k_A j_B|$. Note that being PPT could equivalently be defined with respect to partial transposition over system A, and that being PPT does not depend on the specific basis chosen to take partial transposition (the partial transpose of a state in a different local basis is local-unitarily related to the one in the standard basis). The set of all PPT states shared between A and B is denoted as

$$\text{PPT}(A:B) := \left\{ \rho \in \mathcal{S}_=(AB) \;\middle|\; \rho^{T_B} \geq 0 \right\}. \tag{2.9}$$

With respect to coherence theory, a quantum state is called *incoherent* if it is diagonal in the given reference basis. We denote the set of all incoherent state as

$$\mathcal{I} := \left\{ \rho \in \mathcal{S}_=(A) \mid \rho = \Delta(\rho) \right\}, \tag{2.10}$$

where $\Delta(\rho) := \sum_{i=0}^{d_A-1} |i\rangle\langle i|\rho|i\rangle\langle i|$ is the *completely dephasing channel* (diagonalizing map). A quantum state is *coherent* if it is not incoherent. One of the most important coherent states is the maximally coherent state, denoted as

$$\Psi_k := \frac{1}{k} \sum_{i,j=0}^{k-1} |i\rangle\langle j|, \tag{2.11}$$

where k is the dimension of the quantum system.

2.1.2 Quantum measurement

As a method to extract information from a quantum system, a *quantum measurement* is mathematically described by a collection of measurement operators $\{E_i\}$ acting on the state space of the system being measured. The index i refers to the classical measurement outcomes in the experiment. If the state of the quantum system prior to the measurement is ρ, then the probability that result i occurs is given by $p_i = \operatorname{Tr} E_i \rho E_i^\dagger$ while the post-measurement state is $E_i \rho E_i^\dagger / p_i$. The measurement operators satisfy the completeness equation $\sum_i E_i^\dagger E_i = \mathbb{1}$, which ensures that the probabilities of the measurement outcomes sum to 1.

Without considering the post-measurement states, a quantum measurement can be defined in terms of a *positive operator-valued measure* (POVM). A POVM is given by a set of positive semidefinite operators $\{M_i\}$ satisfying the completeness condition $\sum_i M_i = \mathbb{1}$. The probability of measurement outcome i is given by $\operatorname{Tr} M_i \rho$.

2.1.3 Quantum operations

A *quantum operation* (sometimes called *quantum channel*) is a mathematical formalism used to describe a broad class of transformations that a quantum mechanical system can undergo.

Definition 2.3 *A linear map $\mathcal{E}$ from $\mathcal{L}(A)$ to $\mathcal{L}(B)$ is called a quantum operation if it satisfies the completely positive (CP) and trace-preserving (TP) conditions:*

- **(CP)** $\mathrm{id}_k \otimes \mathcal{E}(X) \geq 0$, for all $k \geq 0, X \in \mathcal{P}(RA)$, where id_k is the identity map on the reference system R with dimension k [2].
- **(TP)** $\operatorname{Tr} \mathcal{E}(Y) = \operatorname{Tr} Y$, for all $Y \in \mathcal{P}(A)$.

A relaxation of the TP condition is the *trace-nonincreasing* (TNI) condition which is given by $\operatorname{Tr} \mathcal{E}(Y) \leq \operatorname{Tr} Y$, for all $Y \in \mathcal{P}(A)$. A linear map $\mathcal{E}$ is called a *subnormalized quantum operation* or *subchannel* [3] if it is completely positive and trace-nonincreasing.

There are several equivalent representations for a quantum operation, including the Stinespring representation, Choi-Kraus representation, and Choi-Jamiołkowski representation. Here we introduce the last two in details.

---

[2] The dimension of the reference system in the CP condition can be restricted to $k = d_A$ [Cho75].

[3] Some authors use the term "quantum operation" to refer specifically to completely positive and trace nonincreasing maps, and the term "quantum channel" to refer to the subset of those that are trace preserving. Here we explicitly use the prefix "sub-" to indicate the trace non-increasing condition.

Choi-Kraus representation

A linear map $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ is completely positive if and only if there exists a finite set of linear operators $\{E_i\}$, $E_i \in \mathcal{L}(A, B)$ such that

$$\mathcal{E}(X) = \sum_i E_i X E_i^\dagger, \quad \forall\, X \in \mathcal{L}(A). \tag{2.12}$$

Furthermore, it is trace-preserving if and only if $\sum_i E_i^\dagger E_i = \mathbb{1}_A$ and it is trace-nonincreasing if and only if $\sum_i E_i^\dagger E_i \leq \mathbb{1}_A$ [Kra71, Cho75].

The Choi-Kraus representation (2.12) is not unique for a given quantum operation. However, there is always a representation (2.12) with at most $d_A d_B$ Kraus operators if the map is completely positive. Different representations can also be related by unitary transformations [NC11].

Choi-Jamiołkowski isomorphism

The Choi-Jamiołkowski isomorphism maps a given linear operation $\mathcal{E} : \mathcal{L}(A') \to \mathcal{L}(B)$ to a bipartite linear operator [Cho75, Jam72],

$$J_{\mathcal{E}} := \sum_{ij} |i_A\rangle\langle j_A| \otimes \mathcal{E}\left(|i_{A'}\rangle\langle j_{A'}|\right) \in \mathcal{L}(AB), \tag{2.13}$$

where $\{i_A\}$ and $\{i_{A'}\}$ are orthonormal bases on the isomorphic Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_{A'}$ respectively. The operator $J_{\mathcal{E}}$ is the so-called Choi-Jamiołkowski operator. On the other hand, the inverse Choi-Jamiołkowski isomorphism maps any bipartite operator $J_{\mathcal{E}} \in \mathcal{L}(AB)$ to a linear operation by

$$\mathcal{E}(X) := \mathrm{Tr}_A\, J_{\mathcal{E}} \left(X_A^T \otimes \mathbb{1}_B\right), \quad \forall\, X_A \in \mathcal{L}(A). \tag{2.14}$$

Based on the Choi-Jamiołkowski isomorphism, the completely positive, trace-preserving or trace-nonincreasing conditions for a linear map can be equivalently represented via its corresponding Choi-Jamiołkowski operator as follows:

- $\mathcal{E}$ is completely positive $\iff$ $J_{\mathcal{E}} \geq 0$;
- $\mathcal{E}$ is trace-preserving $\iff$ $\mathrm{Tr}_B\, J_{\mathcal{E}} = \mathbb{1}_A$;
- $\mathcal{E}$ is trace-nonincreasing $\iff$ $\mathrm{Tr}_B\, J_{\mathcal{E}} \leq \mathbb{1}_A$.

Note that above conditions on the r.h.s. are all linear conditions with respect to $J_{\mathcal{E}}$. This provides the opportunity to study resource transformation under different operations via semidefinite programming.

A hierarchy of bipartite quantum operations

The class of LOCC operations consists of all bipartite operations that can be realized by *local operations and classical communication*. If only *one-way* classical communication is allowed, say, classical information can only be sent from Alice to Bob, we call it 1-LOCC. Furthermore, if no classical communication is allowed, it is called local operation (LO), which admits the tensor product decomposition

$$\Pi_{AB\to A'B'} = \mathcal{E}_{A\to A'} \otimes \mathcal{F}_{B\to B'}, \tag{2.15}$$

where $\mathcal{E}$ and $\mathcal{F}$ represents the local operations performed by Alice and Bob respectively.

While the LO, 1-LOCC and LOCC possess clear operational interpretations, they are generally difficult to optimize over. Related classes that are more tractable are often studied instead. A bipartite quantum operation $\Pi_{AB\to A'B'}$ is said to be a positive partial transpose preserving (or separable) operation if its Choi-Jamiołkowski matrix [4]

$$J_\Pi = \sum_{i,j,m,k} |i_A j_B\rangle\langle m_A k_B| \otimes \Pi\left(|i_A j_B\rangle\langle m_A k_B|\right) \tag{2.16}$$

is positive under partial transpose (or separable) with the system partition $AA' : BB'$, where $\{|i_A\rangle\}$ and $\{|j_B\rangle\}$ are orthonormal basis on Hilbert spaces A and B, respectively. The set of all positive partial transpose preserving and separable operations are denoted as PPT and SEP respectively. Here we use the same notations for quantum states. But it is easy to tell from the context which one is referred to.

These bipartite operations play an important role in entanglement theory [HHHH09], and a well-known fact is that they obey the following strict inclusions [BDF+99],

$$\text{LO} \subsetneq \text{1-LOCC} \subsetneq \text{LOCC} \subsetneq \text{SEP} \subsetneq \text{PPT}. \tag{2.17}$$

Other two sets of bipartite operations are extensively used in quantum channel coding theory. An operation $\Pi_{AB\to A'B'}$ is called *entanglement-assisted* (EA) operation if it can be implemented via local operations with shared quantum entanglement, i.e., there exists a quantum state $\Phi_{A''B''}$ such that for any quantum state ρ_{AB}, it holds

$$\Pi_{AB\to A'B'}\left(\rho_{AB}\right) = \mathcal{E}_{AA''\to A'} \otimes \mathcal{F}_{BB''\to B'}\left(\Phi_{A''B''} \otimes \rho_{AB}\right), \tag{2.18}$$

where $\mathcal{E}$ and $\mathcal{F}$ are quantum operations performed by Alice and Bob respectively. If the shared entangled state is trivial (dimension one), then the entanglement-assisted operation reduce to a local operation.

[4] We will use the label A more than once, when there is no risk of confusion.

A bipartite quantum operation $\Pi_{AB\to A'B'}$ is *A to B non-signalling* ($A \not\to B$) if A cannot send classical information to B by using Π. Mathematically, for any quantum states $\rho_A, \sigma_A \in \mathcal{S}_=(A)$ and $\tau_B \in \mathcal{S}_=(B)$, we have $\mathrm{Tr}_{A'}\,\Pi_{AB\to A'B'}(\rho_A \otimes \tau_B) = \mathrm{Tr}_{A'}\,\Pi_{AB\to A'B'}(\sigma_A \otimes \tau_B)$. Or equivalently it can be characterized by its Choi-Jamiołkowski operator [LM15a, DW16],

$$\mathrm{Tr}_{A'}\,J_\Pi = \frac{\mathbb{1}_A}{d_A} \otimes \mathrm{Tr}_{AA'}\,J_\Pi. \tag{2.19}$$

Similarly, Π is said to be *B to A non-signalling* ($B \not\to A$) if for any quantum states $\rho_B, \sigma_B \in \mathcal{S}_=(B)$ and $\tau_A \in \mathcal{S}_=(A)$, we have $\mathrm{Tr}_{B'}\,\Pi_{AB\to A'B'}(\tau_A \otimes \rho_B) = \mathrm{Tr}_{B'}\,\Pi_{AB\to A'B'}(\tau_A \otimes \sigma_B)$. It can be equivalently characterized by its Choi-Jamiołkowski operator

$$\mathrm{Tr}_{B'}\,J_\Pi = \frac{\mathbb{1}_B}{d_B} \otimes \mathrm{Tr}_{BB'}\,J_\Pi. \tag{2.20}$$

Furthermore, $\Pi_{AB\to A'B'}$ is a *non-signalling* operation if it is non-signalling from A to B and vice versa. The class of non-signalling operations is strictly larger than the class of all entanglement-assisted operations [LM15b]. We summarize the relation of bipartite operations in the following Figure 2.1.

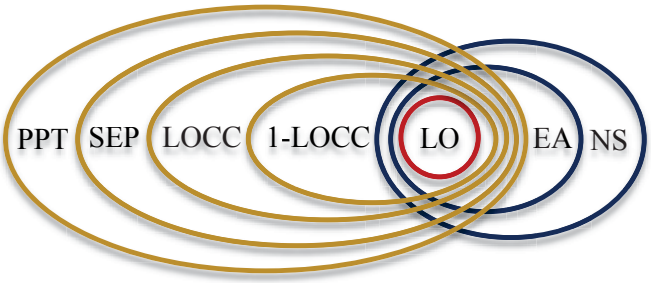


Figure 2.1 : A hierarchy of bipartite quantum operations.

A hierarchy of incoherent quantum operations

A peculiar aspect of the resource theory of coherence is that, although the set of free resource states is defined unambiguously as $\mathcal{I} = \{\rho \geq 0 \mid \Delta(\rho) = \rho\}$, there is no unique physically-motivated choice of allowed free operations.

The *maximally incoherent operations* (MIO) (also known as *incoherence preserving operations*) are defined as any quantum operation $\mathcal{E}$ such that $\mathcal{E}(\mathcal{I}) \subseteq \mathcal{I}$ [Abe06]. This is the largest class of free operations compatible with coherence theory.

A smaller and more relevant class of free operations is called *incoherent operations* (IO) which are characterized as the set of quantum operations admitting a set of Kraus operators

$\{E_i\}$ such that for any E_i [BCP14],

$$\frac{E_i \rho E_i^\dagger}{\operatorname{Tr} E_i \rho E_i^\dagger} \in \mathcal{I}, \quad \forall\, \rho \in \mathcal{I}. \tag{2.21}$$

These Kraus operators E_i are then called *incoherent Kraus operators*. The definition of IO makes sure that in any of the possible outcomes of an IO operation, coherence can never be generated from an incoherent input state, not even probabilistically.

We can be more stringent by adding further desirable properties to the set of free operations. An operation $\mathcal{E}$ is called *strictly incoherent* (SIO) if it can be written in terms of a set of incoherent Kraus operators $\{E_i\}$ such that $E_i^\dagger$ are also incoherent [WY16]. Or equivalently, an operation $\mathcal{E}$ is SIO if and only if it has a set of Kraus operators $\{E_i\}$ such that the operation $\mathcal{E}_i : \rho \to E_i \rho E_i^\dagger$ commute with the completely dephasing channel Δ, i.e., $[\mathcal{E}_i, \Delta] = 0$.

The sets MIO, IO and SIO in general do not have a free dilation, i.e., they cannot be implemented by coupling the system to an environment in an incoherent state followed by a global incoherent unitary. Motivated by this observation, Chitambar and Gour introduced the set of *physical incoherent operations* (PIO) [CG16a, CG16b]. These are all operations which can instead be implemented in the aforementioned way, additionally allowing for incoherent measurements on the environment and classical postprocessing of the measurement outcomes.

Another interesting set is given by the *dephasing-covariant incoherent operations* (DIO), which were introduced independently by Chitambar and Gour [CG16a, CG16b] and Marvian and Spekkens [MS16]. These are all quantum operations $\mathcal{E}$ which commute with the completely dephasing channel, i.e., $\mathcal{E}(\Delta(\rho)) = \Delta(\mathcal{E}(\rho))$ for any quantum state ρ. Or it can be equivalently characterized by

$$\mathcal{E}(|i\rangle\langle i|) \in \mathcal{I}, \quad \forall\, i \quad \text{and} \quad \Delta(\mathcal{E}(|i\rangle\langle j|)) = 0, \quad \forall\, i \neq j. \tag{2.22}$$

Similar to the sets of bipartite operations, there are two strict inclusions for incoherent operations [CG16a, CG16b, dVS17]: PIO $\subsetneq$ SIO $\subsetneq$ IO $\subsetneq$ MIO and PIO $\subsetneq$ SIO $\subsetneq$ DIO $\subsetneq$ MIO. We summarize these relations in Figure 2.2. In the following chapters we will focus more on MIO and DIO.

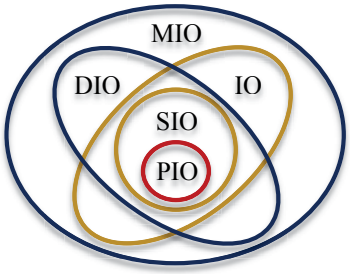


Figure 2.2 : A hierarchy of incoherent quantum operations.

2.1.4 Distance measures

Distance measure between quantum states

There are two most commonly used measures of distance between quantum states, *trace distance* and *fidelity*. Both of them can be generalized to subnormalized quantum states as follows [Tom16].

Definition 2.4 *The generalized trace distance between two subnormalized quantum states ρ and σ is defined as*

$$D(\rho,\sigma) := \frac{1}{2}\|\rho-\sigma\|_1 + \frac{1}{2}|\operatorname{Tr}\rho - \operatorname{Tr}\sigma|. \tag{2.23}$$

The trace distance has a physical interpretation as the distinguishing advantage between the two states. That is, the maximal probability of correctly distinguishing between two uniformly distributed quantum states ρ and σ is given by $\frac{1}{2}(1+D(\rho,\sigma))$.

Definition 2.5 *The generalized fidelity between two subnormalized quantum states ρ and σ is defined as*

$$F(\rho,\sigma) := \|\rho^{1/2}\sigma^{1/2}\|_1 + \sqrt{(1-\operatorname{Tr}\rho)(1-\operatorname{Tr}\sigma)}. \tag{2.24}$$

Proposition 2.6 (Uhlmann's theorem [Uhl76]) *For any two subnormalized quantum states ρ and σ, let $|\varphi\rangle$ be a purification of ρ, then there exists a purification $|\theta\rangle$ of σ such that*

$$F(\rho,\sigma) = F(\varphi,\theta). \tag{2.25}$$

Based on the generalized fidelity, we can define the purified distance.

Definition 2.7 *The purified distance between two subnormalized quantum states ρ and σ is defined as*

$$P(\rho,\sigma) := \sqrt{1-F^2(\rho,\sigma)}. \tag{2.26}$$

The purified distance has simple upper and lower bounds in terms of the generalized trace distance.

Proposition 2.8 *For any two subnormalized quantum states ρ and σ, it holds [Tom16]*

$$D(\rho,\sigma) \leq P(\rho,\sigma) \leq \sqrt{2D(\rho,\sigma)}. \tag{2.27}$$

Distance measure between quantum channels

We use a norm on the set of quantum channels which measures the bias in distinguishing two such channels. The norm is known as the *diamond norm* in quantum information theory [Kit97]. Here, we present it in a formulation which highlights that it is dual to the well-known *completely bounded (cb) norm* [Pau02].

Definition 2.9 *For any linear map $\mathcal{E}: \mathcal{L}(A) \to \mathcal{L}(B)$, its diamond norm is defined as*

$$\|\mathcal{E}\|_{\diamond} := \sup_{k\in\mathbb{N}} \|\mathrm{id}_k \otimes \mathcal{E}\|_1, \tag{2.28}$$

where id_k is the identity map with dimension k and $\|\mathcal{F}\|_1 = \sup_\sigma \|\mathcal{F}(\sigma)\|_1$ with $\sigma_A \in \mathcal{S}_{\leq}(A)$.

The supremum in (2.28) is obtained by $k = d_A$ [Kit97, Pau02]. Based on the diamond norm, the distance between two quantum channels $\mathcal{E}_1$ and $\mathcal{E}_2$ is usually defined as $\frac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_{\diamond}$. Then the minimum error probability to distinguish these two channels is given by $\frac{1}{2}\left(1 - \frac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_{\diamond}\right)$ [BS10].

Since the definition of channel distance involves the optimization over all input states, the following technique can help to fix the input state when considering channels with symmetry.

Proposition 2.10 (Post-selection technique [CKR09]) *Let $\varepsilon > 0$ and $\mathcal{E}_A^n$ and $\mathcal{F}_A^n$ be CPTP maps from $\mathcal{L}(A^{\otimes n})$ to $\mathcal{L}(B)$. If there exists a CPTP map K_π for any permutation π such that $(\mathcal{E}_A^n - \mathcal{F}_A^n)\circ\pi = K_\pi \circ (\mathcal{E}_A^n - \mathcal{F}_A^n)$, then $\|\mathcal{E}_A^n - \mathcal{F}_A^n\|_{\diamond} \leq \varepsilon$ whenever*

$$\|(\mathcal{E}_A^n - \mathcal{F}_A^n) \otimes \mathrm{id}_{RR'}(\omega_{ARR'}^n)\|_1 \leq \varepsilon (n+1)^{-\left(|A|^2-1\right)}, \tag{2.29}$$

where $\omega_{ARR'}^n$ is a purification of the de Finetti state $\omega_{AR}^n = \int \sigma_{AR}^{\otimes n} d(\sigma_{AR})$ with $\sigma_{AR} = |\sigma\rangle\langle\sigma|_{AR} \in \mathcal{S}_{=}(AR)$, $\mathcal{H}_A \cong \mathcal{H}_R$ and $d(\cdot)$ is the measure on the normalized pure states on $\mathcal{H}_A \otimes \mathcal{H}_R$ induced by the Haar measure on the unitary group acting on $\mathcal{H}_A \otimes \mathcal{H}_R$, normalized to $\int d(\cdot) = 1$. Furthermore we can assume without loss of generality that $|R'| \leq (n+1)^{|A|^2-1}$.

The de Finetti state can be written as convex combinations of finite number of i.i.d states.

Lemma 2.11 ([BCR11]) *Let* $\omega_{AR}^n = \int \sigma_{AR}^{\otimes n} d(\sigma_{AR})$ *as in Proposition 2.10. Then*

$$\omega_{AR}^n = \sum_i p_i \left(\sigma_{AR}^i\right)^{\otimes n}, \tag{2.30}$$

with $\sigma_{AR}^i = |\sigma^i\rangle\langle\sigma^i|_{AR} \in \mathcal{S}_=(AR)$, $i \in \left\{1, 2, \cdots, (n+1)^{2|A||R|-2}\right\}$ *and* p_i *a probability distribution.*

2.2 Quantum entropies

Quantum relative entropy and its child entropies

As with many other objects in quantum information theory, *quantum entropies* are defined by extending the classical definitions from probability distributions to density matrices. Let ρ be a quantum state. The von Neumann entropy of ρ, which is the quantum mechanical analog of the Shannon entropy [Sha48], is given by

$$H(\rho) := -\operatorname{Tr}\rho\log\rho. \tag{2.31}$$

Definition 2.12 *For any two quantum states* ρ *and* σ*, if* $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$*, the quantum relative entropy of* ρ *with respect to* σ *is defined by*

$$D(\rho\|\sigma) := \operatorname{Tr}\rho(\log\rho - \log\sigma). \tag{2.32}$$

Otherwise we define $D(\rho\|\sigma) = +\infty$.

Note that the quantum relative entropy is not symmetric, i.e., $D(\rho\|\sigma) = D(\sigma\|\rho)$ does not hold in general. The von Neumann entropy can be recovered by $H(\rho) = -D(\rho\|\mathbb{1})$. The quantum relative entropy also acts as a parent quantity for other entropies:

$$\text{Conditional entropy:} \quad H(A|B)_\rho := -D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B), \tag{2.33}$$

$$\text{Coherent information:} \quad I(A\rangle B)_\rho := D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B), \tag{2.34}$$

$$\text{Mutual information:} \quad I(A:B)_\rho := D(\rho_{AB}\|\rho_A \otimes \rho_B), \tag{2.35}$$

$$\text{Quantum information variance:} \quad V(\rho\|\sigma) := \operatorname{Tr}\rho(\log\rho - \log\sigma)^2 - D(\rho\|\sigma)^2, \tag{2.36}$$

$$\text{Coherent information variance:} \quad V(A\rangle B)_\rho := V(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B). \tag{2.37}$$

The conditional entropy is continuous with respect to the input state.

Lemma 2.13 (Alicki-Fannes Inequality [AF04]) *For any quantum states* ρ_{AB} *and* σ_{AB} *with* $\|\rho_{AB} - \sigma_{AB}\|_1 \leq \varepsilon \leq 1$*, it holds*

$$|H(A|B)_\rho - H(A|B)_\sigma| \leq 4\varepsilon\log|A| + 2h_2(\varepsilon), \tag{2.38}$$

where $h_2(\varepsilon) := -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ *is the binary entropy.*

The quantum mutual information can be written in multiple ways in terms of the quantum relative entropy due to the following result.

Proposition 2.14 ([BD10b]) *For any quantum state ρ_{AB}, it holds*

$$I(A:B)_\rho = \min_{\sigma_B \in \mathcal{S}_=(B)} D(\rho_{AB} \| \rho_A \otimes \sigma_B) = \min_{\substack{\sigma_A \in \mathcal{S}_=(A) \\ \sigma_B \in \mathcal{S}_=(B)}} D(\rho_{AB} \| \sigma_A \otimes \sigma_B). \tag{2.39}$$

Min- and Max-relative entropy and their child entropies

Definition 2.15 ([DKF+14]) *The min-relative entropy of a subnormalized quantum state ρ with respect to $\sigma \geq 0$ is defined as*

$$D_{\min}(\rho \| \sigma) := -\log \| \rho^{1/2} \sigma^{1/2} \|_1^2. \tag{2.40}$$

Definition 2.16 ([Dat09b, Ren05]) *The max-relative entropy of a subnormalized quantum state ρ with respect to $\sigma \geq 0$ is defined as*

$$D_{\max}(\rho \| \sigma) := \inf \left\{ t \geq 0 \mid \rho \leq 2^t \cdot \sigma \right\}, \tag{2.41}$$

if $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$. *Otherwise,* $D_{\max}(\rho \| \sigma) = +\infty$.

Note that Eq. (2.41) can be equivalently given by $D_{\max}(\rho \| \sigma) = \log \| \sigma^{-1/2} \rho \sigma^{-1/2} \|_\infty$.

The *conditional min-* and *max-entropy* are defined as

$$H_{\min}(A|B)_\rho := -\min_{\sigma_B \in \mathcal{S}_\leq(B)} D_{\max}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B), \tag{2.42}$$

$$H_{\max}(A|B)_\rho := -\min_{\sigma_B \in \mathcal{S}_\leq(B)} D_{\min}(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B). \tag{2.43}$$

The above entropies can be extended to corresponding smoothed version, i.e., optimizations of the underlying entropies over a ball of states close to the states under consideration. This closeness is commonly measured in terms of the purified distance. Specifically, we say $\widetilde{\rho}$ is ε-close to ρ, denoted as $\widetilde{\rho} \approx^\varepsilon \rho$, if $P(\widetilde{\rho}, \rho) \leq \varepsilon$. The *smoothed* (conditional) min-/max-relative entropy are respectively defined as follows:

$$D^\varepsilon_{\min}(\rho \| \sigma) := \max_{\widetilde{\rho} \approx^\varepsilon \rho} D_{\min}(\widetilde{\rho} \| \sigma), \quad H^\varepsilon_{\min}(A|B)_\rho := \max_{\widetilde{\rho} \approx^\varepsilon \rho} H_{\min}(A|B)_{\widetilde{\rho}} \tag{2.44}$$

$$D^\varepsilon_{\max}(\rho \| \sigma) := \min_{\widetilde{\rho} \approx^\varepsilon \rho} D_{\max}(\widetilde{\rho} \| \sigma), \quad H^\varepsilon_{\max}(A|B)_\rho := \min_{\widetilde{\rho} \approx^\varepsilon \rho} H_{\max}(A|B)_{\widetilde{\rho}}. \tag{2.45}$$

It is known that the minimization of max-relative entropy is closely related to another resource measure, *log robustness* $LR_g(\rho) := \log(1 + R_g(\rho))$ where

$$R_g(\rho) := \min \left\{ t \geq 0 \;\middle|\; \frac{\rho + t \cdot \omega}{1+t} \in \mathcal{G}, \omega \in \mathcal{S}_=(A) \right\} \tag{2.46}$$

and $\mathcal{G}$ is a given set of (useless) operators. As the name suggests, log robustness quantifies the minimal mixing required to make the resource useless. A geometric interpretation of the robustness $R_g(\rho)$ can be illustrated in Figure 2.3.

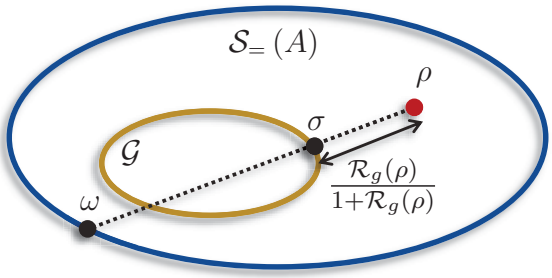


Figure 2.3 : Geometric interpretation of the robustness defined in Eq. (2.46).

Proposition 2.17 ([Dat09a]) *For any quantum state ρ and a set of operators $\mathcal{G}$, it holds*

$$LR_g(\rho) = \min_{\sigma\in\mathcal{G}} D_{\max}(\rho\|\sigma). \tag{2.47}$$

The max-relative entropy is usually related with the cost of a resource. In chapter 5 we will generalize the max-relative entropy as well as the log robustness to quantum channels and provide their nature operational interpretations with respect to the channel simulation cost.

Sandwiched Rényi relative entropy

There are several quantum generalizations of the family of Rényi entropy that contain the quantum relative entropy, min- and max-relative entropies as special cases [MLDS+13, AD15]. Here we introduce the Sandwiched version.

Definition 2.18 *Let ρ and σ be two subnormalized quantum states. For any $\alpha \in (0,1)\cup(1,\infty)$, the order-$\alpha$ sandwiched Rényi relative entropy is defined as*

$$\widetilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log \mathrm{Tr}\left(\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right), \tag{2.48}$$

if $\mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma)$ *and it is equal to* $+\infty$ *otherwise.*

Figure 2.4 : Relations between the sandwiched Rényi relative entropies.

The Sandwiched Rényi relative entropy is monotonically increasing with respect to α, i.e., $\widetilde{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\beta(\rho\|\sigma)$ if $\alpha \geq \beta$. We also have the following reduction [MLDS+13]:

$$D(\rho\|\sigma) = \lim_{\alpha\to 1^-} \widetilde{D}_\alpha(\rho\|\sigma) = \lim_{\alpha\to 1^+} \widetilde{D}_\alpha(\rho\|\sigma), \tag{2.49}$$

$$D_{\min}(\rho\|\sigma) = \widetilde{D}_{1/2}(\rho\|\sigma), \tag{2.50}$$

$$D_{\max}(\rho\|\sigma) = \lim_{\alpha\to\infty} \widetilde{D}_\alpha(\rho\|\sigma). \tag{2.51}$$

Max-information

Note that Proposition 2.14 for quantum mutual information does not hold in general for other entropy measures. In fact, if we replace the quantum relative entropy with the max-relative entropy, the values of the expressions in Proposition 2.14 can lie arbitrarily far apart. Thus we have several definitions for max-information as follows.

Definition 2.19 ([CBR14]) *For any quantum state ρ_{AB}, its max-informations that B has about A are defined as*

$$^1I_{\max}(A:B)_\rho := D_{\max}(\rho_{AB}\|\rho_A\otimes\rho_B), \tag{2.52}$$

$$^2I_{\max}(A:B)_\rho := \min_{\sigma_B\in\mathcal{S}_=(B)} D_{\max}(\rho_{AB}\|\rho_A\otimes\sigma_B), \tag{2.53}$$

$$^3I_{\max}(A:B)_\rho := \min_{\substack{\sigma_A\in\mathcal{S}_=(A)\\ \sigma_B\in\mathcal{S}_=(B)}} D_{\max}(\rho_{AB}\|\sigma_A\otimes\sigma_B). \tag{2.54}$$

The corresponding smoothed versions are defined by

$$^iI^\varepsilon_{\max}(A:B)_\rho := \min_{\widetilde{\rho}\approx^\varepsilon\rho} {}^iI_{\max}(A:B)_{\widetilde{\rho}} \quad \text{for } i\in\{1,2,3\}. \tag{2.55}$$

Unlike the non-smoothed cases, the smoothed max-information are equivalent up to logarithmic terms in the smoothing parameters.

Proposition 2.20 ([CBR14]) *For any quantum state ρ_{AB} and $\varepsilon > 0$, $\varepsilon' \geq 0$, it holds*

$$^3I^{\varepsilon+\varepsilon'}_{\max}(A:B)_\rho \leq {}^2I^{\varepsilon+\varepsilon'}_{\max}(A:B)_\rho \leq {}^3I^{\varepsilon'}_{\max}(A:B)_\rho + f(\varepsilon,\varepsilon'), \quad \textit{and} \tag{2.56}$$

$$^2I^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'}_{\max}(A:B)_\rho \leq {}^1I^{\varepsilon+2\sqrt{\varepsilon}+\varepsilon'}_{\max}(A:B)_\rho \leq {}^2I^{\varepsilon'}_{\max}(A:B)_\rho + g(\varepsilon), \tag{2.57}$$

where

$$f(\varepsilon,\varepsilon') := \log\left(\frac{1}{1-\sqrt{1-\varepsilon^2}} + \frac{1}{1-\varepsilon'}\right),\ g(\varepsilon) := \log\left(\frac{2(1-\varepsilon)+3}{(1-\varepsilon)\left(1-\sqrt{1-\varepsilon^2}\right)}\right). \tag{2.58}$$

There are some basic properties of the max-information that will be used later.

Lemma 2.21 ([BCR11]) *Let $\varepsilon \geq 0$ and $\rho_{AB} = \sum_{i\in I} p_i \rho^i_{AB}$ with $\rho^i_{AB} \in \mathcal{S}_{\leq}(AB)\, \forall\, i \in I$, p_i a probability distribution. Then*

$$ {}^2I^{\varepsilon}_{\max}(A:B)_{\rho} \leq \max_{i\in I} {}^2I^{\varepsilon}_{\max}(A:B)_{\rho^i} + \log|I|. \tag{2.59}$$

Lemma 2.22 ([BCR11]) *Let $\varepsilon \geq 0$ and $\rho_{ABC} \in \mathcal{S}_{=}(ABC)$. Then*

$$ {}^2I^{\varepsilon}_{\max}(A:BC)_{\rho} \leq {}^2I^{\varepsilon}_{\max}(A:B)_{\rho} + 2\cdot\log|C|. \tag{2.60}$$

Quantum hypothesis testing

Quantum hypothesis testing is the task of distinguishing two possible states of a system, ρ_0 and ρ_1. We study two hypotheses, the null hypothesis H_0: the state is ρ_0; the alternative hypothesis H_1: the state is ρ_1. Then we can perform a measurement presented by the POVM $\{M, \mathbb{1} - M\}$ with corresponding classical outcome 0 and 1. If the outcome is 0, we accept the null hypothesis. Otherwise, we accept the alternative one. Analog to the classical case, there are two kinds of errors in this test as shown in Figure 2.5. It is called *Type-I* error if the original state is ρ_0 but the measurement outcome is 1. It is called *Type-II* error if the original state is ρ_1 but the measurement outcome is 0. The probabilities of type-I and type-II error are given by $\mathrm{Tr}\,(\mathbb{1} - M)\,\rho_0$ and $\mathrm{Tr}\, M\rho_1$, respectively. Hypothesis testing relative entropy considers minimizing the type-II error while keeping type-I error within a given error tolerance.

Definition 2.23 *For any two quantum states* [5] *ρ_0 and ρ_1, the quantum hypothesis testing relative entropy is defined as*

$$ D^{\varepsilon}_H(\rho_0||\rho_1) := -\log \beta_{\varepsilon}(\rho_0||\rho_1), \quad \textit{where} \tag{2.61}$$

$$ \beta_{\varepsilon}(\rho_0||\rho_1) := \min\left\{ \mathrm{Tr}\, M\rho_1 \,\middle|\, 0 \leq M \leq \mathbb{1}, 1 - \mathrm{Tr}\, M\rho_0 \leq \varepsilon \right\}. \tag{2.62}$$

Quantum hypothesis testing is an important task with many applications in quantum information theory [Hay17], and it will reduce to the classical case if ρ_0 and ρ_1 commute.

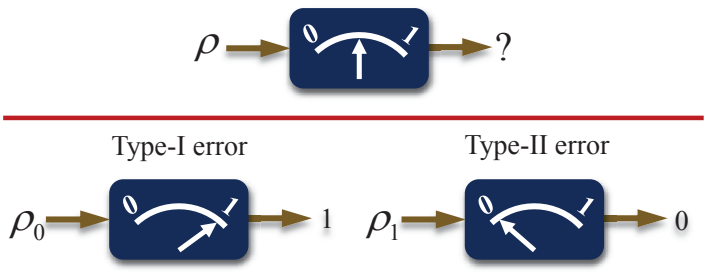


Figure 2.5 : The task of quantum hypothesis testing.

[5] We can generalize the definition to any Hermitian operators. It is still well-defined in the sense of a convex optimization problem.

Asymptotic equipartition property (AEP) and second-order asymptotics

Quantum relative entropy can be recovered in the independent and identically distributed (i.i.d.) asymptotic limit of the smoothed entropies [HP91, ON00, DMHB13, Dat09b]: for any $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} D_H^\varepsilon \left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) = D\left(\rho\|\sigma\right), \tag{2.63}$$

$$\lim_{n\to\infty} \frac{1}{n} D_{\min}^\varepsilon \left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) = D\left(\rho\|\sigma\right), \tag{2.64}$$

$$\lim_{n\to\infty} \frac{1}{n} D_{\max}^\varepsilon \left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) = D\left(\rho\|\sigma\right). \tag{2.65}$$

The conditional entropy of a bipartite state can be recovered via the smoothed conditional min-/max-entropy in the i.i.d. asymptotic limit [TCR09, Tom12], for any $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} H_{\max}^\varepsilon \left(A^n : B^n\right)_{\rho^{\otimes n}} = H\left(A|B\right)_\rho, \tag{2.66}$$

$$\lim_{n\to\infty} \frac{1}{n} H_{\min}^\varepsilon \left(A^n : B^n\right)_{\rho^{\otimes n}} = H\left(A|B\right)_\rho. \tag{2.67}$$

The quantum mutual information of a bipartite state can be recovered by its smoothed max-information in the i.i.d. asymptotic limit [BCR11],

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{1}{n}\left[{}^i I_{\max}^\varepsilon \left(A:B\right)_{\rho^{\otimes n}}\right] = I\left(A:B\right)_\rho, \quad \text{for } i \in \{1,2,3\}. \tag{2.68}$$

The r.h.s. of above limits are called the *first-order asymptotics*. A more accurate estimation can be done via the *second-order* expansion [TH13, Li14]:

$$D_H^\varepsilon \left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) = nD\left(\rho\|\sigma\right) + \sqrt{nV\left(\rho\|\sigma\right)}\,\Phi^{-1}\left(\varepsilon\right) + O\left(\log n\right), \tag{2.69}$$

$$D_{\min}^\varepsilon \left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) = nD\left(\rho\|\sigma\right) + \sqrt{nV\left(\rho\|\sigma\right)}\,\Phi^{-1}\left(\varepsilon\right) + O\left(\log n\right), \tag{2.70}$$

$$D_{\max}^\varepsilon \left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) = nD\left(\rho\|\sigma\right) - \sqrt{nV\left(\rho\|\sigma\right)}\,\Phi^{-1}\left(\varepsilon^2\right) + O\left(\log n\right), \tag{2.71}$$

where Φ^{-1} is the inverse function of cumulative distribution of a standard normal random variable. The *second-order term* is given by the quantum information variance $V\left(\rho\|\sigma\right)$.

Note that recent progress can show a more general result of the asymptotic limits for conditional min/max-entropy beyond the i.i.d. setting [DFR16].

2.3 Semidefinite programming

2.3.1 Basics of semidefinite programming

Semidefinite programming (SDP) is a subfield of convex optimization concerned with the optimization of a linear objective function (a user-specified function that the user wants to minimize

or maximize) over the intersection of the cone of positive semidefinite matrices with an affine space. Semidefinite programming unifies several standard problems (e.g., linear and quadratic programming) and finds applications in various fields. Although semidefinite programs are much more general than linear programs, they are just as easy to solve. Most *interior-point methods* for linear programming have been generalized to semidefinite programs. As in linear programming, these methods have polynomial worst-case complexity, and perform very well in practice. Please refer to a review paper [VB96] for the theory and applications of semidefinite programs and an introduction to primal-dual interior-point methods for their solution.

In this subsection we introduce some basics of semidefinite programming and we refer to John Watrous' lecture note [Wat11] and book [Wat17] for more details.

Definition 2.24 *A semidefinite program is a triple* $(\mathcal{E}, C, D)$*, where* $C \in \mathcal{H}erm(A)$*,* $D \in \mathcal{H}erm(B)$ *and* $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ *is a Hermiticity-preserving linear map, i.e., it maps any Hermitian operators to Hermitian operators.*

Associated with the triple $(\mathcal{E}, C, D)$, there are two optimization problems as follows:

Primal problem		**Dual problem**	
maximize: $\operatorname{Tr} CX$	(2.72a)	minimize: $\operatorname{Tr} DY$	(2.73a)
subject to: $\mathcal{E}(X) = D,$	(2.72b)	subject to: $\mathcal{E}^*(Y) \geq C,$	(2.73b)
$X \geq 0.$	(2.72c)	$Y \in \mathcal{H}erm(B).$	(2.73c)

where $\mathcal{E}^*$ is the dual map by $\operatorname{Tr} \mathcal{E}^*(Y) \cdot X = \operatorname{Tr} Y \cdot \mathcal{E}(X)$ for any X, Y. Note that in practice the optimization may not be explicitly written as the standard form above. We may also write the primal problem as minimization and the dual problem as maximization. The steps of obtaining the dual problem are usually omitted for simplicity.

Any operator X satisfying $\mathcal{E}(X) = D, X \geq 0$ is called *primal feasible* while any operator Y satisfying $\mathcal{E}^*(Y) \geq C, Y \in \mathcal{H}erm(B)$ is called *dual feasible*. Furthermore, a primal problem is *strictly feasible* if there is a primal feasible solution such that $X > 0$. A dual problem is called strictly feasible if there is a dual feasible solution such that $\mathcal{E}^*(Y) > C$. Denote α and β as the *optimal value* of the primal and dual problem respectively. Let $\alpha = -\infty$ and $\beta = +\infty$ when there exist no feasible solutions. A feasible solution achieving the optimal value is then called *optimal solution*.

Semidefinite programs have associated with them a notion of duality, which refers to the special relationship between the primal and dual problems. The property of *weak duality*, which holds for all semidefinite programs, is that the primal optimum can never exceed the dual optimum.

Proposition 2.25 (Weak duality) *For every semidefinite program $(\mathcal{E}, C, D)$, it holds $\alpha \leq \beta$.*

Proof For every primal and dual feasible solution X and Y, we have

$$\operatorname{Tr} CX \leq \operatorname{Tr} \mathcal{E}^* (Y) X = \operatorname{Tr} Y \mathcal{E} (X) = \operatorname{Tr} DY. \tag{2.74}$$

Optimizing overall all feasible solutions, we have $\alpha \leq \beta$. ■

The condition that $\alpha = \beta$ is called *strong duality*, which does not hold in general. However, it does typically hold for semidefinite programs that arise in practice. There are various conditions allowing for verification of strong duality. The following proposition provides one set of conditions under which strong duality is guaranteed.

Proposition 2.26 (Slater's theorem) *For any semidefinite program, the strong duality holds if one of the following conditions satisfied:*

- *the primal problem is feasible and the dual problem is strictly feasible;*
- *the dual problem is feasible and the primal problem is strictly feasible.*

2.3.2 SDP duality proof techniques

In the following chapters, we will frequently need to prove the optimal value of an SDP or show the equivalence of two SDPs. These proofs can be sketched as standard processes due to the SDP duality, as shown in the following two Tables.

- Problem: prove the optimal value of an SDP $(\mathcal{E}, C, D)$ is x.
- Steps:
  1. Check strong duality holds, thus $\alpha = \beta$.
  2. Construct a primal feasible solution X, such that $\operatorname{Tr} CX = x$. This implies $\alpha \geq x$.
  3. Construct a dual feasible solution Y, such that $\operatorname{Tr} DY = x$. This implies $\beta \leq x$.
  4. Combining above three steps, we have $\alpha = \beta = x$.

Table 2.1 : Steps to prove the optimal value of an SDP.

- Problem: show the optimal value of two SDPs $(\mathcal{E}_1, C_1, D_1)$ and $(\mathcal{E}_2, C_2, D_2)$ are equal.
- Steps:
  1. Check strong duality holds, thus $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$.
  2. Suppose the primal optimal solution of SDP $(\mathcal{E}_1, C_1, D_1)$ is given by X_1. Construct a primal feasible solution X_2 (usually based on X_1) for $(\mathcal{E}_2, C_2, D_2)$ such that $\operatorname{Tr} C_2 X_2 \geq \alpha_1$. This implies $\alpha_1 \leq \alpha_2$.

     Or suppose the dual optimal solution of SDP $(\mathcal{E}_2, C_2, D_2)$ is given by Y_2. Construct a dual feasible solution Y_1 (ususlly based on Y_2) for $(\mathcal{E}_1, C_1, D_1)$ such that $\operatorname{Tr} D_1 Y_1 \leq \beta_2$. This implies $\beta_1 \leq \beta_2$.
  3. Swap the role of $(\mathcal{E}_1, C_1, D_1)$ and $(\mathcal{E}_2, C_2, D_2)$ and repeat the second step. We can show that $\alpha_1 \geq \alpha_2$ or $\beta_1 \geq \beta_2$.
  4. Combining above three steps, we have $\alpha_1 = \beta_1 = \alpha_2 = \beta_2$.

Table 2.2 : Steps to prove the optimal values of two SDPs are equal.

2.3.3 A list of frequently used SDPs

The advantages of using semidefinite programming are multifold. First, it can be efficiently computed [6] via interior point methods. With the assistance of various toolboxes (e.g. CVX, YALMIP) as well as solvers (e.g. SDPT3, SeDuMi, Mosek) we can easily perform numerical experiments and make interesting observations. Second, the SDP duality technique can help us simplify the proofs and unify them into standard processes as shown above. However, the difficulty lies in constructing suitable feasible solutions, which on the other hand can be assisted via numerical observations. Finally, SDP provides us a new perspective to study some fundamental quantities as we will see in the following chapters.

It is worth mentioning that not all problems can be reformulated via SDPs and that most SDP reformulations can be rather tricky. In the following, we list some of them that will be

[6] Note that many one-shot quantities, including those will be introduced in this thesis, are characterized via SDPs and thus claimed to be efficiently computable. There is nothing wrong with such claims. But it is easily being misunderstood that those quantities are also efficiently computable for n-shot of resources. Since tensoring exponentially increases the size of SDP, we cannot guarantee its efficiency for the n-shot case.

used in the following chapters.

For any Hermitian operator X, its trace norm $\|X\|_1$ is given by [Wat11]

Primal		**Dual**	
maximize: $\operatorname{Tr} MX$	(2.75a)	minimize: $\operatorname{Tr} X_1 + X_2$	(2.76a)
subject to: $-\mathbb{1} \leq M \leq \mathbb{1},$	(2.75b)	subject to: $X = X_1 - X_2,$	(2.76b)
$M \in \mathcal{H}erm$	(2.75c)	$X_1 \geq 0, X_2 \geq 0.$	(2.76c)

Its infinity norm $\|X\|_\infty$ is given by [Wat11]

Primal		**Dual**	
maximize: $\operatorname{Tr} MX$	(2.77a)	minimize: t	(2.78a)
subject to: $\operatorname{Tr} M \leq 1,$	(2.77b)	subject to: $-t\mathbb{1} \leq X \leq t\mathbb{1},$	(2.78b)
$M \in \mathcal{H}erm.$	(2.77c)	$t \geq 0.$	(2.78c)

For any quantum states ρ and σ, their fidelity $F(\rho, \sigma)$ can be given by [Wat11]

Primal		**Dual**	
maximize: $\frac{1}{2}\operatorname{Tr}\left(X + X^\dagger\right)$	(2.79a)	minimize: $\frac{1}{2}\operatorname{Tr}(\rho Y + \sigma Z)$	(2.80a)
subject to: $\begin{pmatrix} \rho & X \\ X^\dagger & \sigma \end{pmatrix} \geq 0.$	(2.79b)	subject to: $\begin{pmatrix} Y & -\mathbb{1} \\ -\mathbb{1} & Z \end{pmatrix} \geq 0.$	(2.80b)

For any two quantum channels $\mathcal{N}_1, \mathcal{N}_2$ from $\mathcal{L}(A)$ to $\mathcal{L}(B)$, with Choi-Jamiołkowski operators $J_{\mathcal{N}_1}$ and $J_{\mathcal{N}_2}$ respectively, the diamond norm of their difference, i.e., $\frac{1}{2}\|\mathcal{N}_1 - \mathcal{N}_2\|_\diamond$ can be expressed as an SDP of the form [Wat09],

Primal		**Dual**	
maximize: $\operatorname{Tr}(J_{\mathcal{N}_1} - J_{\mathcal{N}_2}) W_{AB}$	(2.81a)	minimize: λ	(2.82a)
subject to: $\rho_A \geq 0,$	(2.81b)	subject to: $\operatorname{Tr}_B Y_{AB} \leq \lambda \mathbb{1}_A,$	(2.82b)
$\operatorname{Tr} \rho_A = 1,$	(2.81c)	$Y_{AB} \geq J_{\mathcal{N}_1} - J_{\mathcal{N}_2},$	(2.82c)
$0 \leq W_{AB} \leq \rho_A \otimes \mathbb{1}_B$	(2.81d)	$Y_{AB} \geq 0.$	(2.82d)

The max-relative entropy $D_{\max}(\rho\|\sigma)$ can be given by

	Primal			**Dual**	
log	maximize: $\operatorname{Tr}\rho X$	(2.83a)	log	minimize: t	(2.84a)
	subject to: $\operatorname{Tr}\sigma X \leq 1,$	(2.83b)		subject to: $\rho \leq t\sigma,$	(2.84b)
	$X \geq 0.$	(2.83c)		$t \geq 0.$	(2.84c)

The quantum hypothesis testing relative entropy $D_H^{\varepsilon}(\rho\|\sigma)$ can be given by

	Primal			**Dual**	
$-\log$	maximize: $\operatorname{Tr} X + (1-\varepsilon)\, x$	(2.85a)	$-\log$	minimize: $\operatorname{Tr} M\sigma$	(2.86a)
	subject to: $\sigma - X - x\rho \geq 0,$	(2.85b)		subject to: $0 \leq M \leq \mathbb{1},$	(2.86b)
	$X \leq 0, x \geq 0.$	(2.85c)		$\operatorname{Tr} M\rho \geq 1-\varepsilon.$	(2.86c)

Finally we summarize the semidefinite characterizations of a quantum operation in the following two tables.

Operation	Semidefinite conditions
CP	$J_{\mathcal{E}} \geq 0$
TP	$\operatorname{Tr}_B J_{\mathcal{E}} = \mathbb{1}_A$
TNI	$\operatorname{Tr}_B J_{\mathcal{E}} \leq \mathbb{1}_A$
MIO	CP $\wedge$ TP $\wedge \operatorname{Tr}_A J_{\mathcal{E}}\left(\lvert i\rangle\langle i\rvert \otimes \mathbb{1}_B\right) \in \mathcal{I}, \forall\, i$
DIO	MIO $\wedge\, \Delta\left(\operatorname{Tr}_A J_{\mathcal{E}}\left(\lvert i\rangle\langle j\rvert \otimes \mathbb{1}_B\right)\right) = 0, \forall\, i \neq j$

Table 2.3 : Semidefinite conditions for single system quantum operation $\mathcal{E}_{A\to B}$.

Operation	Semidefinite conditions
CP	$J_{\mathcal{E}} \geq 0$
TP	$\operatorname{Tr}_{A_2B_2} J_{\mathcal{E}} = \mathbb{1}_{A_1B_1}$
$A \not\to B$	$\operatorname{Tr}_{A_2} J_{\mathcal{E}} = \mathbb{1}_{A_1}/d_{A_1} \otimes \operatorname{Tr}_{A_1A_2} J_{\mathcal{E}}$
$B \not\to A$	$\operatorname{Tr}_{B_2} J_{\mathcal{E}} = \mathbb{1}_{B_1}/d_{B_1} \otimes \operatorname{Tr}_{B_1B_2} J_{\mathcal{E}}$
NS	CP $\wedge$ TP $\wedge\, A \not\to B \wedge B \not\to A$
PPT	CP $\wedge$ TP $\wedge\, J_{\mathcal{E}}^{T_{B_1B_2}} \geq 0$

Table 2.4 : Semidefinite conditions for bipartite quantum operation $\mathcal{E}_{A_1B_1\to A_2B_2}$.

Chapter 3

Entanglement distillation

Entanglement distillation, an essential quantum information processing task, refers to the conversion from multiple copies of noisy entangled states to a smaller number of highly entangled ones. In this chapter, we study the non-asymptotic fundamental limits for entanglement distillation. We investigate the optimal tradeoff between the distillation rate, the number of prepared states, and the error tolerance.

3.1 Introduction

3.1.1 Background

Quantum entanglement is a striking feature of quantum mechanics and a key ingredient in many quantum information processing tasks, such as the efficient transmission of information via quantum dense coding [BW92, MWKZ96] or quantum teleportation [BBC$^+$93], the security of transmitted data through entanglement-based quantum cryptography [Eke91], including the recent development of device-independent quantum cryptography [ABG$^+$07], as well as both device-independent randomness generation [Col09, PAM$^+$10] or amplification [CR12, GMDLT$^+$13, BRG$^+$16], quantum metrology [EdMFD11, GLM11] and the efficient solution of the factorization problem [Sho94]. All these protocols necessarily rely on entanglement resources, especially the maximally entangled states. It is thus of great importance to study the transformation of less entangled states into more suitable ones such as maximally entangled states. This procedure is known as *entanglement distillation*.

In general, the task of entanglement distillation allows two parties (Alice and Bob) to perform local operations and classical communication (LOCC). The *distillable entanglement* characterizes the rate at which one can asymptotically obtain maximally entangled states from a collection of identically and independently distributed (i.i.d) prepared entangled states by LOCC [BBP$^+$96, Rai99b]. Entanglement distillation from non-i.i.d prepared states has also

been considered recently [WGCE16]. Distillable entanglement is a fundamental entanglement measure which captures the resource character of entanglement. Up to now, it remains unknown how to compute distillable entanglement for general quantum states and various approaches [VP98, Rai99a, VW02, Rai01, HHH00, CW04, LDS17, WD16a] have been developed to evaluate this important quantity.

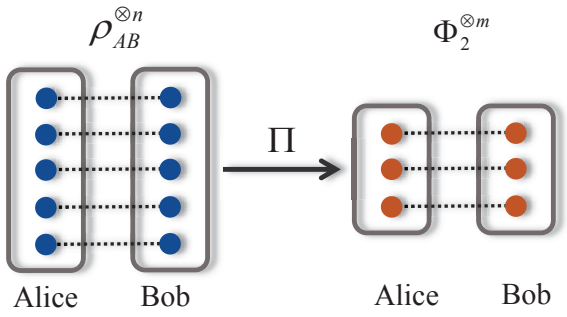


Figure 3.1 : The task of entanglement distillation.

Let Ω represent one of the classes of operation 1-LOCC, LOCC, SEP or PPT. The concise definition of distillable entanglement by the class of operation Ω is given as follows:

$$E_{\mathrm{D},\Omega}\left(\rho_{AB}\right) := \sup\left\{ r \in \mathbb{R}_+ \;\middle|\; \lim_{n\to\infty}\left(\inf_{\Pi\in\Omega}\left\|\Pi\left(\rho_{AB}^{\otimes n}\right) - \Phi_{2^{rn}}\right\|_1\right) = 0 \right\}, \tag{3.1}$$

where $\Phi_k = (1/k)\sum_{i,j=0}^{k-1} |ii\rangle\langle jj|$ is the k-dimensional maximally entangled state. Due to the inclusion relation of the operation classes 1-LOCC $\subsetneq$ LOCC $\subsetneq$ SEP $\subsetneq$ PPT, for any quantum state ρ_{AB} we have the chain of inequalities

$$E_{\mathrm{D,1\text{-}LOCC}}\left(\rho_{AB}\right) \leq E_{\mathrm{D,LOCC}}\left(\rho_{AB}\right) \leq E_{\mathrm{D,SEP}}\left(\rho_{AB}\right) \leq E_{\mathrm{D,PPT}}\left(\rho_{AB}\right). \tag{3.2}$$

To evaluate distillable entanglement efficiently, one possible way is to find computable bounds. Two well-known upper bounds of the distillable entanglement under LOCC as well as PPT operations are given by the *relative entropy of entanglement* (REE) [VPRK97, VP98] and the *PPT-relative entropy of entanglement* [Rai99a],

$$E_{\mathrm{r,SEP}}\left(\rho_{AB}\right) := \min_{\sigma\in\mathrm{SEP}(A:B)} D\left(\rho_{AB}\|\sigma_{AB}\right), \tag{3.3}$$

$$E_{\mathrm{r,PPT}}\left(\rho_{AB}\right) := \min_{\sigma\in\mathrm{PPT}(A:B)} D\left(\rho_{AB}\|\sigma_{AB}\right), \tag{3.4}$$

which express the minimal distinguishability between the given state and all possible separable states or PPT states. An improved bound is the *Rains bound* [Rai01], which is given by

$$R\left(\rho_{AB}\right) := \min_{\sigma\in\mathcal{S}_=(AB)}\left(D\left(\rho_{AB}\|\sigma_{AB}\right) + \log\left\|\sigma^{T_B}\right\|_1\right). \tag{3.5}$$

Note that Eq. (3.5) is not a convex optimization since the second term (logarithmic negativity) is not convex [VW02]. However, the Rains bound can be reformulated [ADVW02] as a convex optimization over the Rains set $\mathrm{PPT}'(A:B) := \left\{\sigma \in \mathcal{P}(AB) \;\middle|\; \|\sigma^{T_B}\|_1 \leq 1\right\}$, that is,

$$R(\rho_{AB}) = \min_{\sigma \in \mathrm{PPT}'(A:B)} D(\rho_{AB} \| \sigma_{AB}). \tag{3.6}$$

This provides the opportunity to numerically calculate the Rains bound, as discussed in Appendix B. Due to the inclusion relations $\mathrm{SEP}(A:B) \subsetneq \mathrm{PPT}(A:B) \subsetneq \mathrm{PPT}'(A:B)$, for any quantum state ρ_{AB}, it holds

$$E_{\mathrm{D,PPT}}(\rho_{AB}) \leq R(\rho_{AB}) \leq E_{\mathrm{r,PPT}}(\rho_{AB}) \leq E_{\mathrm{r,SEP}}(\rho_{AB}). \tag{3.7}$$

The *logarithmic negativity* [VW02, Ple05] is an efficiently computable upper bound on the PPT-assisted distillable entanglement. The best known SDP upper bound is given by [WD16a]

$$E_W(\rho_{AB}) = \log \min \left\{ \left\| X_{AB}^{T_B} \right\|_1 \;\middle|\; X_{AB} \geq \rho_{AB} \right\}, \tag{3.8}$$

which is an improved version of the logarithmic negativity. Other known upper bounds of distillable entanglement are studied in Refs. [VP98, Rai99a, HHH00, CW04]. Most of these known upper bounds are difficult to compute [Hua14] and usually easily computable only for states with high symmetries, such as Werner states, isotropic states, or the family of "iso-Werner" states [BDSW96, VW01, TV00, Rai99a].

In a realistic setting, the resources are finite and the number of i.i.d. prepared states is necessarily limited. More importantly, it is hard to perform joint state manipulations over a very large number of qubits. Therefore, it is important to characterize how many maximally entangled states can be faithfully distilled from finite copies of prepared states. Since the asymptotic rates are insufficient to give a precise estimation, it is necessary to consider second-order characterizations. In particular, for practical use, efficiently computable bounds are more desirable. In the non-asymptotic setting, one also has to make a tradeoff between the distillation rate and the error tolerance of state transformation.

The study of such non-asymptotic scenarios has recently garnered great interest in classical information theory (e.g., [PPV10, Hay09, Tan14]) as well as in quantum information theory (e.g., [WR12, RR11, TH13, BCR11, LM15a, DH13, MW14, BDL16, Tom16, TBR16, WXD18, CH17, CTT17]). Here we study the setting of entanglement distillation.

A non-asymptotic analysis of entanglement distillation will help us better exploit the power of entanglement in a realistic setting. Previously, the one-shot distillable entanglement was studied in Refs. [BD10a, BD11]. But these bounds are not known to be efficiently computable. Thus it is difficult to apply them as experimental benchmarks. These one-shot bounds are also

not suitable to establish second-order estimations. Datta and Leditzky studied the second-order expansion of distillable entanglement under LOCC operations for pure states [DL15]. Here, we go beyond their results by considering more general operations and states. The Rains bound [Rai01] and the hashing bound [DW05] are arguably the best general upper and lower bounds for distillable entanglement, respectively. However, these bounds do not provide sufficiently precise evaluation about entanglement distillation with finite resources.

3.1.2 Outline

In this chapter, we overcome the limitations of previous results and provide efficiently computable estimation for non-asymptotic distillable entanglement. We first build an exact connection between entanglement distillation under positive-partial-transpose-preserving (PPT) operations and quantum hypothesis testing. In particular, the one-shot distillable entanglement under PPT operations can be given by the minimization of hypothesis testing relative entropy between the given state and some class of operators. Based on this connection, we provide efficiently computable second-order estimations of the distillable entanglement for given finite copies of the state and the error tolerance. As applications, we apply our results to study some states of practical interest, including pure states, mixture of Bell states, maximally correlated states and isotropic states.

3.2 One-shot entanglement distillation

Since the distillation process cannot always be accomplished perfectly, we use the fidelity of distillation to characterize the performance of distillation. Then the one-shot distillable entanglement is defined as the maximal number of Bell state (maximally entangled state with local dimension equal to two) we can obtain while keeping the infidelity of the distillation process within a given tolerance.

Definition 3.1 ([Rai01]) *For any bipartite quantum state ρ_{AB}, the fidelity of distillation under the operation class Ω is defined as,*

$$F_{\Omega}\left(\rho_{AB}, k\right) := \max_{\Pi \in \Omega} \operatorname{Tr} \Pi\left(\rho_{AB}\right) \Phi_k, \tag{3.9}$$

where $\Phi_k = (1/k) \sum_{i,j=0}^{k-1} |ii\rangle\langle jj|$ is the k-dimensional maximally entangled state and the maximization is taken over all possible operation Π in the class Ω.

Definition 3.2 *For any bipartite quantum state ρ_{AB}, the one-shot ε-error distillable entanglement under the operation class Ω is defined as*

$$E_{\mathrm{D},\Omega}^{(1),\varepsilon}\left(\rho_{AB}\right) := \log \max \left\{ k \in \mathbb{N} \,\middle|\, F_{\Omega}\left(\rho_{AB}, k\right) \geq 1-\varepsilon \right\}. \tag{3.10}$$

Then the (asymptotic) distillable entanglement is given by the regularization:

$$E_{\mathrm{D},\Omega}\left(\rho_{AB}\right)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}E_{\mathrm{D},\Omega}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right). \tag{3.11}$$

Exploiting the symmetry of maximally entangled state and the semidefinite conditions of PPT operations, we derive the one-shot distillable entanglement under PPT operations as an SDP which is efficiently computable and suitable for a small scale estimation.

Lemma 3.3 *For any bipartite quantum state ρ_{AB} and error tolerance $\varepsilon\in(0,1)$, it holds*

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho\right)=\log\max\left\{\lfloor 1/\eta\rfloor\;\middle|\;0\leq M\leq\mathbb{1},\mathrm{Tr}\,M\rho\geq 1-\varepsilon,-\eta\mathbb{1}\leq M^{T_B}\leq\eta\mathbb{1}\right\}. \tag{3.12}$$

Proof From the definition of one-shot distillable entanglement, we have

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho\right)=\log\max\left\{k\in\mathbb{N}\;\middle|\;\mathrm{Tr}\,\Pi\left(\rho\right)\Phi_k\geq 1-\varepsilon,\Pi\in\mathrm{PPT}\right\}. \tag{3.13}$$

According to the Choi-Jamiołkowski representation of quantum operations, we can represent the output state of operation $\Pi_{AB\to A'B'}$ via its Choi matrix J_Π as

$$\Pi_{AB\to A'B'}\left(\rho_{AB}\right)=\mathrm{Tr}_{AB}\left(J_\Pi\cdot\rho_{AB}^T\otimes\mathbb{1}_{A'B'}\right). \tag{3.14}$$

Then we have

$$\begin{aligned}\mathrm{Tr}\,\Pi_{AB\to A'B'}\left(\rho_{AB}\right)\Phi_k&=\mathrm{Tr}\left(\mathrm{Tr}_{AB}\left(J_\Pi\cdot\rho_{AB}^T\otimes\mathbb{1}_{A'B'}\right)\right)\Phi_k &(3.15)\\&=\mathrm{Tr}\,J_\Pi\cdot\left(\rho_{AB}^T\otimes\mathbb{1}_{A'B'}\right)\left(\mathbb{1}_{AB}\otimes\Phi_k\right) &(3.16)\\&=\mathrm{Tr}\,J_\Pi\cdot\left(\mathbb{1}_{AB}\otimes\Phi_k\right)\left(\rho_{AB}^T\otimes\mathbb{1}_{A'B'}\right) &(3.17)\\&=\mathrm{Tr}\left(\mathrm{Tr}_{A'B'}\,J_\Pi\cdot\left(\mathbb{1}_{AB}\otimes\Phi_k\right)\right)\rho_{AB}^T. &(3.18)\end{aligned}$$

Recall the condition that Π is a PPT operation if and only if its Choi-Jamiołkowski matrix satisfies

$$J_\Pi\geq 0,\quad\mathrm{Tr}_{A'B'}\,J_\Pi=\mathbb{1}_{AB},\quad J_\Pi^{T_{BB'}}\geq 0. \tag{3.19}$$

Combining Eqs. (3.13), (3.18), (3.19) we have

$$\begin{aligned}E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho\right)=\log\max\;&k &(3.20a)\\\text{s.t.}\;&\mathrm{Tr}\left(\mathrm{Tr}_{A'B'}\,J_\Pi\cdot\left(\mathbb{1}_{AB}\otimes\Phi_k\right)\right)\rho_{AB}^T\geq 1-\varepsilon, &(3.20b)\\&J_\Pi\geq 0,\;\mathrm{Tr}_{A'B'}\,J_\Pi=\mathbb{1}_{AB},\;J_\Pi^{T_{BB'}}\geq 0. &(3.20c)\end{aligned}$$

Suppose one optimal solution in optimization (3.20) is given by $\widetilde{J}_\Pi$. Since Φ_k is invariant under any local unitary $U_{A'}\otimes\overline{U}_{B'}$, i.e., $\left(U_{A'}\otimes\overline{U}_{B'}\right)\Phi_k\left(U_{A'}\otimes\overline{U}_{B'}\right)^\dagger=\Phi_k$, we can verify

that $\left(U_{A'} \otimes \overline{U}_{B'}\right) \widetilde{J}_{\Pi} \left(U_{A'} \otimes \overline{U}_{B'}\right)^{\dagger}$ is also optimal. Since any convex combination of optimal solutions remains optimal, we know that

$$\int dU \left(U_{A'} \otimes \overline{U}_{B'}\right) \widetilde{J}_{\Pi} \left(U_{A'} \otimes \overline{U}_{B'}\right)^{\dagger} \tag{3.21}$$

is also optimal, where dU is the Haar measure. According to Schur's lemma, the result of the above integral gives an optimal solution admitting the structure of $W_{AB} \otimes \Phi_k + Q_{AB} \otimes (\mathbb{1} - \Phi_k)$ with certain linear operators W_{AB} and Q_{AB}. Thus without loss of generality, we can take the ansatz of the optimal Choi-Jamiołkowski matrix as

$$J_{\Pi} = W_{AB} \otimes \Phi_k + Q_{AB} \otimes (\mathbb{1} - \Phi_k) . \tag{3.22}$$

Next we take Eq. (3.22) back to the optimization (3.20) and do some further simplification. Denote P_+ and P_- as the symmetric and anti-symmetric projections respectively. From the spectral decomposition, we have $\Phi_k^{T_{B'}} = \frac{1}{k}(P_+ - P_-)$ and

$$\begin{aligned} J_{\Pi}^{T_{BB'}} &= W^{T_B} \otimes \Phi_k^{T_{B'}} + Q^{T_B} \otimes (\mathbb{1} - \Phi_k)^{T_{B'}} && (3.23)\\ &= W^{T_B} \otimes \frac{1}{k}(P_+ - P_-) + Q^{T_B} \otimes \frac{1}{k}\left((k-1)P_+ + (k+1)P_-\right) && (3.24)\\ &= \left(W^{T_B} + (k-1)Q^{T_B}\right) \otimes \frac{1}{k} P_+ + \left(-W^{T_B} + (k+1)Q^{T_B}\right) \otimes \frac{1}{k} P_- . && (3.25) \end{aligned}$$

Since P_+ and P_- are positive and orthogonal, we have $J_{\Pi}^{T_{BB'}} \geq 0$ if and only if

$$\begin{aligned} W^{T_B} + (k-1)Q^{T_B} &\geq 0, && (3.26)\\ -W^{T_B} + (k+1)Q^{T_B} &\geq 0. && (3.27) \end{aligned}$$

Note that $\operatorname{Tr}\left(\operatorname{Tr}_{A'B'} J_{\Pi} \cdot (\mathbb{1}_{AB} \otimes \Phi_k)\right) \rho_{AB}^T = \operatorname{Tr} W \rho_{AB}^T$. Thus we can simplify the optimization (3.20) as

$$\begin{aligned} E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho) = \log \max\ & k && (3.28a)\\ \text{s.t. } & \operatorname{Tr} W\rho_{AB}^T \geq 1-\varepsilon, && (3.28b)\\ & W,\ Q \geq 0, W + \left(k^2-1\right)Q = \mathbb{1}, && (3.28c)\\ & (1-k)Q^{T_B} \leq W^{T_B} \leq (1+k)Q^{T_B}. && (3.28d) \end{aligned}$$

Eliminating the variable Q via the condition $W + \left(k^2-1\right)Q = \mathbb{1}$ and let $M = W^T$, $\eta = \frac{1}{k}$, we obtain the desired result. ■

Remark 3.4 Note that the calculation in Eq. (3.12) can be implemented with two steps:

(1) computing the SDP: $\min\left\{ \eta \;\middle|\; 0 \leq M \leq \mathbb{1}, \operatorname{Tr} M\rho \geq 1-\varepsilon, -\eta\mathbb{1} \leq M^{T_B} \leq \eta\mathbb{1} \right\}$ and obtaining the optimal value η_0;

(2) the optimal value on the r.h.s of Eq. (3.12) is given by $\log\lfloor 1/\eta_0 \rfloor$.

Remark 3.5 It is easy to check that $\delta = \log x - \log\lfloor x \rfloor \in [0,1]$ for any $x \geq 1$. Thus in Eq. (3.12), we can use the least constant $\delta \in [0,1]$ to adjust the r.h.s. to be the logarithm of an integer, i.e,

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho) = -\log\min\left\{ \eta \,\middle|\, 0 \leq M \leq \mathbb{1}, \operatorname{Tr} M\rho \geq 1-\varepsilon, -\eta\mathbb{1} \leq M^{T_B} \leq \eta\mathbb{1} \right\} - \delta. \tag{3.29}$$

This form will help our analysis in the following discussion.

In the next result, we build an exact connection between the one-shot distillable entanglement under PPT operations and the quantum hypothesis testing relative entropy [1].

Theorem 3.6 *For any bipartite quantum state ρ_{AB} and error tolerance $\varepsilon \in (0,1)$, it holds that*

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho_{AB}) = \min_{\substack{\|G^{T_B}\|_1 \leq 1 \\ G=G^\dagger}} D_H^\varepsilon(\rho_{AB}\|G) - \delta, \tag{3.30}$$

where $\delta \in [0,1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

Proof The main idea to build this connection is to use the norm duality between infinity norm and trace norm. From Eq. (3.29), we know that

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho) = -\log\min_{\mathcal{S}_\rho} \left\| M^{T_B} \right\|_\infty - \delta, \tag{3.31}$$

where the set $\mathcal{S}_\rho := \left\{ M \,\middle|\, 0 \leq M \leq \mathbb{1}, \operatorname{Tr} M\rho \geq 1-\varepsilon \right\}$ and $\delta \in [0,1]$. Then using the norm duality $\|M\|_\infty = \max_{\|G\|_1 \leq 1, G=G^\dagger} \operatorname{Tr} MG$, we substitute infinity norm with trace norm and obtain

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho) = -\log\min_{\mathcal{S}_\rho} \max_{\substack{\|G^{T_B}\|_1 \leq 1 \\ G=G^\dagger}} \operatorname{Tr} MG - \delta. \tag{3.32}$$

Since the objective function is linear with respect to the variable M and G, we can apply Sion's minimax theorem [Sio58] and exchange the minimization with the maximization [2],

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho) = -\log \max_{\substack{\|G^{T_B}\|_1 \leq 1 \\ G=G^\dagger}} \min_{\mathcal{S}_\rho} \operatorname{Tr} MG - \delta \tag{3.33}$$

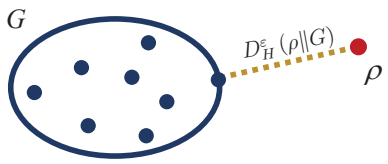
$$= \min_{\substack{\|G^{T_B}\|_1 \leq 1 \\ G=G^\dagger}} -\log\min_{\mathcal{S}_\rho} \operatorname{Tr} MG - \delta. \tag{3.34}$$

[1] Note that the quantum hypothesis testing here only indicates the convex optimization without a clear operational meaning. The second term in the hypothesis testing relative entropy is extended to Hermitian operators.

[2] Note that the result of $\min_{\mathcal{S}_\rho} \operatorname{Tr} MG$ in Eq. (3.34) might be negative. But without loss of generality, we will use the convention $\log x = -\infty$ for $x \leq 0$ throughout this thesis.

Due to the definition of quantum hypothesis testing relative entropy, we have the desired result. ■

This result provides an intuitive interpretation that the one-shot distillable entanglement under PPT operations is given by the "distance" between the given state and a class of useless operators G. Note that Brandão and Datta [BD11] have shown a similar "distance" characterization for (approximately) separability-preserving operations. In terms of coherence theory, Winter and Yang have shown that the distillable coherence under MIO is given by the "distance" between the given state and the class of incoherent states [WY16]. Such "distance" characterization is widely used for various task in quantum information theory while no general result is known. We give a brief summary of the known results in Appendix A.



Remark 3.7 Note that the optimal solution G in the optimization (3.30) is not necessarily positive, as shown in the following example. Here the quantum hypothesis testing relative entropy loses its operational interpretation, but it is still a well-defined quantity as a convex optimization.

Let $\rho_\theta = \frac{3}{4}|\varphi_1\rangle\langle\varphi_1| + \frac{1}{4}|\varphi_2\rangle\langle\varphi_2|$ with $|\varphi_1\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$ and $|\varphi_2\rangle = |10\rangle$. Using the dual SDP for the quantum hypothesis testing relative entropy, the optimal value of $\min_{\|G^{T_B}\|_1\leq 1, G=G^\dagger} D_H^\varepsilon(\rho\|G)$ is equal to

$$-\log\max\left\{\operatorname{Tr}X + t(1-\varepsilon)\,\middle|\, G - X - t\rho \geq 0,\ X \leq 0,\ t \geq 0,\ \left\|G^{T_B}\right\|_1 \leq 1, G = G^\dagger\right\}. \tag{3.35}$$

Without considering the composition of $-\log$, we have the following SDP 1. In Figure 3.2, we show that adding the constraint $G \geq 0$ will change the optimal value of SDP 1. That is, the optimal values of SDP 1 and SDP 2 are different. This tells the optimal solution in Eq. (3.30) is not taken at positive operator G [3].

[3] We implement the SDP 1 and SDP 2 via CVX package, both of which can be solved to a very high (near-machine) precision. The maximal gap in the plot is approximately 1.7×10^{-2}.

SDP 1

$$\begin{aligned} \text{maximize} \quad & \operatorname{Tr} X + t\,(1-\varepsilon) && \text{(3.36a)} \\ \text{subject to} \quad & G - X - t\rho \geq 0, && \text{(3.36b)} \\ & X \leq 0,\ t \geq 0, && \text{(3.36c)} \\ & \|G^{T_B}\|_1 \leq 1,\ G = G^{\dagger}. && \text{(3.36d)} \end{aligned}$$

SDP 2

$$\begin{aligned} \text{maximize} \quad & \operatorname{Tr} X + t\,(1-\varepsilon) && \text{(3.37a)} \\ \text{subject to} \quad & G - X - t\rho \geq 0, && \text{(3.37b)} \\ & X \leq 0,\ t \geq 0, && \text{(3.37c)} \\ & \|G^{T_B}\|_1 \leq 1,\ G \geq 0. && \text{(3.37d)} \end{aligned}$$

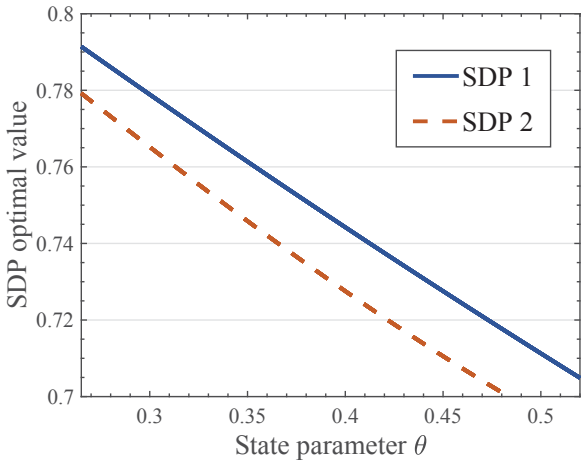


Figure 3.2 : This figure demonstrates the difference of optimal values in SDP 1 (3.36) and SDP 2 (3.37) for the defined quantum state ρ_θ. The solid line depicits the optimal value of SDP 1 while the dashed line depicits the optimal value of SDP 2. The parameter θ ranges from $\pi/12$ to $\pi/6$ and error tolerance is taken at $\varepsilon = 1 - \sqrt{3}/2$.

We stress a bit further the difference between the feasible set in Eq. (3.30) and the Rains set $\mathrm{PPT}'(A:B) = \{\sigma \geq 0 \mid \|\sigma^{T_B}\|_1 \leq 1\}$. If we can show that taking positive solution G will not compromise too much of the distillable entanglement, i.e.,

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) \approx \min_{G\geq 0,\, \|G^{T_B}\|_1 \leq 1} D_H^{\varepsilon}\left(\rho_{AB}^{\otimes n} \| G\right) \tag{3.38}$$

especially for large n, then we can solve a very nice conjecture that

$$E_{\mathrm{D,PPT}}\left(\rho_{AB}\right) = R^{\infty}\left(\rho_{AB}\right), \tag{3.39}$$

according to the result in [BP10].

Remark 3.8 It is also worth mentioning that we can quickly recover the Rains bound [Rai01, ADVW02] from Theorem 3.6. Recall that the the Rains bound is given by

$$R\left(\rho_{AB}\right) = \min_{\sigma\geq 0,\, \|\sigma^{T_B}\|_1 \leq 1} D\left(\rho_{AB} \| \sigma_{AB}\right), \tag{3.40}$$

Let σ be the minimizer of the Rains bound. We have $E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) \leq D_H^{\varepsilon}\left(\rho_{AB}^{\otimes n} \| \sigma_{AB}^{\otimes n}\right) - \delta$ by choosing the feasible solution G as $\sigma_{AB}^{\otimes n}$ and $\delta \in [0,1]$. Taking regularization on both sides, we obtain

$$
\begin{aligned}
E_{\mathrm{D,PPT}}\left(\rho_{AB}\right) &= \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) && (3.41)\\
&\leq \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} D_H^{\varepsilon}\left(\rho_{AB}^{\otimes n} \| \sigma_{AB}^{\otimes n}\right) && (3.42)\\
&= D\left(\rho_{AB} \| \sigma_{AB}\right) && (3.43)\\
&= R\left(\rho_{AB}\right), && (3.44)
\end{aligned}
$$

where the second equality follows from the quantum Stein's lemma [HP91, ON00].

3.3 Non-asymptotic entanglement distillation

We now introduce an efficiently computable estimation of the non-asymptotic distillable entanglement for general quantum states. Such estimation is especially accurate for large n where the term $O\left(\log n\right)$ is negligible.

Recall that the coherent information is defined as $I\left(A\rangle B\right)_{\rho} := D\left(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B\right)$. The quantum information variance is defined as $V\left(\rho \| \sigma\right) := \mathrm{Tr}\, \rho \left(\log \rho - \log \sigma\right)^2 - D\left(\rho \| \sigma\right)^2$. The coherent information variance is defined as $V\left(A\rangle B\right)_{\rho} := V\left(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B\right)$.

Theorem 3.9 *For any given bipartite quantum state ρ_{AB}, number of prepared states n, error tolerance $\varepsilon \in (0,1)$, and operation class $\Omega \in \{$1-LOCC, LOCC, SEP, PPT$\}$, it holds that*

$$
f\left(\rho, n, \varepsilon\right) + O\left(\log n\right) \leq E_{\mathrm{D},\Omega}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) \leq g\left(\rho, n, \varepsilon\right) + O\left(\log n\right), \tag{3.45}
$$

where $f\left(\rho, n, \varepsilon\right)$ and $g\left(\rho, n, \varepsilon\right)$ are efficiently computable functions given by

$$
\begin{aligned}
f\left(\rho, n, \varepsilon\right) &:= nI\left(A\rangle B\right)_{\rho} + \sqrt{nV\left(A\rangle B\right)_{\rho}}\, \Phi^{-1}\left(\varepsilon\right), && (3.46)\\
g\left(\rho, n, \varepsilon\right) &:= nR\left(\rho\right) + \sqrt{nV_R\left(\rho\right)}\, \Phi^{-1}\left(\varepsilon\right), && (3.47)
\end{aligned}
$$

and $V_R\left(\rho\right) = V\left(\rho \| \sigma\right)$, σ is the minimizer [4] of the Rains bound (3.40), Φ^{-1} is the inverse cumulative normal distribution function.

Proof Due to the inclusion relations of the operation classes, we only need to show the upper bound for PPT operations and lower bound for 1-LOCC operations. Each second-order bound

[4] The minimizer σ is not necessarily unique unless ρ is full rank. However, the result of this Theorem holds for any minimizer σ.

can be obtained by applying the corresponding one-shot bound to n-fold tensor product state $\rho^{\otimes n}$ and combining with the second-order expansion of the related entropies.

For the second-order upper bound, denote σ as the minimizer of the Rains boudn (3.40). We first have $E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) \leq D_H^{\varepsilon}\left(\rho_{AB}^{\otimes n}\|\sigma_{AB}^{\otimes n}\right) - \delta$ by taking the feasible solution $G = \sigma_{AB}^{\otimes n}$ in Eq. (3.30) and $\delta \in [0,1]$. Due to the second-order expansion of the quantum hypothesis testing relative entropy [TH13, Li14], we obtain

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) \leq nD\left(\rho||\sigma\right) + \sqrt{nV\left(\rho||\sigma\right)}\,\Phi^{-1}\left(\varepsilon\right) + O\left(\log n\right). \tag{3.48}$$

For the second-order lower bound, we adopt the one-shot hashing bound in Lemma 3.10 below. For the n-fold tensor product state $\rho_{AB}^{\otimes n}$, we choose $\eta = 1/\sqrt{n}$ and have the following result which holds for $n > 1/\varepsilon$,

$$E_{\mathrm{D,1\text{-}LOCC}}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) \geq -H_{\max}^{\sqrt{\varepsilon}-\eta}\left(A^n|B^n\right)_{\rho^{\otimes n}} + 4\log\eta. \tag{3.49}$$

Using the second-order expansion of the smoothed conditional max-entropy [TH13], we have

$$\begin{aligned} E_{\mathrm{D,1\text{-}LOCC}}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) &\geq -nH\left(A|B\right)_{\rho} + \sqrt{nV\left(A\rangle B\right)_{\rho}}\,\Phi^{-1}\Big(\big(\sqrt{\varepsilon} - \frac{1}{\sqrt{n}}\big)^2\Big) + O\left(\log n\right) &(3.50)\\ &= nI\left(A\rangle B\right)_{\rho} + \sqrt{nV\left(A\rangle B\right)_{\rho}}\,\Phi^{-1}\left(\varepsilon\right) + O\left(\log n\right). &(3.51)\end{aligned}$$

The second line follows since $I\left(A\rangle B\right)_{\rho} = -H\left(A|B\right)_{\rho}$. Note that Φ^{-1} is continuously differentiable around $\varepsilon > 0$ and we have $\Phi^{-1}\left(\left(\sqrt{\varepsilon} - 1/\sqrt{n}\right)^2\right) = \Phi^{-1}\left(\varepsilon\right) + O\left(1/\sqrt{n}\right)$.

■

Lemma 3.10 *([WTB16]) For any bipartite quantum state ρ_{AB}, error tolerance $\varepsilon \in (0,1)$ and $0 \leq \eta < \sqrt{\varepsilon}$, it holds*

$$E_{D,1\text{-}LOCC}^{(1),\varepsilon}\left(\rho_{AB}\right) \geq -H_{max}^{\sqrt{\varepsilon}-\eta}\left(A|B\right)_{\rho} + 4\log\eta. \tag{3.52}$$

Remark 3.11 It is worth noting that there are some other one-shot lower bounds [BD13, BD10a] which can be used to establish second-order estimations. One of the reasons we use the one-shot lower bound in Lemma 3.10 is because it gives the same ε dependence as our second-order upper bound. For pure states, there exists a better one-shot lower bound in [BD13]. But note that our second-order bounds are already tight for pure states up to the second order terms.

The difficulty of obtaining good second-order estimation is to find one-shot bounds which give the same epsilon dependence after the second-order expansion. This is necessary to show the tightness of second-order bounds. Our result in Theorem 3.6 works well in this sense.

Remark 3.12 Since the Rains bound in Eq. (3.40) is given as a convex optimization, there are various methods to solve it numerically. We provide an algorithm which can be used to

efficiently calculate the Rains bound and output the minimizer operator in Appendix B. It is also worth noting that our bounds are similar to the second-order bounds on quantum capacity in [TBR16]. However, the bounds in Ref. [TBR16] are not are not easy to calculate in general.

3.4 Examples

In the following, we apply our second-order bounds to estimate the non-asymptotic distillable entanglement of some important classes of states, including pure states, mixtures of Bell states, maximally correlated states and isotropic states.

Pure states

For any bipartite pure state $|\psi\rangle$, denote the reduced state as $\rho_A = \mathrm{Tr}_B \psi_{AB}$. Without loss of generality, we only need to consider a pure state ψ with Schmidt decomposition $|\psi\rangle = \sum \sqrt{p_i}|ii\rangle$, then $\rho_A = \rho_B = \sum p_i |i\rangle\langle i|$. Let $\sigma = \sum p_i |ii\rangle\langle ii|$. We can check that $\sigma \in \mathrm{PPT}'$ and the following equalities are straightforward,

$$D\left(\psi\|\sigma\right) = I\left(A\rangle B\right)_\psi = S\left(\rho_A\right), \tag{3.53}$$

$$V\left(\psi\|\sigma\right) = V\left(A\rangle B\right)_\psi = \mathrm{Tr}\,\rho_A \left(\log \rho_A\right)^2 - S\left(\rho_A\right)^2. \tag{3.54}$$

Since $D\left(\psi\|\sigma\right) = I\left(A\rangle B\right)_\psi \leq R\left(\psi\right) \leq D\left(\psi\|\sigma\right)$, we know that σ is the minimizer of the Rains bound optimization. Then

$$E_{\mathrm{D},\Omega}^{(1),\varepsilon}\left(\psi^{\otimes n}\right) = nS\left(\rho_A\right) + \sqrt{n\left(\mathrm{Tr}\,\rho_A \left(\log \rho_A\right)^2 - S\left(\rho_A\right)^2\right)}\,\Phi^{-1}\left(\varepsilon\right) + O\left(\log n\right), \tag{3.55}$$

for $\Omega \in \{$1-LOCC, LOCC, SEP, PPT$\}$. This estimation has been given by Datta and Leditzky [DL15], which was only for LOCC operations. It is known that the distillable entanglement of a pure state coincide with the von Neumann entropy of its reduced state under 1-LOCC, LOCC, SEP, PPT [LP01a]. The result in Eq. (3.55) shows that not only is the asymptotic distillable entanglement the same but also its convergence speed (second-order term) [5].

Mixture of Bell states

In laboratories, we usually obtain mixed states due to imperfection of operations and decoherence. A common case is the noise dominated by one type of Pauli error [CTSL06, Cam07]. Without loss of generality, we consider the phase noise, which results in the mixture of Bell

[5] We can further prove that for any pure state the one-shot distillable entanglement is also the same under these four classes of operations.

states $\rho_{AB} = p|v_1\rangle\langle v_1| + (1-p)|v_2\rangle\langle v_2|$, where $p \in (0,1)$, $|v_1\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ and $|v_2\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Following similar steps for pure states, we can take the minimizer $\sigma = \frac{1}{2}|v_1\rangle\langle v_1| + \frac{1}{2}|v_2\rangle\langle v_2|$ and show that our second-order bounds in Eq. (3.45) are tight. Thus

$$E_{\mathrm{D},\Omega}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) = n\left(1 - h_2\left(p\right)\right) + \sqrt{np\left(1-p\right)\left(\log\frac{1-p}{p}\right)^2}\,\Phi^{-1}\left(\varepsilon\right) + O\left(\log n\right), \quad (3.56)$$

where $h_2\left(p\right) = -p\log p - \left(1-p\right)\log\left(1-p\right)$ is the binary entropy and the operation class $\Omega \in \{\text{1-LOCC, LOCC, SEP, PPT}\}$.

Maximally correlated states

Besides the mixture of Bell states above, we can also show the tightness for a broader class of state, i.e., all maximally correlated states $\rho_{AB} = \sum_{i,j=0}^{d-1}\widehat{\rho}_{ij}|i_Ai_B\rangle\langle j_Aj_B|$, where $\widehat{\rho}_A = \sum_{i,j=0}^{d-1}\widehat{\rho}_{ij}|i_A\rangle\langle j_A|$ is a quantum state. Taking the minimizer $\sigma = \Delta\left(\rho\right)$ where $\Delta\left(\cdot\right) = \sum_{i,j=0}^{d-1}\langle ij|\cdot|ij\rangle|ij\rangle\langle ij|$ is the completely dephashing channel on the bipartite systems, we have

$$E_{\mathrm{D},\Omega}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) = nD\left(\rho\|\Delta\left(\rho\right)\right) + \sqrt{nV\left(\rho\|\Delta\left(\rho\right)\right)}\,\Phi^{-1}\left(\varepsilon\right) + O\left(\log n\right), \quad (3.57)$$

for $\Omega \in \{\text{1-LOCC, LOCC, SEP, PPT}\}$. Note that if $\widehat{\rho}_A$ is a pure state, then maximally correlated state ρ_{AB} will also reduce to a bipartite pure states. The class of maximally correlated states also contains the mixture of Bell states.

Furthermore, since the resource theory of coherence is closely related to the one of entanglement due to the one-to-one correspondence

$$\rho_{AB} = \sum_{i,j=0}^{d-1}\widehat{\rho}_{ij}|i_Ai_B\rangle\langle j_Aj_B| \Longrightarrow \widehat{\rho}_A = \sum_{i,j=0}^{d-1}\widehat{\rho}_{ij}|i_A\rangle\langle j_A|, \quad (3.58)$$

an important conjecture with various evidences is that any incoherent operation acting on a state $\widehat{\rho}_A$ is equivalent to an LOCC operation acting on the maximally correlated state ρ_{AB} [WY16, SAP17]. If this conjecture holds, Eq. (3.57) will also give the second-order characterization for coherence distillation.

Isotropic states

Another common noise in reality is the so-called depolarizing noise [Cam07, NFB14], which results in the isotropic state,

$$\rho_F = F\cdot\Phi_d + \left(1-F\right)\frac{\mathbb{1} - \Phi_d}{d^2-1}, \quad (3.59)$$

where $0 \leq F \leq 1$ and d is the local dimension of the maximally entangled state Φ_d. The isotropic state is also the Choi-Jamiołkowski state of the depolarizing channel, and its 1-LOCC distillable entanglement is equal to the quantum capacity of the depolarizing channel [BDSW96, PLOB17] $\mathcal{N}(\rho) = p\rho + (1-p)\mathbb{1}/d$, the determination of which is still a big open problem in quantum information theory. Here we study the non-asymptotic distillable entanglement of this particular class of states.

Isotropic states possess the same symmetry as the maximally entangled states, which are invariant under any local unitary $U \otimes \overline{U}$. Exploiting such symmetry, we can simplify the PPT-assisted distillable entanglement for the n-fold isotropic state as a linear program [6].

Lemma 3.13 *For any n-fold isotropic state $\rho_F^{\otimes n}$ with integer n and error tolerance ε, its one-shot distillable entanglement under PPT operations $E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho_F^{\otimes n}\right)$ is given by*

$$\begin{aligned} \log \quad \text{maximize} \quad & \lfloor 1/\eta \rfloor && \text{(3.60a)} \\ \text{subject to} \quad & \sum_{i=0}^{n} \binom{n}{i} F^i (1-F)^{n-i} m_i \geq 1-\varepsilon, && \text{(3.60b)} \\ & -\eta \leq \sum_{i=0}^{n} x_{i,k} m_i \leq \eta, \ \forall\, k = 0, 1, \cdots, n, && \text{(3.60c)} \\ & 0 \leq m_i \leq 1, \ \forall\, i = 0, 1, \cdots, n, && \text{(3.60d)} \end{aligned}$$

where

$$x_{i,k} = \frac{1}{d^n} \sum_{m=\max\{0,i+k-n\}}^{\min\{i,k\}} \binom{k}{m} \binom{n-k}{i-m} (-1)^{i-m} (d-1)^{k-m} (d+1)^{n-k+m-i}. \quad \text{(3.61)}$$

Proof The technique is very similar to the one we use in the proof of Lemma 3.3. Consider the n-fold isotropic state

$$\rho_F^{\otimes n} = \sum_{i=0}^{n} f_i P_i^n \left(\Phi_d, \Phi_d^{\perp}\right), \text{ with } f_i = F^i \left(\frac{1-F}{d^2-1}\right)^{n-i}, \Phi_d^{\perp} = \mathbb{1} - \Phi_d. \quad \text{(3.62)}$$

Here, $P_i^n\left(\Phi, \Phi^{\perp}\right)$ represents the sum of those n-fold tensor product terms with exactly i copies of Φ_d. For example,

$$P_1^3\left(\Phi_d, \Phi_d^{\perp}\right) = \Phi_d^{\perp} \otimes \Phi_d^{\perp} \otimes \Phi_d + \Phi_d^{\perp} \otimes \Phi_d \otimes \Phi_d^{\perp} + \Phi_d \otimes \Phi_d^{\perp} \otimes \Phi_d^{\perp}. \quad \text{(3.63)}$$

Suppose M is the optimal solution of the optimization

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}\left(\rho_F^{\otimes n}\right) = \log \max\left\{ \lfloor 1/\eta \rfloor \,\middle|\, 0 \leq M \leq \mathbb{1}, \operatorname{Tr} M\rho_F^{\otimes n} \geq 1-\varepsilon, -\eta\mathbb{1} \leq M^{T_B} \leq \eta\mathbb{1} \right\}, \quad \text{(3.64)}$$

[6] Note that the optimal fidelity for n-fold isotropic states can also be simplified to a linear program, which has been studied by Rains in [Rai01]. Here, we focus on the distillable rate.

then for any local unitary $U = \bigotimes_{i=1}^{n} \left(U_A^i \otimes \overline{U}_B^i\right)$ where i denotes the i-th copies of corresponding system, $UMU^\dagger$ is a also optimal solution. Convex combinations of optimal solutions are also optimal. So we can take the optimal solution M to be an operator which is invariant under any local unitary $\bigotimes_{i=1}^{n} \left(U_A^i \otimes \overline{U}_B^i\right)$. Moreover, since $\rho_F^{\otimes n}$ is invariant under the symmetric group acting by permuting the tensor factors, we can take the optimal solution M of the form $\sum_{i=0}^{n} m_i P_i^n \left(\Phi_d, \Phi_d^\perp\right)$.

Since $P_i^n \left(\Phi_d, \Phi_d^\perp\right)$ are orthogonal projections, the operator M has eigenvalues $\{m_i\}_{i=0}^n$ without considering degeneracy. Next, we will need to know the eigenvalues of M^{T_B}. Decomposing operators $\Phi_d^{T_B}$ and $\Phi_d^{\perp\, T_B}$ into orthogonal projections, i.e.,

$$\Phi_d^{T_B} = \frac{1}{d}\left(P_+ - P_-\right), \quad \Phi_d^{\perp\, T_B} = \left(1 - \frac{1}{d}\right) P_+ + \left(1 + \frac{1}{d}\right) P_- \tag{3.65}$$

where P_+ and P_- are symmetric and anti-symmetric projections respectively and collecting the terms with respect to $P_k^n \left(P_+, P_-\right)$, we have

$$M^{T_B} = \sum_{i=0}^{n} m_i P_i^n \left(\Phi_d^{T_B}, \Phi_d^{\perp\, T_B}\right) \tag{3.66}$$

$$= \sum_{i=0}^{n} m_i \Big(\sum_{k=0}^{n} x_{i,k} P_k^n \left(P_+, P_-\right) \Big) \tag{3.67}$$

$$= \sum_{k=0}^{n} \Big(\sum_{i=0}^{n} x_{i,k} m_i \Big) P_k^n \left(P_+, P_-\right). \tag{3.68}$$

Since $P_k^n \left(P_+, P_-\right)$ are also orthogonal projections, M^{T_B} has eigenvalues $\{t_k\}_{k=0}^n$ without considering degeneracy, where $t_k = \sum_{i=0}^{n} x_{i,k} m_i$. As for the condition $\operatorname{Tr} M\rho_F^{\otimes n} \geq 1 - \varepsilon$, we have

$$\operatorname{Tr} M\rho_F^{\otimes n} = \operatorname{Tr} \sum_{i=0}^{n} f_i m_i P_i^n \left(\Phi_d, \Phi_d^\perp\right) \tag{3.69}$$

$$= \sum_{i=0}^{n} f_i m_i \binom{n}{i} \left(d^2 - 1\right)^{n-i} \tag{3.70}$$

$$= \sum_{i=0}^{n} \binom{n}{i} F^i \left(1 - F\right)^{n-i} m_i. \tag{3.71}$$

■

This linear program can be solved *exactly* via Mathematica. In Figure 3.3, we plot the one-shot distillable entanglement for the n-fold isotropic state $\rho_F^{\otimes n}$ with $d = 3$, $F = 0.9$, and error tolerance 0.001. The blocklength n ranges from 1 to 100. We observe that even if we were able to coherently manipulate 100 copies of the states with the broad class of PPT assistance, the maximal distillation rate still could not reach the hashing bound $I\left(A\rangle B\right)_{\rho_F}$

which is asymptotically achievable. This implies the asymptotic rate does not give a good estimation in the practical scenario.

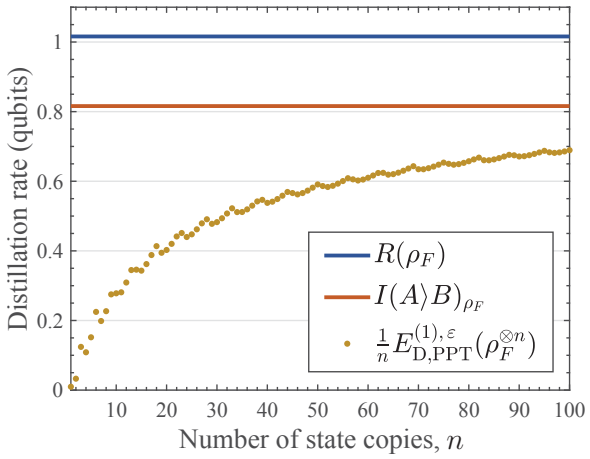


Figure 3.3 : The dotted line shows the exact value of distillation rate for n-fold isotropic state $\rho_F^{\otimes n}$ with $F = 0.9$, local dimension $d = 3$. The error tolerance is taken at $\varepsilon = 0.001$ and the blocklength n ranges from 1 to 100. The solid line below is the hashing bound, while the solid line above is the Rains bound.

For the approximation of large blocklength entanglement distillation, we employ the second-order bounds in Eq. (3.45). In Figure 3.4, we show the second-order estimation for n-fold isotropic state $\rho_F^{\otimes n}$ with $d = 3$, $F = 0.9$, and error tolerance 0.001. In this figure we focus on large blocklength ($n \geq 100$) and use a logarithmic scale of the horizontal axis. The finite blocklength distillation rate lies between the two dashed lines, while the asymptotic rate lies between the two solid lines.

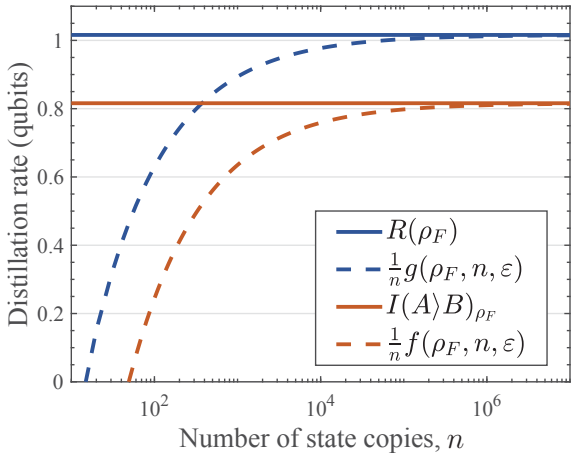


Figure 3.4 : The two dashed lines show the second-order lower bound in Eq. (3.46) and upper bound in Eq. (3.47) for n-fold isotropic state $\rho_F^{\otimes n}$ with $F = 0.9$, local dimension $d = 3$. The error tolerance is taken at $\varepsilon = 0.001$ and the blocklength n ranges from 10^2 to 10^7. The solid line below is the hashing bound, while the solid line above is the Rains bound.

Using the curve fitting via least-squares method, we can construct an ansatz curve

$$c_1 + c_2 \frac{1}{\sqrt{n}} + c_3 \frac{\log n}{n} + c_4 \frac{1}{n}, \tag{3.72}$$

which has the best fit to the series of points $\frac{1}{n} E_{\mathrm{D,PPT}}^{(1),\varepsilon} \left(\rho_F^{\otimes n}\right)$ $(1 \leq n \leq 100)$ in Figure 3.3. Combining with the second-order upper bound in Figure 3.4, we get Figure 3.5. It shows that for small number of copies n, the second-order upper bound does not give a accurate estimation since we ignore the term $O(\log n)$. But for large n $(\geq 10^2)$, the fitting curve almost coincides with the second-order upper bound. Thus the second-order upper bound works better for large blocklength estimation. The convergence of the fitting curve indicates that $E_{\mathrm{D,PPT}}(\rho_F) = R(\rho_F)$ for isotropic states ρ_F. It would be of great interest to find an analytical proof for this conjecture.

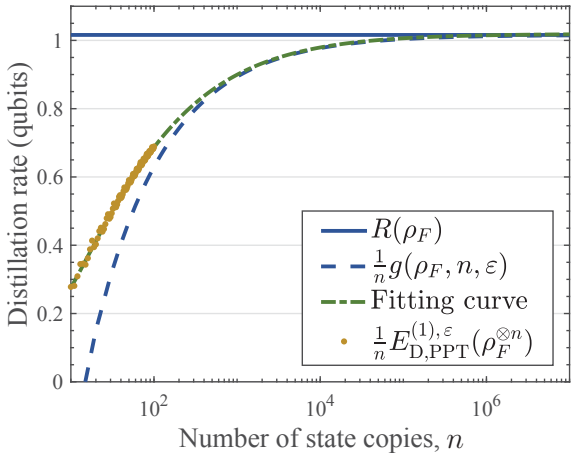


Figure 3.5 : The dash-dotted line is the fitting curve of exact values of distillation rate for n-fold isotropic state $\rho_F^{\otimes n}$ with $F = 0.9$, local dimension $d = 3$. The error tolerance is taken at $\varepsilon = 0.001$. The dashed line is the second-order upper bound in Eq. (3.47) and the solid line is the Rains bound.

3.5 Discussion

3.5.1 Summary of results

1. SDP for one-shot distillable entanglement under PPT operations:

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho) = \log \max \left\{ \lfloor 1/\eta \rfloor \,\middle|\, 0 \leq M \leq \mathbb{1}, \mathrm{Tr}\, M\rho \geq 1-\varepsilon, -\eta\mathbb{1} \leq M^{T_B} \leq \eta\mathbb{1} \right\}.$$

2. Connection between entanglement distillation under PPT operations and the quantum hypothesis testing:

$$E_{\mathrm{D,PPT}}^{(1),\varepsilon}(\rho_{AB}) = \min_{\substack{\|G^{T_B}\|_1 \leq 1 \\ G=G^\dagger}} D_H^\varepsilon(\rho_{AB}\|G) - \delta,$$

   where $\delta \in [0,1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

3. Efficiently computable second-order estimation for general quantum states:

$$f(\rho,n,\varepsilon) + O(\log n) \leq E_{\mathrm{D},\Omega}^{(1),\varepsilon}\left(\rho_{AB}^{\otimes n}\right) \leq g(\rho,n,\varepsilon) + O(\log n),$$

   where f and g are given by Eqs. (3.46) and (3.47) respectively.

4. Various practical example analysis including the pure states, mixture of Bell states, maximally correlated states and isotropic states.

3.5.2 Outlook

We have provided both theoretical and numerical results for the entanglement distillation in the non-asymptotic setting. Since entanglement distillation has become a central building block of quantum network proposals [DBCZ99, CTSL06, GJC12, NFB14], our results can be applied as useful benchmarks for experimentalists to build a reliable quantum network in the future. Theoretically, we have obtained a complete characterization of one-shot entanglement distillation under PPT operations and built a precise connection between one-shot distillable entanglement and hypothesis testing relative entropy in Eq. (3.30). This also provides a potential approach to improve the Rains bound by taking other forms of feasible solution, for example, non-i.i.d. operators. The recent progress on the entropy accumulation theorem [DFR16] can estimate the entropy without the assumption of i.i.d. input. It is thus of great interest to further relax the finite analysis of entanglement distillation to a more general case.

Chapter 4

Coherence distillation

4.1 Introduction

4.1.1 Background

Quantum coherence is a physical resource that is essential for various tasks in quantum computing (e.g. implementing the Deutsch-Jozsa algorithm [Hil16]), quantum cryptography (e.g. quantum key distribution [CML16]), quantum information processing (e.g. quantum state merging [SCR+16], state redistribution [AJS18] and channel simulation [DFW+18]), thermodynamics [LJR15] and quantum metrology [FD11]. A series of efforts have been devoted to building a resource framework of coherence in recent years [Abe06, GS08, LM14, BCP14, SAP17], characterizing in particular the state transformations and operational uses of coherence in fundamental resource manipulation protocols [WY16, CSR+16, CH16, ZLY+18, RFWA18]. As in any physical resource theory, a central problem of the resource theory of quantum coherence is *distillation*: the process of extracting canonical units of coherence, as represented by the maximally coherent state $|\Psi_m\rangle$, from a given quantum state using a choice of free operations.

The free states in the resource theory of quantum coherence, so-called incoherent states, are the density operators which are diagonal in a given reference orthonormal basis $\{|i\rangle\}$. The set of incoherent states is denoted as $\mathcal{I}$. Depending on the choice of operations, one has a different resource theory. There have been different proposals for the class of free operations to be considered [WY16, CG16b, MS16, dVS17, SAP17], and there is no agreement on which one should be considered as the most well motivated physically. This necessitates the investigation of operational capabilities of several different classes of maps. The relevant choices of free operations that we will focus on are: maximally incoherent operations (MIO) [Abe06], defined to be all quantum operations $\mathcal{E}$ such that $\mathcal{E}(\rho) \in \mathcal{I}$ for every $\rho \in \mathcal{I}$; dephasing-covariant incoherent operations (DIO) [CG16b, MS16], which are operations $\mathcal{E}$ such that $[\Delta, \mathcal{E}] = 0$,

or equivalently $\mathcal{E}(|i\rangle\langle i|) \in \mathcal{I}$ and $\Delta(\mathcal{E}(|i\rangle\langle j|)) = 0, \forall i \neq j$. In particular, MIO is the largest possible choice of free operations in the resource theory of coherence, and therefore its operational capabilities establish ultimate limits for transformations with all types of free operations.

The usual asymptotic approach to studying the problem in quantum information theory is to assume that there is an unbounded number of independent and identically distributed copies of a quantum state available and impose that the transformation error asymptotically goes to zero [BBPS96, BBP+96, Rai99b, WY16]. In a realistic setting, these assumptions become unphysical due to our limited access to a finite number of copies of a given state, and it becomes necessary to look at non-asymptotic and, in particular, one-shot regimes in detail. More importantly, since loss and decoherence severely restrict our ability to manipulate large quantum systems, one needs to allow for a finite error in the distillation protocol.

Like any resource transformation, there are two kinds of setting for coherence distillation, i.e., *deterministic* distillation and *probabilistic* distillation. In the deterministic scenario, the transformation process always succeeds and we can study the tradeoff between the distillable rate and the distillation fidelity. However, deterministic protocols such as the ones recently studied in [RFWA18] may not always be the most suitable choice, particularly when the maximal achievable fidelity of distillation is not sufficient for the desired applications. It is thus of importance to consider a more general framework, probabilistic coherence distillation, in which the distillation will succeed only with some probability. Here, the allowed error can be characterized by two key parameters with practical relevance: the *success probability* of the one-shot distillation process, and the *fidelity* between the extracted state and the target state $|\Psi_m\rangle$. To have a systematic understanding of coherence distillation with finite resources and be able to implement practical schemes for this task, it is crucial to describe and optimize the fundamental relations between these two parameters.

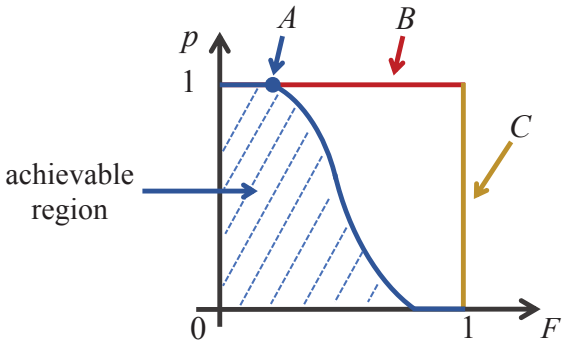


Figure 4.1 : A schematic plot of the tradeoff between success probability (p) and the fidelity of coherence distillation (F).

In Figure 4.1, we show a schematic plot of the tradeoff between the success probability and the fidelity. For a given resource quantum state and target state dimension, we may compromise a bit of the success probability to obtain a quantum state of higher fidelity. The red line (marked with arrow B) corresponds to the deterministic coherence distillation that the success probability is forced to be one. In section 4.2 we will see MIO and DIO have exactly the same power on this line. The blue dot (marked with arrow A) indicates the maximal fidelity of coherence distillation, which is a key parameter in the deterministic scenario. The yellow line (marked with arrow C) corresponds to the zero-error scenario, on which the maximal success probability admits an analytical formula for pure resource state [LP01b, CG16b, DBG15]. In section 4.3, we will show that MIO can be much more powerful than DIO in this zero-error case. More generally, if there exists a quantum operation with which we can distill a coherent state with fidelity F and success with probability p, the tuple (F, p) is called achievable. The shaded area indicates all the achievable tuples. In section 4.3, we will discuss how to obtain the boundary of such achievable region via semidefinite programs.

4.1.2 Outline

In section 4.2, we study the deterministic distillation of quantum coherence in the one-shot setting, that is, the conversion of general quantum states into maximally coherent states under different classes of quantum operations. We show that MIO and DIO have the same power in the task of one-shot coherence distillation. We establish that the one-shot distillable coherence under MIO and DIO is efficiently computable with a semidefinite program, which we show to correspond to a quantum hypothesis testing problem.

In section 4.3, we develop a general framework of probabilistic coherence distillation. We interpret the fundamental relations between the distillation fidelity and the maximal success probability via a gauge function construction, and show that the maximal success probability under MIO and DIO can be efficiently computed via semidefinite programming. We demonstrate that the distillation of perfect coherence from any full-rank state is impossible even probabilistically, while any pure coherent state can always be perfectly distilled with MIO into a maximally coherent state of arbitrary dimension with a non-zero probability. This result highlights an operational advantage of MIO over other classes of operations and contrasts with the case of deterministic distillation. We provided an analytical characterization of distillation with pure input states and in particular described the distillation of qubit maximally coherent states under MIO and DIO. We further explored novel phenomena of coherence distillation such as the breakdown of the tradeoff between the maximal success probability and the fidelity under a certain threshold as well as the catalyst-assisted enhancement by maximally coherent states.

4.2 Deterministic coherence distillation

In this section, we study the deterministic coherence distillation in the one-shot setting, in which one has access only to a single copy of a quantum system and allows for a finite accuracy, reflecting the realistic restrictions on state transformations. In particular, we establish an exact expression for the one-shot distillable coherence under MIO and DIO, which can be efficiently computed as an SDP. Interestingly, we show that the two quantities are in fact the same, demonstrating that MIO and DIO have the same power in the task of coherence distillation, and together with recent results in coherence dilution [ZLY+18, Chi18] shedding light on the asymptotic reversibility of state transformations in the resource of coherence under DIO.

4.2.1 Framework of deterministic coherence distillation

We will denote by $\Psi_m = |\Psi_m\rangle\langle\Psi_m|$ the m-dimensional maximally coherent state where $|\Psi_m\rangle = \sum_{i=0}^{m-1} \frac{1}{\sqrt{m}}|i\rangle$ in the reference basis. The *distillable coherence* $C_{d,\mathrm{IO}}(\rho)$ is the asymptotic rate at which Ψ_2 can be obtained per copy of a given state ρ via incoherent operations. Winter and Yang [WY16] showed that the distillable coherence of an arbitrary mixed state coincides with the *relative entropy of coherence* $C_r(\rho) = \min_{\sigma\in\mathcal{I}} D(\rho\|\sigma)$ introduced in [Abe06]. For any state ρ, the distillable coherence is then given by

$$C_{d,\mathrm{IO}}(\rho) = C_r(\rho) = S(\Delta(\rho)) - S(\rho). \tag{4.1}$$

We now consider the non-asymptotic setting.

Definition 4.1 *For any quantum state ρ, the fidelity of coherence distillation under the class of operations Ω is defined by*

$$F_\Omega(\rho, m) = \max_{\Pi\in\Omega} \mathrm{Tr}\,\Pi(\rho)\,\Psi_m. \tag{4.2}$$

Definition 4.2 *For any quantum state ρ and error tolerance ε, the one-shot ε-error distillable coherence under the class of operation Ω is defined as*

$$C_{d,\Omega}^{(1),\varepsilon}(\rho) := \log\max\left\{m\in\mathbb{N} \,\middle|\, F_\Omega(\rho, m) \geq 1-\varepsilon\right\}. \tag{4.3}$$

As a consequence, the asymptotic distillable coherence can be given as

$$C_{d,\Omega}(\rho) = \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}C_{d,\Omega}^{(1),\varepsilon}\left(\rho^{\otimes n}\right). \tag{4.4}$$

4.2.2 Distillation rate of quantum coherence

One of the main results of this section is that the one-shot distillable coherence can be computed exactly via the following SDP.

Theorem 4.3 *For any state ρ and operation class $\Omega \in \{\mathrm{MIO}, \mathrm{DIO}\}$, the fidelity of coherence distillation and the one-shot distillable coherence can both be written as the following SDPs:*

$$F_{\Omega}(\rho, m) = \max\left\{ \mathrm{Tr}\, G\rho \;\middle|\; 0 \leq G \leq \mathbb{1},\ \Delta(G) = \frac{1}{m}\mathbb{1} \right\}, \tag{4.5}$$

$$C_{d,\Omega}^{(1),\varepsilon}(\rho) = \log \max\left\{ \lfloor 1/\eta \rfloor \;\middle|\; \mathrm{Tr}\, G\rho \geq 1-\varepsilon, 0 \leq G \leq \mathbb{1}, \Delta(G) = \eta\mathbb{1} \right\}. \tag{4.6}$$

Proof Let us consider the class MIO first. Denote J_{Π} as the Choi-Jamiołkowski matrix of operation Π. Then due to the Choi-Jamiołkowski isomorphism, it holds

$$\mathrm{Tr}\,\Pi(\rho)\,\Psi_m = \mathrm{Tr}\left(\mathrm{Tr}_A\, J_{\Pi} \cdot \rho^T \otimes \mathbb{1}\right)\Psi_m = \mathrm{Tr}\left(\mathrm{Tr}_B\left(J_{\Pi} \cdot \mathbb{1} \otimes \Psi_m\right)\right)\rho^T. \tag{4.7}$$

According to the definition, we have the following optimization

$$C_{d,\mathrm{MIO}}^{(1),\varepsilon}(\rho) = \log \max\left\{ m \in \mathbb{N} \;\middle|\; \mathrm{Tr}\left(\mathrm{Tr}_B\left(J_{\Pi} \cdot \mathbb{1} \otimes \Psi_m\right)\right)\rho^T \geq 1-\varepsilon, \Pi \in \mathrm{MIO} \right\}. \tag{4.8}$$

Denote U_{π} as the unitary that permutes the basis $\{|i\rangle\}$ to $\{|\pi(i)\rangle\}$. Then Ψ_m is invariant under all U_{π}. Suppose the optimal solution of optimization (4.8) is taken at $\widetilde{J}_{\Pi}$. We can check that $(\mathbb{1} \otimes U_{\pi})\,\widetilde{J}_{\Pi}\,(\mathbb{1} \otimes U_{\pi})^{\dagger}$ is also optimal. Since any convex combination of optimal solutions remains optimal, we know that $(1/m!)\sum_{\pi}(\mathbb{1} \otimes U_{\pi})\,\widetilde{J}_{\Pi}\,(\mathbb{1} \otimes U_{\pi})^{\dagger}$ is also optimal. According to Schur's lemma, this uniform average will lead to a particular structure of an optimal solution $J_{\Pi} = R \otimes \mathbb{1} + Q \otimes \left(\Psi_m - \frac{1}{m}\mathbb{1}\right)$. Moreover, since $\mathrm{Tr}_B\, J_{\Pi} = \mathbb{1}$, we have $R = \frac{1}{m}\mathbb{1}$. Since the operation maps any incoherent state to an incoherent state, we have $\Delta(Q) = 0$. Thus there exists an optimal operation admitting the structure that

$$J_{\Pi} = Q \otimes \left(\Psi_m - \frac{1}{m}\mathbb{1}\right) + \mathbb{1} \otimes \frac{1}{m}\mathbb{1},\ \Delta(Q) = 0. \tag{4.9}$$

Then $J_{\Pi} \geq 0$ if and only if $-\frac{1}{m-1}\mathbb{1} \leq Q \leq \mathbb{1}$. By direct calculation, $\mathrm{Tr}\,\Pi(\rho)\,\Psi_m = \left(1-\frac{1}{m}\right)\mathrm{Tr}\, Q\rho^T + \frac{1}{m}$. Then we have

$$F_{\mathrm{MIO}}(\rho, m) = \max\left\{ \left(1-\frac{1}{m}\right)\mathrm{Tr}\, Q\rho + \frac{1}{m} \;\middle|\; -\frac{1}{m-1}\mathbb{1} \leq Q \leq \mathbb{1}, \Delta(Q) = 0 \right\}. \tag{4.10}$$

Replacing $G = \frac{m-1}{m}Q + \frac{1}{m}\mathbb{1}$, we have

$$F_{\mathrm{MIO}}(\rho, m) = \max\left\{ \mathrm{Tr}\, G\rho \;\middle|\; 0 \leq G \leq \mathbb{1}, \Delta(G) = \frac{1}{m}\mathbb{1} \right\}. \tag{4.11}$$

The optimal operation is given by

$$\Pi(\rho) = (\mathrm{Tr}\, G\rho)\,\Psi_m + (1 - \mathrm{Tr}\, G\rho)\,\frac{\mathbb{1} - \Psi_m}{m-1},\ 0 \leq G \leq \mathbb{1},\ \Delta(G) = \frac{1}{m}\mathbb{1}. \tag{4.12}$$

Following the definition of one-shot distillable coherence and substituting $\eta = 1/m$, we have the SDP for $C^{(1),\varepsilon}_{d,\mathrm{MIO}}(\rho)$. From Eq. (4.9), we know that the optimal operation is also in DIO. Thus we have $F_{\mathrm{MIO}}(\rho, m) = F_{\mathrm{DIO}}(\rho, m)$ and $C^{(1),\varepsilon}_{d,\mathrm{MIO}}(\rho) = C^{(1),\varepsilon}_{d,\mathrm{DIO}}(\rho)$. ■

This result reveals a fundamental relation between different sets of operations in the resource theory of coherence, showing that MIO and DIO have the same power in the task of coherence distillation. This correspondence is in fact surprising: not only is DIO a strict subset of MIO, it is also known that MIO is strictly more powerful than DIO in state transformations [CG16a, MS16], that there exist entropic coherence monotones under DIO which are not monotones under MIO [CG16a], and that the two sets can exhibit different operational capabilities in tasks such as coherence dilution [ZLY+18]. Furthermore, since MIO constitutes the largest class of free operations in the resource theory of coherence, the result is of practical relevance as it shows that using DIO is sufficient to achieve the best rates of distillation under any class of free operations. It is unknown if there is a corresponding result in entanglement theory — the distillable rate under the (approximately) separability-preserving operations [BD11] can be achieved via its strictly subset.

Remark 4.4 The calculation of the optimization (4.6) can be implemented with two separate steps which are similar to Remark 3.4. With the same reasoning as Remark 3.5, we can use the least constant $\delta \in [0, 1]$ to adjust the r.h.s. of Eq. (4.6) to be the logarithm of an integer, i.e,

$$C^{(1),\varepsilon}_{d,\Omega}(\rho) = -\log \min \left\{ \eta \,\middle|\, \mathrm{Tr}\, G\rho \geq 1-\varepsilon, 0 \leq G \leq \mathbb{1}, \Delta(G) = \eta \mathbb{1} \right\} - \delta. \tag{4.13}$$

We will now show that the quantities introduced in Theorem 4.3 admit alternative characterizations. In particular, we will express the one-shot distillable coherence as a quantum hypothesis testing problem [1] which is similar to the result in Theorem 3.6. To do so, we will need to optimize over a larger set of matrices than the incoherent states $\mathcal{I}$: namely, the set of diagonal Hermitian operators with unit trace,

$$\mathcal{J} = \left\{ X \in \mathcal{H}erm \,\middle|\, \mathrm{Tr}\, X = 1,\ \Delta(X) = X \right\} \tag{4.14}$$

Proposition 4.5 *For any quantum state ρ, error tolerance ε and the operation class $\Omega \in \{\mathrm{MIO}, \mathrm{DIO}\}$, it holds that*

$$C^{(1),\varepsilon}_{d,\Omega}(\rho) = \min_{X \in \mathcal{J}} D^{\varepsilon}_{H}(\rho \| X) - \delta, \tag{4.15}$$

where $\delta \in [0, 1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

[1] Note that the quantum hypothesis testing here only indicates the convex optimization without a clear operational meaning. The second term in the hypothesis testing relative entropy is extended to Hermitian operators.

Proof To begin with, note that

$$\inf\left\{\,\eta \mid \Delta\left(W\right)=\eta\mathbb{1}\,\right\}=\sup\left\{\,\mathrm{Tr}\,M\Delta\left(W\right)\;\middle|\;\mathrm{Tr}\,M=1,M=M^{\dagger}\right\} \tag{4.16}$$

by SDP strong duality. Denote the set of operators $\mathcal{S}_{\rho}:=\left\{\,G\mid\ \mathrm{Tr}\,G\rho\geq 1-\varepsilon,\,0\leq G\leq\mathbb{1}\,\right\}$. From Eq. (4.13), we can obtain

$$\begin{aligned}
C_{d,\Omega}^{(1),\varepsilon}\left(\rho\right) &= -\log\min_{\mathcal{S}_{\rho}}\min\left\{\eta\mid\Delta\left(G\right)=\eta\mathbb{1}\right\}-\delta && (4.17)\\
&= -\log\min_{\mathcal{S}_{\rho}}\max_{\substack{\mathrm{Tr}\,M=1\\M=M^{\dagger}}}\mathrm{Tr}\,M\Delta\left(G\right)-\delta && (4.18)\\
&= -\log\max_{\substack{\mathrm{Tr}\,M=1\\M=M^{\dagger}}}\min_{\mathcal{S}_{\rho}}\mathrm{Tr}\,\Delta\left(M\right)G-\delta && (4.19)\\
&= \min_{\substack{\mathrm{Tr}\,M=1\\M=\Delta(M)\\M=M^{\dagger}}}-\log\min_{\mathcal{S}_{\rho}}\mathrm{Tr}\,MG-\delta && (4.20)\\
&= \min_{M\in\mathcal{J}}D_{H}^{\varepsilon}\left(\rho\|M\right)-\delta && (4.21)
\end{aligned}$$

where the third equality follows by Sion's minimax theorem [Sio58] and the self-duality of the completely dephasing operation Δ. Without loss of generality, we take $\log x=-\infty$ for any $x\leq 0$. ■

This result have important consequences beyond the one-shot regime, in particular for the *asymptotic* reversibility of state transformations in the resource theory of coherence — that is, the question whether the amount of coherence which can be distilled from a number of copies of a state ρ (distillable coherence C_d) is the same as the amount of coherence needed to prepare the same number of copies (coherence cost C_c) in the asymptotic limit of an arbitrarily large number of i.i.d. copies. It is known that the resource theory of coherence is reversible under MIO [BG15, ZLY+18], but irreversible under IO as we have $C_{d,\mathrm{IO}}\left(\rho\right)<C_{c,\mathrm{IO}}\left(\rho\right)$ in general [WY16]. Recently, it has been claimed that $C_{c,\mathrm{DIO}}\left(\rho\right)=C_{c,\mathrm{MIO}}\left(\rho\right)=C_{r}\left(\rho\right)$ [ZLY+18], although a complete proof of this fact did not appear until [Chi18]. Our result in Theorem 4.3 in particular shows that $C_{d,\mathrm{DIO}}^{(1),\varepsilon}\left(\rho\right)=C_{d,\mathrm{MIO}}^{(1),\varepsilon}\left(\rho\right)$ and therefore $C_{d,\mathrm{DIO}}\left(\rho\right)=C_{d,\mathrm{MIO}}\left(\rho\right)$. Combining the known result [WY16] $C_{d,\mathrm{MIO}}\left(\rho\right)=C_{r}\left(\rho\right)$, we have $C_{d,\mathrm{DIO}}\left(\rho\right)=C_{d,\mathrm{MIO}}\left(\rho\right)=C_{r}\left(\rho\right)$, complementing the claims of Ref. [ZLY+18] and strenghtening the asymptotic results of Ref. [Chi18] by showing their applicability even in the one-shot case. The fact that state transformations are indeed reversible under DIO and the maximal set of operations MIO is not necessary for full reversibility contrasts with other resource theories such as entanglement, where the maximal set of operations is the only one known to provide asymptotic reversibility [VC01, VDC02, BP08, WD17a].

4.3 Probabilistic coherence distillation

In this section, we develop the framework of probabilistic coherence distillation, characterizing the relation between the maximum success probability and the fidelity of distillation in the one-shot setting. We describe the quantitative and qualitative aspects of this task under several representative choices of free operations, providing insights into not only the practical aspects of state manipulations in the resource theory of quantum coherence, but also the fundamental differences between physically relevant classes of free operations and their performance in coherence manipulation. Before proceeding, we note that, previously, the framework of probabilistic state transformations has been employed in characterizing entanglement distillation [BBPS96, LP01b, Vid99, JP99, IP05] as well as related settings in the resource theory of thermodynamics [AOP16], and recently found use in the investigation of practical entanglement distillation schemes [RST+18]. Our results fill an important gap in the literature by establishing the probabilistic toolbox for the key resource of quantum coherence.

4.3.1 Framework of probabilistic coherence distillation

The basic task of probabilistic distillation can be understood as follows. For any given quantum state ρ held by a single party A, we aim to transform this state to an m-dimensional maximally coherent state (target state) $|\Psi_m\rangle = \frac{1}{\sqrt{m}} \sum_{i=0}^{m-1} |i\rangle$ with high fidelity. A single-bit classical flag register L is used to indicate whether the transformation succeeds or not. If the flag is in the 0 state, it tells that the distillation process succeeds and the output state σ has fidelity at least $1-\varepsilon$ with the target state. Otherwise, the process has failed, and we discard the unwanted output state ω. Our goal is then to maximize the success probability while keeping the transformation infidelity within some tolerance ε. Let (ρ, m, ε) be a triple with a given initial state ρ, target state dimension m, and error tolerance ε.

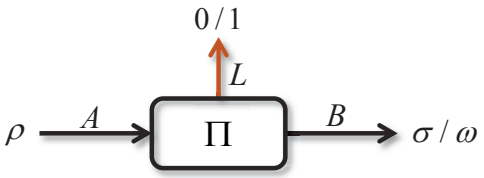


Figure 4.2 : The task of probabilistic coherence distillation with classical register L and quantum registers A, B. If the classical outcome is 0, then the quantum output is σ. Otherwise, it ouputs ω.

Definition 4.6 *For any triple (ρ, m, ε), the maximal success probability of coherence distillation under the operation class $\Omega \in \{\mathrm{SIO}, \mathrm{IO}, \mathrm{DIO}, \mathrm{MIO}\}$ is defined as*

$$P_\Omega\left(\rho \to \Psi_m, \varepsilon\right) = \max p \tag{4.22a}$$

$$\text{s.t. } \Pi_{A\to LB}\left(\rho\right) = p|0\rangle\langle 0|_L \otimes \sigma + \left(1-p\right)|1\rangle\langle 1|_L \otimes \omega, \tag{4.22b}$$

$$F\left(\sigma, \Psi_m\right)^2 \geq 1-\varepsilon,\ 0 \leq p \leq 1, \tag{4.22c}$$

$$\Pi \in \Omega,\ \sigma, \omega \in \mathcal{S}_=\left(B\right). \tag{4.22d}$$

If the distillation fails, we can perform a free operation to make the unwanted state ω completely mixed without changing the success probability. Thus, without loss of generality, we can take $\omega = \mathbb{1}/m$. Exploiting the fact that the target state Ψ_m is invariant under the twirling operation

$$\mathcal{T}\left(\sigma\right) = \frac{1}{m!}\sum_{i=1}^{m!} P_i \sigma P_i, \tag{4.23}$$

where P_i are all the permutations on the system of σ, we can also fix the optimal output state as $\sigma = \Psi_m^\varepsilon$ where

$$\Psi_m^\varepsilon := \left(1-\varepsilon\right)\Psi_m + \varepsilon\frac{\mathbb{1}-\Psi_m}{m-1}. \tag{4.24}$$

Specifically, for any optimal output state σ, we can further perform the free operation $\mathcal{T}$, which gives a new output state $\mathcal{T}\left(\sigma\right)$ always in the form of $a\Psi_m + b\left(\mathbb{1}-\Psi_m\right)/\left(m-1\right)$, where we can choose $a = 1-\varepsilon$ and $b = \varepsilon$ while keeping the fidelity with the target state and the optimal success probability unchanged. This allows us to write

$$P_\Omega\left(\rho \to \Psi_m, \varepsilon\right) = P_\Omega\left(\rho \to \Psi_m^\varepsilon, 0\right), \tag{4.25}$$

meaning that the maximal success probability of coherence distillation is the same as the maximal success probability of transforming the given state to the target Ψ_m^ε with fidelity one.

4.3.2 Computing the maximum distillation probability

We now set out to find efficiently computable expressions for the maximal distillation probability. Consider a generalization of the set Ω to the class Ω_{sub} of subnormalized quantum operations, that is, completely positive and trace-nonincreasing maps. Using this notation, we can conveniently express the maximal success probability as follows (see also [IP05, BG17]).

Proposition 4.7 *For any triplet (ρ, m, ε) and operation class Ω, the maximal success probability $P_\Omega\left(\rho \to \Psi_m, \varepsilon\right)$ is given by*

$$\max\left\{p \in \mathbb{R}_+ \;\middle|\; \mathcal{E}\left(\rho\right) = p \cdot \Psi_m^\varepsilon,\ \mathcal{E} \in \Omega_{\mathrm{sub}}\right\}. \tag{4.26}$$

Then, it holds that

$$P_{\Omega}\left(\rho \to \Psi_m, \varepsilon\right)^{-1} = \min\left\{t \in \mathbb{R}_+ \middle| \Psi_m^{\varepsilon} \in t \cdot \mathcal{S}_{\rho}\right\} \tag{4.27}$$

where $\mathcal{S}_{\rho} := \left\{\mathcal{E}\left(\rho\right) \middle| \mathcal{E} \in \Omega_{\text{sub}}\right\}$ *is the set of all output operators of* ρ *under the operation class* Ω_{sub}.

Proof For any quantum operation $\Pi_{A\to LB}$ such that $\Pi_{A\to LB}\left(\rho\right) = |0\rangle\langle 0|_L \otimes \mathcal{E}_0\left(\rho\right) + |1\rangle\langle 1|_L \otimes \mathcal{E}_1\left(\rho\right)$ where $\mathcal{E}_0$ and $\mathcal{E}_1$ are two subnormalized operations, we can show that $\Pi_{A\to LB} \in \Omega$ if and only if $\mathcal{E}_0, \mathcal{E}_1 \in \Omega_{\text{sub}}$ and $\mathcal{E}_0 + \mathcal{E}_1$ is trace preserving. Thus finding the optimal solution in the optimization (4.22) is equivalent to find the optimal subnormalized operations $\mathcal{E}_0$ and $\mathcal{E}_1$ such that $\mathcal{E}_0\left(\rho\right) = p \cdot \Psi_m^{\varepsilon}$, $\mathcal{E}_1\left(\rho\right) = \left(1-p\right)\mathbb{1}/m$ and $\mathcal{E}_0 + \mathcal{E}_1$ trace-preserving. Since we can always take $\mathcal{E}_1\left(\rho\right) = \left(\operatorname{Tr}\rho - \operatorname{Tr}\mathcal{E}_0\left(\rho\right)\right)\mathbb{1}/m$ without compromising the success probability, the maximal success probability of coherence distillation is only dependent on $\mathcal{E}_0$, and the result follows. ■

This result simplifies the optimization of the maximal success probability via subnormalized free operations, providing a geometric interpretation for the maximal success probability as a gauge function [Roc70, Reg18], as shown in Figure 4.3.

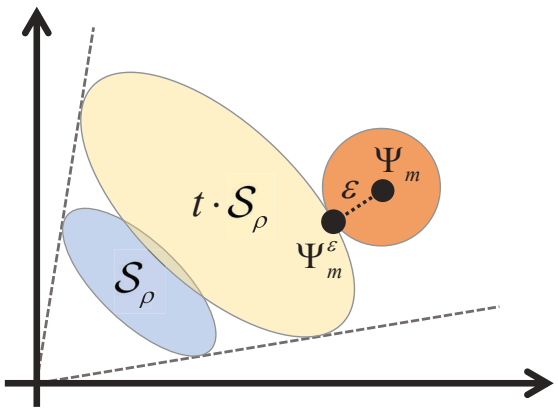


Figure 4.3 : Geometric interpretation of the maximal success probability of coherence distillation based on Eq. (4.27).

By further exploiting the symmetry of Ψ_m^{ε}, we can compute the maximal success probability under MIO/DIO via the following semidefinite programs.

Theorem 4.8 *For any triplet* (ρ, m, ε)*, the maximal success probability of distillation under*

MIO *and* DIO *are respectively given by*

$$
\begin{aligned}
P_{\mathrm{MIO}}\left(\rho \to \Psi_m, \varepsilon\right) = \max\ & \operatorname{Tr} G\rho \\
\text{s.t. } & \Delta\left(G\right) = m\Delta\left(C\right), && \text{(4.28a)} \\
& 0 \leq C \leq G \leq \mathbb{1}, && \text{(4.28b)} \\
& \operatorname{Tr} C\rho \geq (1-\varepsilon) \operatorname{Tr} G\rho, && \text{(4.28c)} \\
P_{\mathrm{DIO}}\left(\rho \to \Psi_m, \varepsilon\right) = \max\ & \operatorname{Tr} G\rho \\
\text{s.t. } & \text{Eqs. } (4.28a, 4.28b, 4.28c)\,,\ G = \Delta\left(G\right).
\end{aligned}
$$

Proof For any optimal subnormalized quantum operation $\mathcal{E}$ in Proposition 4.7, the operation $\widetilde{\mathcal{E}} = \mathcal{T} \circ \mathcal{E}$ is also optimal since Ψ_m^{ε} is invariant under the twirling operation $\mathcal{T}$ in (4.23). Denoting $J_{\mathcal{N}}$ as the Choi-Jamiołkowski matrix of the operation $\mathcal{N}$, we then have $J_{\widetilde{\mathcal{E}}} = C \otimes \Psi_m + D \otimes (\mathbb{1} - \Psi_m)$ for some operators C and D. Taking this form of Choi-Jamiołkowski matrix into the conditions of the optimization in Proposition 4.7, we obtain the desired forms of the SDPs. ■

These SDPs provide us with an efficient way to numerically calculate the maximal success probability for general triplets (ρ, m, ε), and allow us to obtain fundamental results about the capabilities of the different sets of operations for probabilistic coherence distillation.

It is known that operations in the class DIO can never increase the diagonal rank of a pure state, while it is known that MIO allow for the rank to increase [CG16b], suggesting that MIO is a much stronger class. It is therefore surprising that MIO and DIO have exactly the same power in the task of deterministic coherence distillation in Theorem 4.3, and that the two sets of operations lead to the same asymptotic transformation rates for all states [Chi18]. In the following, we will instead explicitly show crucial differences between MIO and DIO when one goes beyond deterministic transformations, highlighting the increased capabilities of MIO in probabilistic distillation, as well as establishing the limitations on coherence distillation in general.

Theorem 4.9 *For any triplet* $(\rho, m, 0)$ *with full-rank state* ρ *and* $m \geq 2$*, it holds*

$$
P_{\mathrm{MIO}}\left(\rho \to \Psi_m, 0\right) = 0. \tag{4.29}
$$

For any triplet $(\varphi, m, 0)$ *with coherent pure state* $|\varphi\rangle = \sum_{i=0}^{n-1} \varphi_i |i\rangle$*,* $\varphi_i \neq 0$ *and* $m, n \geq 2$*, it*

holds [2]

$$P_{\text{MIO}}\left(\varphi \to \Psi_m, 0\right) \geq \frac{n^2}{\sum_{i=0}^{n-1} |\varphi_i|^{-2}} \left\| \frac{n-m}{n-1}\widetilde{\varphi} + \frac{n\,(m-1)}{n-1}\Delta\left(\widetilde{\varphi}\right) \right\|_\infty^{-1} \geq \frac{n^2}{m\left(\sum_{i=0}^{n-1} |\varphi_i|^{-2}\right)} > 0 \quad (4.30)$$

where

$$|\widetilde{\varphi}\rangle := \frac{1}{\sqrt{s}} \sum_{i=0}^{n-1} \frac{\varphi_i}{|\varphi_i|^2} |i\rangle \quad \text{with} \quad s = \sum_{j=0}^{n-1} |\varphi_j|^{-2}. \quad (4.31)$$

Proof We prove the Eq. (4.29) first. From the SDP characterization in Theorem 4.8, we know that $G - C \geq 0$ and $\operatorname{Tr}\left(G - C\right)\rho = 0$ due to $\varepsilon = 0$. Since ρ is full-rank, we have $G = C$. Together with $\Delta\left(G\right) = m\Delta\left(C\right)$ and $m \geq 2$, we have $G = C = 0$, implying $P_{\text{MIO}}\left(\rho \to \Psi_m, 0\right) = 0$.

As for Eq. (4.30), let us choose

$$C = c\widetilde{\varphi}\,, \quad G = c\widetilde{\varphi} + \frac{(m-1)\,c}{n-1}\left(n\Delta\left(\widetilde{\varphi}\right) - \widetilde{\varphi}\right), \quad (4.32)$$

where

$$c = \left\| \frac{n-m}{n-1}\widetilde{\varphi} + \frac{n\,(m-1)}{n-1}\Delta\left(\widetilde{\varphi}\right) \right\|_\infty^{-1}. \quad (4.33)$$

We check the SDP constraints for MIO one by one. The first condition trivially holds by the construction. The last condition holds since $\langle\varphi|n\Delta\left(\widetilde{\varphi}\right) - \widetilde{\varphi}|\varphi\rangle = 0$, which implies that $\langle\varphi|C|\varphi\rangle = \langle\varphi|G|\varphi\rangle$. We now move on to the second condition. Clearly $C \geq 0$ and furthermore $G \geq C$ as follows from $\varphi \leq n\Delta\left(\varphi\right)$. To show that $G \leq \mathbb{1}$, just observe that

$$\|G\|_\infty = c\left\| \widetilde{\varphi} + \frac{m-1}{n-1}\left(n\Delta\left(\widetilde{\varphi}\right) - \widetilde{\varphi}\right) \right\|_\infty = c\left\| \frac{n-m}{n-1}\widetilde{\varphi} + \frac{n\,(m-1)}{n-1}\Delta\left(\widetilde{\varphi}\right) \right\|_\infty = 1\,. \quad (4.34)$$

Hence, C, G as defined above form a valid ansatz for the semidefinite program and

$$P_{\text{MIO}}\left(\varphi, m, 0\right) \geq \operatorname{Tr} G\varphi = \frac{n^2 c}{s}, \quad (4.35)$$

which yields the first lower bound in (4.30). As for the second bound, it suffices to show that $c \geq 1/m$, i.e. that $c^{-1} \leq m$. This can be done thanks to the triangle inequality:

$$c^{-1} = \left\| \widetilde{\varphi} + \frac{m-1}{n-1}\left(n\Delta\left(\widetilde{\varphi}\right) - \widetilde{\varphi}\right) \right\|_\infty \quad (4.36)$$

$$\leq \|\widetilde{\varphi}\|_\infty + (m-1)\left\| \frac{n\Delta\left(\widetilde{\varphi}\right) - \widetilde{\varphi}}{n-1} \right\|_\infty \quad (4.37)$$

$$\leq 1 + (m-1) = m\,, \quad (4.38)$$

[2] I would like to thank Lami Ludovico for letting me include these lower bounds in this thesis.

where we have used the fact that $\frac{n\Delta(\widetilde{\varphi})-\widetilde{\varphi}}{n-1}$ is a valid density matrix. ■

This result establishes a no-go theorem for coherence distillation, showing that no class of free operations preserving incoherent states can allow to distill any perfect coherence from a full-rank state, even probabilistically. Note that any generic density matrix has full rank, and so does Ψ_m^{ε} for any $\varepsilon > 0$. Thus

$$|P_{\text{MIO}}\left(\Psi_m^{\varepsilon} \to \Psi_m, 0\right) - P_{\text{MIO}}\left(\Psi_m \to \Psi_m, 0\right)| = 1, \tag{4.39}$$

even though Ψ_m^{ε} can be arbitrarily close to Ψ_m, implying that the maximal success probability is not continuous with respect to the input state. In practical terms, any amount of depolarizing noise will result in a full-rank state, so in a scenario where the coherent state Ψ_m is stored in a quantum memory exposed to depolarizing noise, it is impossible to recover it perfectly using free operations with any non-zero probability.

However, for any pure coherent state, it is always possible to probabilistically distill a maximally coherent state of arbitrary dimension via MIO. The first lower bound gives

$$P_{\text{MIO}}(\Psi_n \to \Psi_m, 0) \geq \frac{n-1}{m-1}, \quad \text{if } m > n. \tag{4.40}$$

Observe [3] that instead $P_{\text{DIO}}\left(\Psi_n \to \Psi_m, 0\right) = 0$ for $m > n$. This tells us that, as the dimension n increases, there are n-dimensional density matrices ρ_n such that

$$P_{\text{MIO}}\left(\rho_n \to \Psi_{n+1}, 0\right) \to 1, \tag{4.41}$$

$$P_{\text{DIO}}\left(\rho_n \to \Psi_{n+1}, 0\right) = 0, \tag{4.42}$$

for all n, i.e., P_{MIO} and P_{DIO} can exhibit an arbitrarily large gap. This shows that in the probabilistic distillation scenario, MIO can be much more powerful than DIO in general, in a stark contrast with the case of deterministic coherence distillation.

In the task of distilling maximally coherent qubit states Ψ_2, we can extend the above result and characterize analytically the maximal probability of distillation under DIO and MIO with arbitrary infidelity ε. In this particular case, MIO does not provide any advantage over DIO.

Lemma 4.10 ([LP01b, CG16b, DBG15]) *For any pure state $|\varphi\rangle = \sum_{i=1}^{n} \varphi_i |i\rangle$, it holds*

$$P_{\text{(S)IO}}\left(\varphi \to \Psi_m, 0\right) = \begin{cases} 0 & \text{if } \operatorname{rank}\Delta\left(\varphi\right) < m, \\ \min\limits_{k\in[1,m]} \dfrac{m}{k} \displaystyle\sum_{i=m-k+1}^{d} \varphi_i^2 & \text{otherwise.} \end{cases} \tag{4.43}$$

[3] It was proved in the following Lemma 4.15.

Proposition 4.11 *For any pure state* $|\varphi\rangle = \sum_{i=1}^{n} \varphi_i |i\rangle$ *with* $\varphi_1 \geq \ldots \geq \varphi_n > 0$, *and* $\Omega \in \{\mathrm{DIO}, \mathrm{MIO}\}$, *it holds that*

$$P_\Omega(\varphi \to \Psi_2, \varepsilon) = \begin{cases} 1 & \text{if } \varepsilon \geq \varepsilon_0(\varphi_1), \\ 2\left(1-\varphi_1^2\right)\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2 & \text{otherwise,} \end{cases} \tag{4.44}$$

$$\text{where} \quad \varepsilon_0(\varphi_1) = \begin{cases} 0 & \text{if } \varphi_1 \leq \frac{1}{\sqrt{2}} \\ \frac{1}{2} - \varphi_1\sqrt{1-\varphi_1^2} & \text{otherwise.} \end{cases} \tag{4.45}$$

Proof For the case of $\varphi_1 \leq \frac{1}{\sqrt{2}}$, we have

$$P_\Omega(\varphi \to \Psi_2, \varepsilon) \geq P_{\mathrm{SIO}}(\varphi \to \Psi_2, 0) = 1. \tag{4.46}$$

From the result in [RFWA18], we know that if $\varepsilon \geq \frac{1}{2} - \varphi_1\sqrt{1-\varphi_1^2}$, then $P_{\mathrm{DIO}}(\varphi \to \Psi_2, \varepsilon) = 1$. In the following, we therefore only consider the case $\varepsilon < \frac{1}{2} - \varphi_1\sqrt{1-\varphi_1^2}$. We prove this result by explicit constructing feasible solutions in both primal and dual SDPs. The primal SDP under DIO is given by

$$\begin{aligned} P_{\mathrm{DIO}}(\varphi \to \Psi_2, \varepsilon) = \max \ & \mathrm{Tr}\, G\varphi && \text{(4.47a)} \\ \text{s.t. } & \mathrm{Tr}\, C\varphi \geq (1-\varepsilon)\,\mathrm{Tr}\, G\varphi, && \text{(4.47b)} \\ & 0 \leq C \leq G \leq \mathbb{1}, && \text{(4.47c)} \\ & \Delta(G) = 2\Delta(C),\ G = \Delta(G). && \text{(4.47d)} \end{aligned}$$

We take the ansatz

$$G = \mathbb{1} - x|0\rangle\langle 0|, \quad C = \frac{1}{2}G + y\sum_{i=2}^{n} \varphi_1\varphi_n\left(|1\rangle\langle i| + |i\rangle\langle 1|\right). \tag{4.48}$$

Then we have

$$\mathrm{Tr}\, G\rho = 1 - x\varphi_1^2, \quad \mathrm{Tr}\, C\rho = \frac{1}{2}\left(1 - x\varphi_1^2\right) + 2y\varphi_1^2\left(1-\varphi_1^2\right), \tag{4.49}$$

and the spectrum of C and $G - C$ are given by

$$\left\{\underbrace{\frac{1}{2}, \cdots, \frac{1}{2}}_{n-2 \text{ fold}}, \frac{1}{2} - \frac{x}{4} - \sqrt{y^2\varphi_1^2\left(1-\varphi_1^2\right) + \frac{x^2}{16}}, \frac{1}{2} - \frac{x}{4} + \sqrt{y^2\varphi_1^2\left(1-\varphi_1^2\right) + \frac{x^2}{16}}\right\}. \tag{4.50}$$

Then we have the relaxation

$$\begin{aligned} P_{\mathrm{DIO}}(\varphi \to \Psi_2, \varepsilon) \geq \max \ & 1 - x\varphi_1^2 \\ \text{s.t. } & 4y\varphi_1^2\left(1-\varphi_1^2\right) = (1-2\varepsilon)\left(1 - x\varphi_1^2\right), \\ & 1 - x \geq 4y^2\varphi_1^2\left(1-\varphi_1^2\right), \\ & 0 \leq x \leq 1. \end{aligned} \tag{4.51}$$

By choosing

$$x = \frac{1 - 2\left(1 - \varphi_1^2\right)\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2}{\varphi_1^2}, \quad y = \frac{\left(\sqrt{1-\varepsilon}+\sqrt{\varepsilon}\right)^2}{2\varphi_1^2\left(1-2\varepsilon\right)}, \tag{4.52}$$

we can verify that this is a feasible solution and obtain

$$P_{\mathrm{DIO}}\left(\varphi \to \Psi_2, \varepsilon\right) \geq 2\left(1 - \varphi_1^2\right)\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2. \tag{4.53}$$

As for the dual problem, we consider the dual SDP under MIO,

$$\begin{aligned} P_{\mathrm{MIO}}\left(\varphi \to \Psi_2, \varepsilon\right) = \min \quad & \mathrm{Tr}\, Y \\ \text{s.t.} \quad & \left(1 - \left(1-\varepsilon\right)x\right)\varphi + X + \Delta\left(Z\right) \leq Y, \\ & x\varphi - X - 2\Delta\left(Z\right) \leq 0, \\ & x \geq 0, X \geq 0, Y \geq 0 \end{aligned} \tag{4.54}$$

Taking

$$x = \frac{\left(\sqrt{1-\varepsilon}+\sqrt{\varepsilon}\right)^2}{\sqrt{1-\varepsilon}\sqrt{\varepsilon}\left(1-2\varepsilon\right)}, \quad Y = 2\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2 \left(\mathbb{1} - |0\rangle\langle 0|\right)\varphi\left(\mathbb{1} - |0\rangle\langle 0|\right), \tag{4.55}$$

$$Z = \frac{2}{1-2\varepsilon}|0\rangle\langle 0|\varphi|0\rangle\langle 0|, \quad X = Y - \Delta\left(Z\right) - \left(1 - \left(1-\varepsilon\right)x\right)\varphi, \tag{4.56}$$

we can verify that $\{x, X, Y, Z\}$ is a valid feasible solution. Thus

$$P_{\mathrm{MIO}}\left(\varphi \to \Psi_2, \varepsilon\right) \leq 2\left(1 - \varphi_1^2\right)\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2. \tag{4.57}$$

Combining Eqs. (4.53) and (4.57), we have the desired result. ■

Remark 4.12 The function ε_0 can be related to the so-called m-distillation norm [RFWA18], characterizing the fidelity of deterministic distillation.

Remark 4.13 Using this analytical result, we can give a concrete example to show that the probabilistic distillation framework can outperform the deterministic one. Suppose we need to distill a maximally coherent qubit state Ψ_2 from the input state $|\varphi\rangle = \left(3|0\rangle + |1\rangle\right)/\sqrt{10}$ with acceptable fidelity at least 0.9. The input state becomes useless in the deterministic scenario, since the maximal fidelity achievable via deterministic protocols is given by 0.8. However, probabilistic operations allow us to achieve the required distillation fidelity with success probability 0.5, which is significantly better than always yielding nothing as in the deterministic case. In the other case, if the acceptable fidelity is 0.8, we can gain better performance (higher fidelity) by compromising a bit of the success probability despite the fact that deterministic protocols are sufficient to accomplish the task. Such a scenario can be dubbed "gambling with coherence", based on similar terminology used in [BBPS96, LP01b].

4.3.3 Relation between distillation fidelity and probability

For any given input state ρ and target state dimension m, the maximal success probability is only dependent on the transformation fidelity. The higher the fidelity we require from our output state, the lower the probability that we will succeed. Intuitively, one would expect that the success probability will smoothly decrease as the fidelity increases; however, we find a phenomenon which breaks this expectation. Specifically, we will now show that the maximal success probability may happen to vanish if the fidelity goes beyond some threshold. This is analogous to the strong converse theorem in channel coding theory [Wol78, ON99, Win99], which says that the coding success probability goes to zero if the coding rate exceeds the capacity of the channel. Note that this phenomenon will not occur in the case of distillation from pure input states under MIO due to Theorem 4.9. In the following result, we completely characterize this property for pure input states under DIO.

Proposition 4.14 *For any pure state $|\varphi\rangle = \sum_{i=1}^{n} \varphi_i |i\rangle$ with nonzero coefficients φ_i, it holds that*

$$P_{\mathrm{DIO}}\left(\varphi \to \Psi_m, \varepsilon\right) \begin{cases} > 0 & \text{if } n \geq m \text{ or if } n < m \text{ and } \varepsilon \geq 1 - \frac{n}{m}, \\ = 0 & \text{if } n < m \text{ and } \varepsilon < 1 - \frac{n}{m}. \end{cases} \tag{4.58}$$

Proof If $n \geq m$, we know that

$$P_{\mathrm{DIO}}\left(\varphi \to \Psi_m, \varepsilon\right) \geq P_{\mathrm{SIO}}\left(\varphi \to \Psi_m, 0\right) > 0, \tag{4.59}$$

where the second inequality follows from Eq. (4.43).

Note that if $P_{\Omega}\left(\sigma_1 \to \sigma_2, 0\right) = 1$, then $P_{\Omega}\left(\rho \to \sigma_2, 0\right) \geq P_{\Omega}\left(\rho \to \sigma_1, 0\right)$ since we can first transform ρ to σ_1 perfectly and then get σ_2. If $\varepsilon \geq 1 - \frac{n}{m}$, we have

$$P_{\mathrm{DIO}}\left(\Psi_n \to \Psi_m^{\varepsilon}, 0\right) = P_{\mathrm{DIO}}\left(\Psi_n \to \Psi_m, \varepsilon\right) = 1. \tag{4.60}$$

The first equality follows from the fact that $P_{\Omega}\left(\rho \to \Psi_m, \varepsilon\right) = P_{\Omega}\left(\rho \to \Psi_m^{\varepsilon}, 0\right)$. The second equality follows from Lemma 4.15 below. Then

$$\begin{aligned} P_{\mathrm{DIO}}\left(\varphi \to \Psi_m, \varepsilon\right) &= P_{\mathrm{DIO}}\left(\varphi \to \Psi_m^{\varepsilon}, 0\right) \\ &\geq P_{\mathrm{DIO}}\left(\varphi \to \Psi_n, 0\right) \\ &\geq P_{\mathrm{SIO}}\left(\varphi \to \Psi_n, 0\right) \\ &> 0. \end{aligned} \tag{4.61}$$

The first inequality follows from Eq. (4.60). The last inequality follows from Eq. (4.43). If $\varepsilon \leq 1 - \frac{n}{m}$, we have $P_{\mathrm{DIO}}\left(\varphi \to \Psi_m, \varepsilon\right) \leq P_{\mathrm{DIO}}\left(\Psi_n \to \Psi_m, \varepsilon\right) = 0$, where the second equality follows from Lemma 4.15. ■

This result shows that if the input dimension is larger than the target state dimension, there is always a tradeoff as expected between the maximal success probability and the transformation fidelity. However, the tradeoff curve will always be truncated at the fidelity threshold $\frac{n}{m}$. At the point of $\varepsilon = 1 - \frac{n}{m}$, demanding a slightly higher fidelity will make the probabilistic distillation impossible, as shown in Figure 4.4.

Lemma 4.15 *For any integer $n \leq m$, it holds that*

$$P_{\mathrm{DIO}}\left(\Psi_n \to \Psi_m, \varepsilon\right) = \begin{cases} 1, & \varepsilon \geq 1 - \frac{n}{m}, \\ 0, & \varepsilon < 1 - \frac{n}{m}. \end{cases} \tag{4.62}$$

Proof For $\varepsilon \geq 1 - \frac{n}{m}$, we can take feasible solution $G = \mathbb{1}$, $C = \frac{n}{m}\psi_n$, which gives feasible value 1 in the primal problem. For $\varepsilon < 1 - \frac{n}{m}$, we can take feasible solution $x = \frac{1}{1-\frac{n}{m}-\varepsilon}$, $X = Y = 0$, $Z = \frac{1}{m-n-m\varepsilon}\mathbb{1}$, $W = \frac{1}{m-n-m\varepsilon}\left(n\psi_n - \mathbb{1}\right)$, which gives feasible value 0 in the dual problem. ■

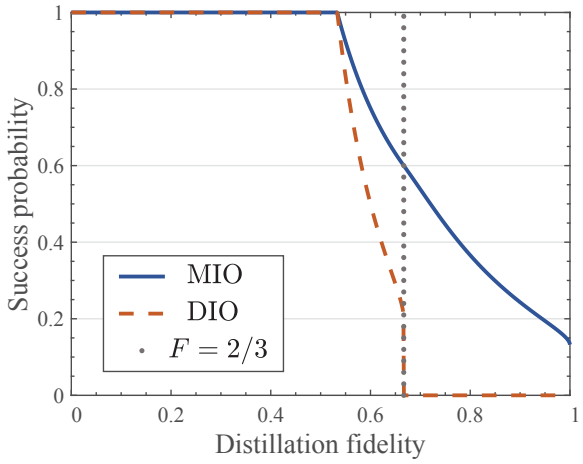


Figure 4.4 : An example of the threshold in the tradeoff between the fidelity ($F = 1 - \varepsilon$) and the success probability of distillation (p) for the transformation $\left(|0\rangle + 3|1\rangle\right)/\sqrt{10} \to \Psi_3$. The auxiliary dotted line $F = 2/3$ is perpendicular to the horizontal axis.

4.3.4 Probabilistic distillation with catalytic assistance

A more general coherence distillation setting is to consider the scenario with catalytic assistance [JP99], where the input to the protocol consists of the resource state ρ together with another state γ (catalyst). As suggested by its name, we need to reproduce γ untouched [4] in the

[4] In the published paper [FWL+18], we consider a more general framework which allows small perturbation of the returning catalyst.

output regardless of whether the distillation process succeeds or not. In [BSW16], the authors study catalytic coherence transformations without enforcing the preservation of the catalyst when the transformation fails — it is then not surprising that catalytic assistance improves the success probability, since we take the risk to sacrifice our catalyst. However, we can show that using catalysts can enhance probabilistic distillation even when we require them to be reproduced regardless of the outcome.

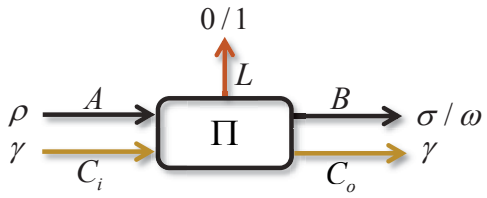


Figure 4.5 : The task of probabilistic coherence distillation with catalytic assistance, where L is the classical register and A, B, C_i, C_o are quantum registers.

Formally, we denote the the catalysis-assisted maximal success probability of coherence distillation under the operation class Ω as $P_\Omega\left(\rho \xrightarrow{\gamma} \Psi_m, \varepsilon\right)$, which is given by

$$P_\Omega\left(\rho \xrightarrow{\gamma} \Psi_m, \varepsilon\right) = \max p \tag{4.63a}$$

$$\text{s.t. } \Pi\left(\rho \otimes \gamma\right) = \left(p|0\rangle\langle 0| \otimes \sigma + (1-p)\,|1\rangle\langle 1| \otimes \omega\right) \otimes \gamma, \tag{4.63b}$$

$$F\left(\sigma, \Psi_m\right)^2 \geq 1 - \varepsilon,\ 0 \leq p \leq 1, \tag{4.63c}$$

$$\Pi \in \Omega,\ \sigma, \omega \in \mathcal{S}_=\left(B\right). \tag{4.63d}$$

Since we can always choose not to interact with the catalyst, it is clear that

$$P_\Omega\left(\rho \xrightarrow{\gamma} \Psi_m, \varepsilon\right) \geq P_\Omega\left(\rho \to \Psi_m, \varepsilon\right). \tag{4.64}$$

Taking as an example the two-qubit state $\rho = \frac{1}{2}\left(v_1 + v_2\right)$ with

$$|v_1\rangle = \frac{1}{2}\left(|00\rangle - |01\rangle - |10\rangle + |11\rangle\right), \tag{4.65}$$

$$|v_2\rangle = \frac{1}{5\sqrt{2}}\left(2|00\rangle + 6|01\rangle - 3|10\rangle + |11\rangle\right), \tag{4.66}$$

it turns out that the catalytic assistance of $\gamma = \Psi_2$ can enhance the success probability (at least 12%) of distilling one coherent bit via DIO reliably ($\varepsilon \leq 0.01$). This example shows that the maximally coherent state can be used as a catalyst, manifesting a difference with the case of deterministic state transformation, where no transformation can be catalyzed by a maximally coherent state [JP99, DBG15].

Note that the catalyst-assisted probability under DIO can still be written as an SDP. We show the difference between $P_{\mathrm{DIO}}\left(\rho \xrightarrow{\Psi_2} \Psi_2, 0.01\right)$ and $P_{\mathrm{DIO}}\left(\rho \to \Psi_2, 0.01\right)$ in the following Figure 4.6. On the right hand side, the enhancement ratio is given by $\left[P_{\mathrm{DIO}}\left(\rho \xrightarrow{\Psi_2} \Psi_2, 0.01\right) - P_{\mathrm{DIO}}\left(\rho \to \Psi_2, 0.01\right)\right] / P_{\mathrm{DIO}}\left(\rho \to \Psi_2, 0.01\right)$.

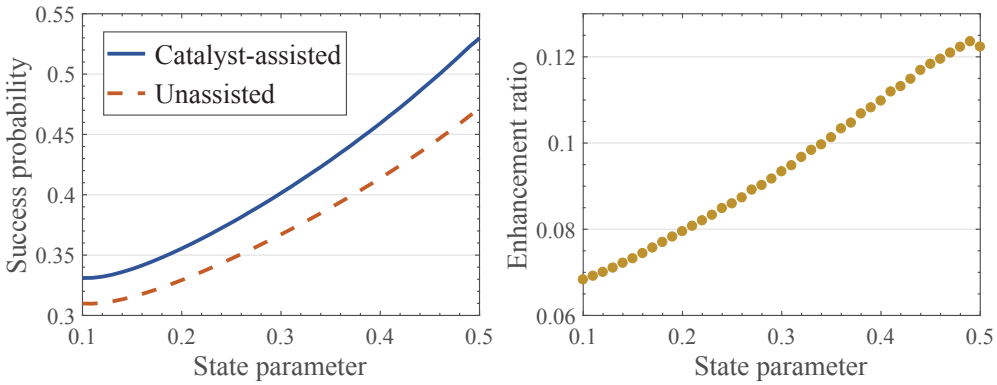


Figure 4.6 : Examples of catalyst-assisted probabilistic coherence distillation.

4.4 Discussion

4.4.1 Summary of results

1. For deterministic coherence distillation, MIO and DIO have the same power:

$$F_{\mathrm{MIO}}(\rho, m) = F_{\mathrm{DIO}}(\rho, m), \quad C_{d,\mathrm{MIO}}^{(1),\varepsilon}(\rho) = C_{d,\mathrm{DIO}}^{(1),\varepsilon}(\rho),$$
$$C_{d,\mathrm{DIO}}(\rho) = C_{d,\mathrm{MIO}}(\rho) = C_r(\rho).$$

   For probabilistic coherence distillation, the probability gap under MIO and DIO can be arbitrarily large.

2. Connection between the one-shot coherence distillation and the quantum hypothesis testing: for $\Omega \in \{\mathrm{MIO}, \mathrm{DIO}\}$ it holds

$$C_{d,\Omega}^{(1),\varepsilon}(\rho) = \min_{X \in \mathcal{J}} D_H^{\varepsilon}(\rho \| X) - \delta,$$

   where $\delta \in [0, 1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

3. The maximal success probability of coherence distillation is given by a gauge function

$$P_{\Omega}(\rho \to \Psi_m, \varepsilon)^{-1} = \min\left\{t \in \mathbb{R}_+ \middle| \Psi_m^{\varepsilon} \in t \cdot \mathcal{S}_{\rho}\right\}$$

   where $\mathcal{S}_{\rho} := \left\{\mathcal{E}(\rho) \,\middle|\, \mathcal{E} \in \Omega_{\mathrm{sub}}\right\}$ is the set of all output operators of ρ under the operation class Ω_{sub}.

4. No-go theorem: no class of free operations preserving incoherent states can allow to distill any perfect coherence from a full-rank state, even probabilistically.

5. The maximal success probability of coherence distillation may happen to vanish if the fidelity goes beyond some threshold.

6. The maximally coherent state can be used as catalyst to enhance the success probability, which is contrast to the deterministic scenario.

4.4.2 Outlook

Our work opens new perspectives for the investigation of quantum coherence manipulation with finite resources, unveils several new features of coherence from a resource theoretic viewpoint, and contributes to an increased understanding of the fundamental properties of the different sets of free operations. More generally, our results establish fundamental limitations to the processing of coherence in realistic settings, which is of direct relevance to applications in quantum information processing and quantum technology tasks that exploit coherence as a resource. It would be of interest to analyze as well the task of probabilistic coherence dilution under different free operations, whose deterministic case was studied in [ZLY+18]. Another interesting perspective for future work may be to apply the framework of probabilistic coherence distillation developed here to the study of other important resource theories, such as those of asymmetry, magic states, and thermodynamics.

Chapter 5

Quantum channel simulation

Quantum channels are valuable resources for sending quantum information between different quantum nodes. As a fundamental problem in quantum information theory, quantum channel simulation considers using other quantum resources to simulate a given quantum channel. In this chapter, we study the problem of simulating a quantum channel via quantum coherence and another quantum channel, respectively.

5.1 Introduction

5.1.1 Background

As a fundamental ingredient of quantum information science, a quantum channel acts as an irreplaceable component during the information transmission process, such as exchanging data among different nodes in a quantum computer or sending qubits across the quantum internet. Since every channel has different power of sending information, the quantification of the information theoretic power of a channel becomes one of the most fundamental problems in information theory. One particular interest is the study of *channel capacities*, that is to ask how many quantum (classical) bits a channel can reliably transmit per channel use. However, we can unify the problem of channel capacities into a broader framework, i.e., *channel simulation* — it studies using other resources to imitate the functionality of a channel. In this framework, the process of sending quantum (classical) bits reliably using a quantum channel can be understood as a simulation of quantum (classical) noiseless channels [KW04]. Depending on the different resources available, the channel simulation problem has many variants.

For classical channels, Shannon's noisy channel coding theorem determines the capability of noisy classical channels to simulate noiseless ones [Sha48]. Dual to this famous coding theorem, the "reverse Shannon theorem" concerns the use of noiseless channels to simulate noisy ones as well as the use of a channel to simulate another [BSST02]. Specifically, ev-

ery channel can be simulated using an amount of classical communication equal to the capacity of the channel when there is free shared randomness between A and B in the asymptotic setting [BSST02]. For quantum channels, the case when A and B share an unlimited amount of entanglement has been completely solved by the *quantum reverse Shannon theorem* (QRST) [BDH+14, BCR11], which states that the rate to optimally simulate a quantum channel in the asymptotic setting is determined by its entanglement-assisted classical capacity. In the zero-error setting [Sha56], using one channel to simulate another exactly with the aid of non-signalling correlations has been studied recently in [CLMW11, DW16, WD16b]. Moreover, quantum channel simulations via quantum coherence and quantum entanglement have been studied in [BBCW13, BGMW17].

In realistic settings, the number of simulated channels and the amount of resource for simulation are necessarily limited. Therefore, it is important to characterize how well we can simulate a quantum channel with finite resources. The first step in this direction is to consider the one-shot setting. One-shot analysis has recently attracted great interest in classical information theory (see, e.g., [PPV10, Hay09]) and quantum information theory (see, e.g., [Tom12, DH13, BD10a, MW14, WXD18, RR11, DRRW13, BCT16, AJM+16]). In one-shot information theory, the smooth max-information of a quantum state [BCR11] and its generalizations [CBR14] are all basic and useful quantities, which have various applications in quantum rate distortion theory as well as the physics of quantum many-body systems. In this chapter we will introduce some new entropies naturally arisen from the simulation process.

5.1.2 Outline

In section 5.2 we study the framework of quantum channel simulation via quantum coherence, discussing the simulation via MIO in details. We first show that the minimum error of coherence simulation and the one-shot coherence simulation cost under MIO can be both efficiently calculated via semidefinite programs. Second, we prove that the one-shot zero-error coherence simulation cost is additive. Notably the zero-error coherence simulation cost is exactly equal to the maximal coherence generated from the channel, i.e., the cohering power of the channel. This leads to a conjecture that the coherence simulation cost of a channel is equal to its coherence generating capacity — an analogous result of the quantum reverse Shannon theorem. Finally, we introduce a channel's version of the max-relative entropy and builds its connection with the coherence simulation cost.

In section 5.3 we study the general framework of quantum channel simulation via another quantum channel, that is, the ability of a quantum channel to simulate another one using different classes of codes. First, we show that the minimum error of simulation and the one-

shot quantum simulation cost under non-signalling assisted codes are efficiently computable via semidefinite programming. Second, we introduce the channel's smooth max-information, which can be seen as a one-shot generalization of the mutual information of a quantum channel. We provide an exact operational interpretation of the channel's smooth max-information as the one-shot quantum simulation cost. We further introduce the channel's log-robustness and elaborate its relation with the channel's max-information. Third, we derive the asymptotic equipartition property (AEP) of the channel's smooth max-information, i.e., it converges to the quantum mutual information of the channel in the independent and identically distributed asymptotic limit. This implies the quantum reverse Shannon theorem in the presence of non-signalling correlations. As applications, we explore finite blocklength simulation cost of fundamental quantum channels and provide both numerical and analytical solutions.

5.2 Channel simulation via quantum coherence

5.2.1 General framework

In this section we consider simulating a quantum channel via quantum coherence. The general framework is shown as Figure 5.1. For a given quantum channel $\mathcal{N}_{A\to B}$, our task is to simulate this channel with maximally coherent states and free operations. In the following, we denote $\Psi_k = (1/k)\sum_{i,j=0}^{k-1} |i\rangle\langle j|$ as the maximally coherent state with dimension k. Then the effective channel is given by

$$\widetilde{\mathcal{N}}_{A\to B}(\rho) = \Pi_{AR\to B}(\rho \otimes \Psi_k), \quad \forall\, \rho \in \mathcal{S}_{=}(A). \tag{5.1}$$

We use the diamond norm distance between the effective channel and the target channel to characterize the performance of a simulation and introduce the minimum error of coherence simulation as follows.

Figure 5.1 : The task of channel simulation via quantum coherence.

Definition 5.1 *For a given quantum channel $\mathcal{N}_{A\to B}$, the minimum error of coherence simula-*

tion via Ψ_k under the operation class Ω is defined as

$$\omega_{\Omega}\left(\Psi_k, \mathcal{N}\right) := \frac{1}{2} \inf \left\| \widetilde{\mathcal{N}} - \mathcal{N} \right\|_{\diamondsuit} \tag{5.2a}$$

$$\text{s.t. } \widetilde{\mathcal{N}}_{A\to B}\left(\rho\right) = \Pi_{AR\to B}\left(\rho \otimes \Psi_k\right), \forall\, \rho \in \mathcal{S}_{=}\left(A\right), \tag{5.2b}$$

$$\Pi_{AR\to B} \in \Omega. \tag{5.2c}$$

where $\left\|\mathcal{F}\right\|_{\diamondsuit} := \sup_{k\in\mathbb{N}} \left\|\mathcal{F} \otimes \mathrm{id}_k\right\|_1$ denotes the diamond norm and $\left\|\cdot\right\|_1$ is induced by the trace norm. The free operation class $\Omega \in \{\mathrm{PIO}, \mathrm{SIO}, \mathrm{IO}, \mathrm{DIO}, \mathrm{MIO}\}$.

Definition 5.2 *For a given quantum channel $\mathcal{N}_{A\to B}$, the one-shot ε-error coherence simulation cost under the operation class Ω is defined as*

$$S_{c,\Omega}^{(1),\varepsilon}\left(\mathcal{N}\right) := \log \min \left\{ k \in \mathbb{N}_+ \,\middle|\, \omega_{\Omega}\left(\Psi_k, \mathcal{N}\right) \leq \varepsilon \right\}. \tag{5.3}$$

The asymptotic coherence simulation cost is defined as

$$S_{c,\Omega}\left(\mathcal{N}\right) := \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} S_{c,\Omega}^{(1),\varepsilon}\left(\mathcal{N}^{\otimes n}\right). \tag{5.4}$$

5.2.2 One-shot characterizations

Based on above definitions, we will discuss the case for $\Omega = \mathrm{MIO}$ in details.

Proposition 5.3 *For a given quantum channel $\mathcal{N}_{A\to B}$ and maximally coherent state Ψ_k, the minimum error of coherence simulation under* MIO *can be given by the following SDP,*

$$\omega_{\mathrm{MIO}}\left(\Psi_k, \mathcal{N}\right) = \inf \gamma \tag{5.5a}$$

$$\text{s.t. } \mathrm{Tr}_B\, Y_{AB} \leq \gamma \mathbb{1}_A, \tag{5.5b}$$

$$Y_{AB} \geq J_{\widetilde{\mathcal{N}}} - J_{\mathcal{N}},\ Y_{AB} \geq 0, \tag{5.5c}$$

$$J_{\widetilde{\mathcal{N}}} \geq 0,\ \mathrm{Tr}_B\, J_{\widetilde{\mathcal{N}}} = \mathbb{1}_A, \tag{5.5d}$$

$$J_{\widetilde{\mathcal{N}}} \leq k \cdot J_{\mathcal{M}}, \tag{5.5e}$$

$$J_{\mathcal{M}} \geq 0,\ \mathrm{Tr}_B\, J_{\mathcal{M}} = \mathbb{1}_A, \tag{5.5f}$$

$$\mathrm{Tr}_A\, J_{\mathcal{M}} |i\rangle\langle i|_A \in \mathcal{I},\ \forall\, i. \tag{5.5g}$$

Proof From the Choi-Jamiołkowski isomorphism, we have

$$\widetilde{\mathcal{N}}_{A\to B}\left(\rho\right) = \mathrm{Tr}_A\, J_{\widetilde{\mathcal{N}}} \cdot \rho_A^T, \tag{5.6}$$

$$\Pi_{AR\to B}\left(\rho \otimes \Psi_k\right) = \mathrm{Tr}_{AR}\, J_{\Pi} \cdot \rho_A^T \otimes \Psi_k^T = \mathrm{Tr}_A\left(\mathrm{Tr}_R\, J_{\Pi} \cdot \mathbb{1}_A \otimes \Psi_k^T\right) \rho_A^T. \tag{5.7}$$

Since $\widetilde{\mathcal{N}}_{A\to B}\left(\rho\right) = \Pi_{AR\to B}\left(\rho \otimes \Psi_k\right)$ holds for any state $\rho \in \mathcal{S}_{=}\left(A\right)$, we have the Choi-Jamiołkowski operator relation between the effective channel $\widetilde{\mathcal{N}}$ and the free operation Π as

$J_{\widetilde{\mathcal{N}}} = \mathrm{Tr}_R\, J_\Pi \cdot \Psi_k^T$. From the definition of minimum error of coherence simulation and the SDP (2.82) for diamond norm, we obtain

$$\omega_{\mathrm{MIO}}\left(\Psi_k, \mathcal{N}\right) = \inf \| \mathrm{Tr}_B\, Y_{AB} \|_\infty \tag{5.8a}$$

$$\text{s.t. } Y_{AB} \geq J_{\widetilde{\mathcal{N}}} - J_{\mathcal{N}}, Y_{AB} \geq 0, \tag{5.8b}$$

$$J_{\widetilde{\mathcal{N}}} = \mathrm{Tr}_R\, J_\Pi \cdot \Psi_k^T, \tag{5.8c}$$

$$\Pi_{AR\to B} \in \mathrm{MIO}. \tag{5.8d}$$

If J_Π is an optimal solution of (5.8), then $\frac{1}{|R|!}\sum_\pi U_\pi J_\Pi U_\pi$ is also an optimal solution, where U_π are all the permutation operator on the system R. Thus without loss of generality, we can take J_Π in the form of

$$J_\Pi = P_{AB} \otimes \mathbb{1}_R + Q_{AB} \otimes \left(\Psi_k - \frac{1}{k}\mathbb{1}_R\right). \tag{5.9}$$

Since Π is a CP map, we have

$$J_\Pi = \left(P_{AB} - \frac{1}{k}Q_{AB}\right) \otimes \left(\mathbb{1}_R - \Psi_k\right) + \left(P_{AB} + \left(1 - \frac{1}{k}\right) Q_{AB}\right) \otimes \Psi_k \geq 0, \tag{5.10}$$

which is equivalent to $P_{AB} - \frac{1}{k}Q_{AB} \geq 0$ and $P_{AB} + \left(1 - \frac{1}{k}\right) Q_{AB} \geq 0$. Since Π is a TP map, we have

$$\mathrm{Tr}_B\, J_\Pi = \mathrm{Tr}_B\, P_{AB} \otimes \mathbb{1}_R + \mathrm{Tr}_B\, Q_{AB} \otimes \left(\Psi_k - \frac{1}{k}\mathbb{1}_R\right) = \mathbb{1}_{AR}, \tag{5.11}$$

which is equivalent to $\mathrm{Tr}_B\, Q_{AB} = 0$ and $\mathrm{Tr}_B\, P_{AB} = \mathbb{1}_A$. Since $\Pi \in \mathrm{MIO}$, for any input state $|ij\rangle\langle ij|_{AR}$, we have

$$\Pi\left(|ij\rangle\langle ij|_{AR}\right) = \mathrm{Tr}_{AR}\left(P_{AB} \otimes \mathbb{1}_R + Q_{AB} \otimes \left(\Psi_k - \frac{1}{k}\mathbb{1}_R\right)\right)|ij\rangle\langle ij|_{AR} \tag{5.12}$$

$$= \mathrm{Tr}_A\, P_{AB}|i\rangle\langle i|_A \in \mathcal{I}. \tag{5.13}$$

By direct calculation, we also have $J_{\widetilde{\mathcal{N}}} = \mathrm{Tr}_R\, J_\Pi \cdot \Psi_k^T = P + \left(1 - \frac{1}{k}\right) Q$. Eliminating the variable Q and denoting $J_{\mathcal{M}} = P_{AB}$, we have the desired result. ■

From Proposition 5.3, we denote $J_{\widetilde{\mathcal{M}}} = kJ_{\mathcal{M}}$ and have the following result.

Proposition 5.4 *For a given quantum channel $\mathcal{N}_{A\to B}$, the one-shot ε-error coherence simula-*

tion cost under MIO *is given by the following SDP,*

$$S_{c,\mathrm{MIO}}^{(1),\varepsilon}\left(\mathcal{N}\right) = \log \min \lceil \mathrm{Tr}\, J_{\widetilde{\mathcal{M}}} \,/\, d_A \rceil \tag{5.14a}$$

$$\text{s.t. } 0 \leq J_{\widetilde{\mathcal{N}}} \leq J_{\widetilde{\mathcal{M}}}, \tag{5.14b}$$

$$\mathrm{Tr}_B\, J_{\widetilde{\mathcal{N}}} = \mathbb{1}_A,\ \mathrm{Tr}_B\, J_{\widetilde{\mathcal{M}}} = \mathrm{Tr}\, J_{\widetilde{\mathcal{M}}} \,/\, d_A \cdot \mathbb{1}_A, \tag{5.14c}$$

$$\mathrm{Tr}_A\, J_{\widetilde{\mathcal{M}}}\ (|i\rangle\langle i| \otimes \mathbb{1}_B) \in \mathcal{I}_H, \tag{5.14d}$$

$$\mathrm{Tr}_B\, Y_{AB} \leq \varepsilon \mathbb{1}_A, \tag{5.14e}$$

$$Y_{AB} \geq J_{\widetilde{\mathcal{N}}} - J_{\mathcal{N}}, Y_{AB} \geq 0, \tag{5.14f}$$

where $\mathcal{I}_H$ *is the set of diagonal Hermitian operators.*

Remark 5.5 It is easy to check that $\delta = \log\lceil x \rceil - \log x \in [0,1]$ for any $x \geq 1$. Thus we can use the least constant $\delta \in [0,1]$ to adjust the r.h.s. of Eq. (5.14) to be the logarithm of an integer. That is,

$$S_{c,\mathrm{MIO}}^{(1),\varepsilon}\left(\mathcal{N}\right) = \delta + \log \min \mathrm{Tr}\, J_{\widetilde{\mathcal{M}}} \,/\, d_A \tag{5.15a}$$

$$\text{s.t. } 0 \leq J_{\widetilde{\mathcal{N}}} \leq J_{\widetilde{\mathcal{M}}}, \tag{5.15b}$$

$$\mathrm{Tr}_B\, J_{\widetilde{\mathcal{N}}} = \mathbb{1}_A,\ \mathrm{Tr}_B\, J_{\widetilde{\mathcal{M}}} = \mathrm{Tr}\, J_{\widetilde{\mathcal{M}}} \,/\, d_A \cdot \mathbb{1}_A, \tag{5.15c}$$

$$\mathrm{Tr}_A\, J_{\widetilde{\mathcal{M}}}\ (|i\rangle\langle i| \otimes \mathbb{1}_B) \in \mathcal{I}_H, \tag{5.15d}$$

$$\mathrm{Tr}_B\, Y_{AB} \leq \varepsilon \mathbb{1}_A, \tag{5.15e}$$

$$Y_{AB} \geq J_{\widetilde{\mathcal{N}}} - J_{\mathcal{N}}, Y_{AB} \geq 0, \tag{5.15f}$$

where $\delta \in [0,1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

Besides the above SDP characterization, we have the following simpler refinement for the zero-error case.

Proposition 5.6 *For a given quantum channel* $\mathcal{N}_{A\to B}$*, the one-shot zero-error coherence simulation cost under* MIO *is given by*

$$S_{c,\mathrm{MIO}}^{(1),0}\left(\mathcal{N}\right) = \max_i C_{\max}\left(\mathcal{N}\left(|i\rangle\langle i|\right)\right) + \delta, \tag{5.16}$$

where $C_{\max}\left(\rho\right) := \min_{\sigma\in\mathcal{I}} D_{\max}\left(\rho\|\sigma\)$*,* $\mathcal{I} = \{\rho \geq 0 \mid \mathrm{Tr}\,\rho = 1, \Delta\left(\rho\right) = \rho\}$*, and* $\delta \in [0,1]$ *is the least constant such that the r.h.s. is the logarithm of an integer.*

Proof We prove this result via SDP duality. From Proposition 5.4, we take $\varepsilon = 0$ and obtain

the following SDP,

$$S_{c,\mathrm{MIO}}^{(1),0}(\mathcal{N}) = \delta + \log \min \operatorname{Tr} J_{\widetilde{\mathcal{M}}} \,/\, d_A \tag{5.17a}$$

$$\text{s.t. } 0 \leq J_{\mathcal{N}} \leq J_{\widetilde{\mathcal{M}}}, \tag{5.17b}$$

$$\operatorname{Tr}_B J_{\widetilde{\mathcal{M}}} = \operatorname{Tr} J_{\widetilde{\mathcal{M}}}/d_A \cdot \mathbb{1}_A, \tag{5.17c}$$

$$\operatorname{Tr}_A J_{\widetilde{\mathcal{M}}} |i\rangle\langle i| \otimes \mathbb{1}_B \in \mathcal{I}_H. \tag{5.17d}$$

Its dual SDP is given by

$$S_{c,\mathrm{MIO}}^{(1),0}(\mathcal{N}) = \delta + \log \max \ \operatorname{Tr} J_{\mathcal{N}} X \tag{5.18a}$$

$$\text{s.t. } X = Y \otimes \mathbb{1}_B + \sum\nolimits_{i=0}^{d_A-1} |i\rangle\langle i| \otimes Z_i, \tag{5.18b}$$

$$\operatorname{Tr} Y = 1,\ \Delta(Y) = Y,\ \Delta(Z_i) = 0, \tag{5.18c}$$

$$X \geq 0, Y, Z_i \text{ Hermitian}. \tag{5.18d}$$

For convenience, we denote $r = \max_i C_{\max}(\mathcal{N}(|i\rangle\langle i|))$. According to the dual SDP of $C_{\max}$ presented later in Eq. (5.31), r can also be given by an SDP,

$$r = \log \max_i \max \ \operatorname{Tr} J_{\mathcal{N}} |i\rangle\langle i| \otimes P_i \tag{5.19a}$$

$$\text{s.t. } P_i = \mathbb{1}_B + Q_i,\ \Delta(Q_i) = 0, \tag{5.19b}$$

$$P_i \geq 0, Q_i \text{ Hermitian}. \tag{5.19c}$$

Suppose the optimal solution of SDP (5.19) is given by $\{P_{i_0}, Q_{i_0}\}$. Let $X = |i_0\rangle\langle i_0| \otimes P_{i_0}$, $Y = |i_0\rangle\langle i_0|$ and $Z_i = \delta_{i,i_0} \cdot Q_{i_0}$. We can verify that $\{X, Y, Z_i\}$ is a feasible solution of SDP (5.18) which implies

$$S_{c,\mathrm{MIO}}^{(1),0}(\mathcal{N}) - \delta \geq \operatorname{Tr} J_{\mathcal{N}} |i_0\rangle\langle i_0| \otimes P_{i_0} = r. \tag{5.20}$$

On the other hand, suppose the optimal solution of SDP (5.18) is given by $\{X, Y, Z_i\}$. If $y_i = \operatorname{Tr} Y|i\rangle\langle i| = 0$, we have $Z_i \geq 0$. This implies $Z_i = 0$ due to $\Delta(Z_i) = 0$. Without loss of generality, we suppose $y_i > 0$ for any i. Then we have the optimal value

$$\operatorname{Tr} J_{\mathcal{N}} X = \sum_{i=0}^{d_A-1} \operatorname{Tr} J_{\mathcal{N}} |i\rangle\langle i| \otimes (y_i \mathbb{1}_B + Z_i) \tag{5.21}$$

$$= \sum_{i=0}^{d_A-1} y_i \operatorname{Tr} J_{\mathcal{N}} |i\rangle\langle i| \otimes \left(\mathbb{1}_B + \frac{Z_i}{y_i}\right) \tag{5.22}$$

$$\leq \max_i \operatorname{Tr} J_{\mathcal{N}} |i\rangle\langle i| \otimes \left(\mathbb{1}_B + \frac{Z_i}{y_i}\right). \tag{5.23}$$

Suppose the maximum is taken at $i = i_0$. Let $P_{i_0} = \mathbb{1}_B + \frac{Z_{i_0}}{y_{i_0}}$ and $Q_{i_0} = \frac{Z_{i_0}}{y_{i_0}}$. We know that $\{P_{i_0}, Q_{i_0}\}$ is a feasible solution to SDP (5.19). Thus

$$r \geq \operatorname{Tr} J_{\mathcal{N}} |i_0\rangle\langle i_0| \otimes \left(\mathbb{1}_B + \frac{Z_{i_0}}{y_{i_0}} \right) \geq S^{(1),0}_{c,\mathrm{MIO}} (\mathcal{N}) - \delta. \tag{5.24}$$

■

Due to the additivity (Lemma 5.8) and the quasi-convexity of max-relative entropy of coherence [BSF+17] , we have the following asymptotic result.

Theorem 5.7 *The asymptotic zero-error coherence simulation cost under* MIO *is given by*

$$S^0_{c,\mathrm{MIO}} (\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} S^{(1),0}_{c,\mathrm{MIO}} \left(\mathcal{N}^{\otimes n}\right) \tag{5.25}$$

$$= \max_i C_{\max} \left(\mathcal{N} \left(|i\rangle\langle i|\right)\right) \tag{5.26}$$

$$= \max_{\rho\in\mathcal{I}} C_{\max} \left(\mathcal{N} (\rho)\right). \tag{5.27}$$

Proposition 5.6 and Theorem 5.7 are interesting since they present that the coherence we need to simulate a channel is given by the maximum coherence that can be generated from this channel—the *cohering power of the channel* [BCP14, MK15]. These results are similar to the quantum reverse Shannon theory as we will discuss in the next section. It is of great interest to make a conjecture for the general case that the asymptotic coherence simulation cost of a channel is equal to its coherence generating capacity introduce in [DDMW17].

Lemma 5.8 *The max-relative entropy of coherence is additive, i.e.,*

$$C_{\max} (\rho_1 \otimes \rho_2) = C_{\max} (\rho_1) + C_{\max} (\rho_2). \tag{5.28}$$

Proof Due to the additivity of max-relative entropy, it is easy to verify that

$$C_{\max} (\rho_1 \otimes \rho_2) \leq C_{\max} (\rho_1) + C_{\max} (\rho_2). \tag{5.29}$$

We prove the other direction via SDP. The max-relative entropy of coherence can be given by the following SDP[1],

$$\textbf{Primal:} \quad C_{\max} (\rho) = \log \min \left\{ \operatorname{Tr} V \,\middle|\, \rho \leq V,\ V = \Delta (V),\ V \geq 0 \right\}, \tag{5.30}$$

$$\textbf{Dual:} \quad C_{\max} (\rho) = \log \max \left\{ \operatorname{Tr} \rho X \,\middle|\, 0 \leq X \leq \mathbb{1} + Y,\ \Delta (Y) = 0 \right\}. \tag{5.31}$$

Suppose the optimal solutions for ρ_1 and ρ_2 in the dual SDP is given by $\{X_1, Y_1\}$ and $\{X_2, Y_2\}$, respectively. Then we can verify that $\{X_1 \otimes X_2, \mathbb{1}_1 \otimes Y_2 + Y_1 \otimes \mathbb{1}_2 + Y_1 \otimes Y_2\}$ is a feasible solution for $\rho_1 \otimes \rho_2$. Thus

$$C_{\max} (\rho_1 \otimes \rho_2) \geq \log \operatorname{Tr} (\rho_1 \otimes \rho_2) (X_1 \otimes X_2) \geq C_{\max} (\rho_1) + C_{\max} (\rho_2). \tag{5.32}$$

This concludes the proof. ■

[1] It is clear that the condition $X \leq \mathbb{1} + Y$ in the dual SDP could be replaced with equality.

5.2.3 The channel's max-relative entropy

The max-relative entropy between two quantum states has been extensively used in the quantum information theory. Here we introduce a channel's version of this entropy and use it to characterize the one-shot coherence channel simulation cost.

Definition 5.9 *For any two quantum channels $\mathcal{N}_{A\to B}$ and $\mathcal{M}_{A\to B}$ with corresponding Choi-Jamiołkowski matrix $J_{\mathcal{N}}$ and $J_{\mathcal{M}}$, we define the channel's max-relative entropy as*

$$D_{\max}\left(\mathcal{N}\|\mathcal{M}\right) := D_{\max}\left(J_{\mathcal{N}}\|J_{\mathcal{M}}\right). \tag{5.33}$$

The smoothed channel's max-relative is defined as

$$D^{\varepsilon}_{\max}\left(\mathcal{N}\|\mathcal{M}\right) := \inf_{\substack{\frac{1}{2}\|\widetilde{\mathcal{N}}-\mathcal{N}\|_{\diamond}\leq\varepsilon \\ \widetilde{\mathcal{N}}\in\mathrm{CPTP(A:B)}}} D_{\max}\big(\widetilde{\mathcal{N}}\big\|\mathcal{M}\big). \tag{5.34}$$

From SDP (5.5), we know that $J_{\widetilde{\mathcal{N}}}$ corresponds to a CPTP map $\widetilde{\mathcal{N}}$ according to the Choi-Jamiołkowski isomorphism. Then we have [2]

$$\begin{aligned}\omega_{\mathrm{MIO}}\left(\Psi_k,\mathcal{N}\right) = \frac{1}{2}\inf\ &\|\widetilde{\mathcal{N}}-\mathcal{N}\|_{\diamond} \\ &\text{s.t. } J_{\widetilde{\mathcal{N}}}\leq k\cdot J_{\mathcal{M}},\ \widetilde{\mathcal{N}}\in\mathrm{CPTP},\mathcal{M}\in\mathrm{MIO}.\end{aligned} \tag{5.35}$$

Together with the definition of the coherence simulation cost, we can characterize this cost via the smoothed channel's max-relative entropy.

Proposition 5.10 *For a given quantum channel $\mathcal{N}_{A\to B}$, the one-shot ε-error coherence simulation cost under* MIO *can be given by*

$$S^{(1),\varepsilon}_{c,\mathrm{MIO}}\left(\mathcal{N}\right) = \min_{\mathcal{M}\in\mathrm{MIO}} D^{\varepsilon}_{\max}\left(\mathcal{N}||\mathcal{M}\right)+\delta, \tag{5.36}$$

where $\delta\in[0,1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

This result also gives the distance characterization of the one-shot coherence simulation cost as shown in Figure 5.2.

[2] From the optimization (5.35), we can see that a quantum channel $\mathcal{N}_{A\to B}$ can be written as $\Pi_{AR\to B}\left((\cdot)\otimes\Psi_k\right)$ with Π in MIO if and only if there exists a channel $\mathcal{M}$ in MIO such that $J_{\mathcal{N}}\leq kJ_{\mathcal{M}}$.

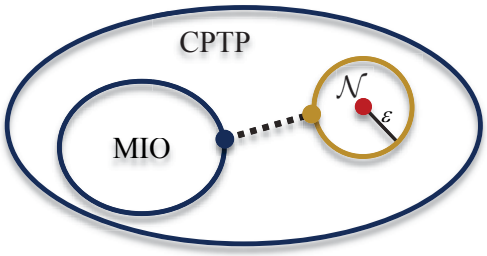


Figure 5.2 : Distance characterization of one-shot coherence simulation cost under MIO.

The Proposition 5.10 gives the characterization in the one-shot scenario. From the perspective of information theory, we may wonder what it converges to asymptotically. Inspired by the main result in Ref. [BP10], we provide a conjecture for this as follows.

Define the channel distance with respect to the quantum relative entropy as

$$D\left(\mathcal{N}\|\mathcal{M}\right) := \max_{\rho_A} D\left(\mathcal{N}_{A\to B}\left(\phi_{AA'}\right)\|\mathcal{M}_{A\to B}\left(\phi_{AA'}\right)\right) \tag{5.37}$$

where $\phi_{AA'}$ is a purification of ρ_A. Then for any quantum channel, we denote its relative entropy distance to the set of MIO as

$$K_{\mathrm{MIO}}\left(\mathcal{N}\right) := \min_{\mathcal{M}\in\mathrm{MIO}} D\left(\mathcal{N}\|\mathcal{M}\right), \tag{5.38}$$

and its regularization form

$$K^{\infty}_{\mathrm{MIO}}\left(\mathcal{N}\right) := \lim_{n\to\infty}\frac{1}{n}K_{\mathrm{MIO}}\left(\mathcal{N}^{\otimes n}\right). \tag{5.39}$$

Conjecture 5.11 *For any quantum channel $\mathcal{N}$, its asymptotic coherence simulation cost under* MIO *is given by*

$$S_{c,\mathrm{MIO}}\left(\mathcal{N}\right) = K^{\infty}_{\mathrm{MIO}}\left(\mathcal{N}\right). \tag{5.40}$$

5.3 Channel simulation via quantum channels

In this section, we focus on quantum channel simulation via other quantum channels. The central quantity we introduce is the channel's smooth max-information.

5.3.1 General framework and codes

Let us first formally introduce the task of channel simulation and some notations. A quantum channel (quantum operation) $\mathcal{N}_{A_o\to B_i}$ is a completely positive (CP) and trace-preserving (TP)

linear map from operators on a finite-dimensional Hilbert space A_o to operators on a finite-dimensional Hilbert space B_i. As shown in Figure 5.3, Alice and Bob share a quantum channel $\mathcal{N}_{A_o \to B_i}$. By adding encoding and decoding scheme, they can use the channel $\mathcal{N}$ to simulate another channel $\mathcal{M}$. Composing with the encoder and decoder, their effective channel is given by $\widetilde{\mathcal{N}}_{A_i \to B_o} = \Pi_{A_i B_i \to A_o B_o} \mathcal{N}_{A_o \to B_i}$, where Π is a bipartite quantum operation that generalizes the usual encoding scheme $\mathcal{E}$ and decoding scheme $\mathcal{D}$. Note that the bipartite quantum operation Π here is required to be B to A non-signalling, which makes the composition of Π and $\mathcal{N}$ feasible [CDP08, DW16]. We say such Π is an Ω-assisted code if it can be implemented by local operations with Ω-assistance. In the following, we eliminate Ω for the case of unassisted codes. We write $\Omega = \mathrm{NS}$ and $\Omega = \mathrm{PPT}$ for NS-assisted and PPT-assisted codes, respectively. These codes have also been applied to other basic tasks of quantum information processing (see e.g., [LM15a, APE03, WFD17, Rai01, FWTD17]). In particular:

- an unassisted code reduces to the product of encoder and decoder, $\Pi = \mathcal{D}_{B_i \to B_o} \mathcal{E}_{A_i \to A_o}$;
- a NS-assisted code corresponds to a bipartite quantum operation which is non-signalling from Alice to Bob and vice-versa;
- a PPT-assisted code corresponds to a bipartite operation whose Choi-Jamiołkowski operator is positive under partial transpose over systems $B_i B_o$.

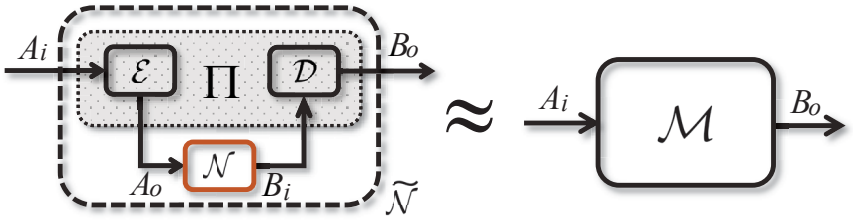


Figure 5.3 : The task of quantum channel simulation via another quantum channel. The subscript i stands for "input" while o stands for "output".

Definition 5.12 *For any two quantum channels $\mathcal{N}$ and $\mathcal{M}$, the minimum error of simulation from $\mathcal{N}$ to $\mathcal{M}$ under Ω-assisted codes is defined as*

$$\omega_\Omega(\mathcal{N}, \mathcal{M}) := \frac{1}{2} \inf_{\Pi \in \Omega} \|\Pi \circ \mathcal{N} - \mathcal{M}\|_\diamond, \tag{5.41}$$

where $\|\cdot\|_\diamond$ is the diamond norm.

Definition 5.13 *The channel simulation rate from $\mathcal{N}$ to $\mathcal{M}$ under Ω-assisted codes is defined as*

$$S_{\Omega}\left(\mathcal{N},\mathcal{M}\right):=\lim_{\varepsilon\to 0}\inf\left\{\frac{n}{m}\;\middle|\;\omega_{\Omega}\left(\mathcal{N}^{\otimes n},\mathcal{M}^{\otimes m}\right)\leq\varepsilon\right\}, \tag{5.42}$$

where the infimum is taken over ratios $\frac{n}{m}$ with $n,m\in\mathbb{N}$.

In this framework of channel simulation, the classical capacity $C\left(\mathcal{N}\right)$ and the quantum capacity $Q\left(\mathcal{N}\right)$ of the channel $\mathcal{N}$ are given by

$$C\left(\mathcal{N}\right)=S\big(\mathcal{N},\widehat{\mathrm{id}}_2\big)^{-1}\text{ and }\;Q\left(\mathcal{N}\right)=S\left(\mathcal{N},\mathrm{id}_2\right)^{-1}, \tag{5.43}$$

where $\widehat{\mathrm{id}}_2$ is the one-bit noiseless channel and id_2 is the one-qubit noiseless channel.

If we consider simulating the given channel $\mathcal{N}$ via a m-dimensional noiseless quantum channel id_m, then the one-shot ε-error quantum simulation cost under Ω-assisted codes is defined as

$$S_{\Omega,\varepsilon}^{(1)}\left(\mathcal{N}\right):=\log\min\left\{m\in\mathbb{N}\mid\omega_{\Omega}\left(\mathrm{id}_m,\mathcal{N}\right)\leq\varepsilon\right\}. \tag{5.44}$$

The asymptotic quantum simulation cost is given by

$$S_{\Omega}\left(\mathcal{N}\right)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}S_{\Omega,\varepsilon}^{(1)}\left(\mathcal{N}^{\otimes n}\right). \tag{5.45}$$

5.3.2 Channel simulation via noisy quantum channels

We show that the minimum error of simulation under NS-assisted (and PPT-assisted) codes can be given by SDPs. The one-shot ε-error quantum simulation cost under NS-assisted codes can also be given by an SDP.

Lemma 5.14 ([LM15a, DW16]) *Let J_{Π} and $J_{\mathcal{N}}$ denote the Choi-Jamiołkowski operators of the quantum channel $\mathcal{N}$ and code Π, respectively. Then the Choi-Jamiołkowski operator of the effective channel $\widetilde{\mathcal{N}}_{A_i\to B_o}=\Pi_{A_iB_i\to A_oB_o}\circ\mathcal{N}_{A_o\to B_i}$ is given by*

$$J_{\widetilde{\mathcal{N}}}=\mathrm{Tr}_{A_oB_i}\left(J_{\mathcal{N}}^{T}\otimes\mathbb{1}_{A_iB_o}\right)J_{\Pi}. \tag{5.46}$$

Combining with the SDP (2.82) for diamond norm and the constraints of the code Π, we have the following result.

Proposition 5.15 *For any two quantum channels $\mathcal{N}$ and $\mathcal{M}$ with the corresponding Choi-Jamiołkowski matrices $J_{\mathcal{N}}$ and $J_{\mathcal{M}}$, the minimum error of simulation from $\mathcal{N}$ to $\mathcal{M}$ under*

NS-*assisted codes* $\omega_{\mathrm{NS}}(\mathcal{N},\mathcal{M})$ *is given by the following SDP,*

$$\begin{aligned}
&\min \gamma && \text{(5.47a)}\\
&\text{s.t. } \mathrm{Tr}_{B_o} Y_{A_iB_o} \le \gamma \mathbb{1}_{A_i}, && \text{(5.47b)}\\
&\quad Y_{A_iB_o} \ge J_{\widetilde{\mathcal{N}}} - J_{\mathcal{M}},\ Y_{A_iB_o} \ge 0, && \text{(5.47c)}\\
&\quad J_{\widetilde{\mathcal{N}}} = \mathrm{Tr}_{A_oB_i}\left(J_{\mathcal{N}}^T \otimes \mathbb{1}_{A_iB_o}\right) J_\Pi, && \text{(5.47d)}\\
&\quad J_\Pi \ge 0,\ \mathrm{Tr}_{A_oB_o} J_\Pi = \mathbb{1}_{A_iB_i}, \qquad \text{(CPTP)} && \text{(5.47e)}\\
&\quad \mathrm{Tr}_{A_o} J_\Pi = \frac{\mathbb{1}_{A_i}}{|A_i|} \otimes \mathrm{Tr}_{A_oA_i} J_\Pi, \qquad (\mathrm{A} \not\to \mathrm{B}) && \text{(5.47f)}\\
&\quad \mathrm{Tr}_{B_o} J_\Pi = \frac{\mathbb{1}_{B_i}}{|B_i|} \otimes \mathrm{Tr}_{B_iB_o} J_\Pi. \qquad (\mathrm{B} \not\to \mathrm{A})\,. && \text{(5.47g)}
\end{aligned}$$

To obtain $\omega_{\mathrm{NS}\cap\mathrm{PPT}}(\mathcal{N},\mathcal{M})$*, we only need to add the PPT constraint* $J_\Pi^{T_{B_iB_o}} \ge 0$*, where* $T_{B_iB_o}$ *denotes the partial transpose over systems* B_iB_o*.*

The constraints in Eq. (5.47e) represent the CP and TP conditions of the bipartite operation Π. The constraints in Eqs. (5.47f) and (5.47g) represent the non-signalling conditions that A cannot signal to B and B cannot signal to A, respectively.

Corollary 5.16 *The minimum error to simulate a quantum channel* $\mathcal{N}$ *from a noiseless quantum channel* id_m *under* NS*-assisted codes* $\omega_{\mathrm{NS}}(\mathrm{id}_m,\mathcal{N})$ *is given by the following SDP,*

$$\begin{aligned}
&\min \gamma && \text{(5.48a)}\\
&\text{s.t. } \mathrm{Tr}_{B_o} Y_{A_iB_o} \le \gamma \mathbb{1}_{A_i}, && \text{(5.48b)}\\
&\quad Y_{A_iB_o} \ge J_{\widetilde{\mathcal{N}}} - J_{\mathcal{N}},\ Y_{A_iB_o} \ge 0, && \text{(5.48c)}\\
&\quad J_{\widetilde{\mathcal{N}}} \ge 0,\ \mathrm{Tr}_{B_o} J_{\widetilde{\mathcal{N}}} = \mathbb{1}_{A_i}, && \text{(5.48d)}\\
&\quad J_{\widetilde{\mathcal{N}}} \le \mathbb{1}_{A_i} \otimes V_{B_o},\ \mathrm{Tr}\, V_{B_o} = m^2. && \text{(5.48e)}
\end{aligned}$$

To obtain $\omega_{\mathrm{NS}\cap\mathrm{PPT}}(\mathrm{id}_m,\mathcal{N})$*, we only need to add the* PPT *constraint*

$$-\mathbb{1}_{A_i} \otimes V_{B_o}^T \le m J_{\widetilde{\mathcal{N}}}^{T_B} \le \mathbb{1}_{A_i} \otimes V_{B_o}^T. \tag{5.49}$$

Proof The Choi-Jamiołkowski operator of the operation id_m is given by

$$J_m = \sum_{i,j=0}^{m-1} |ii\rangle\langle jj|_{A_oB_i}, \tag{5.50}$$

which is the unnormalized maximally entangled state. The main idea is to exploit the symmetry of J_m and simplify the SDP (5.47).

Note that J_m is invariant under any local unitary $U_{A_o} \otimes \overline{U}_{B_i}$. If $\widetilde{J}_\Pi$ is optimal for SDP (5.47), we can verify that $\left(U_{A_o} \otimes \overline{U}_{B_i}\right) \widetilde{J}_\Pi \left(U_{A_o} \otimes \overline{U}_{B_i}\right)^\dagger$ is also optimal. Any convex combination of optimal solutions remains optimal. Thus, without loss of generality we can take

$$J_\Pi = \int dU \left(U_{A_o} \otimes \overline{U}_{B_i}\right) \widetilde{J}_\Pi \left(U_{A_o} \otimes \overline{U}_{B_i}\right)^\dagger \tag{5.51}$$

$$= \frac{J_m}{m} \otimes C_{A_i B_o} + \left(\mathbb{1} - \frac{J_m}{m}\right) \otimes D_{A_i B_o}, \tag{5.52}$$

where the integral is taken over the Haar measure and C, D are operators on system $A_i B_o$. Combining Eq. (5.51) and SDP (5.47), we know that Eq. (5.47d) is equivalent to $J_{\widetilde{\mathcal{N}}} = mC$. Eq. (5.47e) is equivalent to $C \geq 0$, $D \geq 0$ and $\mathrm{Tr}_{B_o}\left(C + \left(m^2 - 1\right) D\right) = m\mathbb{1}_{A_i}$. Since $J_{\widetilde{\mathcal{N}}}$ is the Choi-Jamiołkowski matrix of the effective channel, we have $\mathrm{Tr}_{B_o} J_{\widetilde{\mathcal{N}}} = \mathrm{Tr}_{B_o} mC = \mathbb{1}_{A_i}$ and $\mathrm{Tr}_{B_o} mD = \mathbb{1}_{A_i}$. This implies that Eq. (5.47g) is trivial and Eq. (5.47f) is equivalent to $C + \left(m^2 - 1\right) D = \frac{\mathbb{1}_{A_i}}{|A_i|} \otimes \mathrm{Tr}_{A_i}\left(C + \left(m^2 - 1\right) D\right)$. Denote $V_{B_o} = \frac{m}{|A_i|} \mathrm{Tr}_{A_i}\left(C + \left(m^2 - 1\right) D\right)$. We have $J_{\widetilde{\mathcal{N}}} + \left(m^2 - 1\right) mD = \mathbb{1}_{A_i} \otimes V_{B_o}$. Eliminating variable D, we have the desired SDP. ■

From the definition of $S^{(1)}_{\mathrm{NS},\varepsilon}(\mathcal{N})$ and SDP (5.48), we have the following SDP for the one-shot quantum simulation cost.

Proposition 5.17 *For any quantum channel $\mathcal{N}_{A_i \to B_o}$ and error tolerance $\varepsilon \geq 0$, the one-shot ε-error quantum simulation cost under* NS*-assisted codes is given by the following SDP,*

$$S^{(1)}_{\mathrm{NS},\varepsilon}(\mathcal{N}) = \log \min \left\lceil \sqrt{\mathrm{Tr}\, V_{\mathrm{B_o}}} \right\rceil \tag{5.53a}$$

$$\text{s.t. } \mathrm{Tr}_{B_o} Y_{A_i B_o} \leq \varepsilon \mathbb{1}_{A_i}, \tag{5.53b}$$

$$Y_{A_i B_o} \geq J_{\widetilde{\mathcal{N}}} - J_{\mathcal{N}},\ Y_{A_i B_o} \geq 0, \tag{5.53c}$$

$$J_{\widetilde{\mathcal{N}}} \geq 0,\ \mathrm{Tr}_{B_o} J_{\widetilde{\mathcal{N}}} = \mathbb{1}_{A_i}, \tag{5.53d}$$

$$J_{\widetilde{\mathcal{N}}} \leq \mathbb{1}_{A_i} \otimes V_{B_o}. \tag{5.53e}$$

Remark 5.18 With the same reasoning as Remark 5.5, we can also write Eq. (5.53) as

$$S^{(1)}_{\mathrm{NS},\varepsilon}(\mathcal{N}) = \delta + \frac{1}{2} \log \min \mathrm{Tr}\, V_{\mathrm{B_o}} \tag{5.54a}$$

$$\text{s.t. } \mathrm{Tr}_{B_o} Y_{A_i B_o} \leq \varepsilon \mathbb{1}_{A_i}, \tag{5.54b}$$

$$Y_{A_i B_o} \geq J_{\widetilde{\mathcal{N}}} - J_{\mathcal{N}},\ Y_{A_i B_o} \geq 0, \tag{5.54c}$$

$$J_{\widetilde{\mathcal{N}}} \geq 0,\ \mathrm{Tr}_{B_o} J_{\widetilde{\mathcal{N}}} = \mathbb{1}_{A_i}, \tag{5.54d}$$

$$J_{\widetilde{\mathcal{N}}} \leq \mathbb{1}_{A_i} \otimes V_{B_o}, \tag{5.54e}$$

where $\delta \in [0, 1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

Note that the one-shot quantum simulation cost under NS∩PPT-assisted codes is not an SDP, since the objective function m appears in the conditions $\operatorname{Tr} V_{B_o} = m^2$ and $-\mathbb{1}_{A_i} \otimes V_{B_o}^T \leq m J_{\widetilde{\mathcal{N}}}^{T_B} \leq \mathbb{1}_{A_i} \otimes V_{B_o}^T$ with different powers. We do not see a way to obtain a linear objective function.

It is also worth mentioning that the zero-error quantum simulation cost was studied by Duan and Winter in [DW16]. The authors show that the zero-error NS-assisted simulation cost is given by the conditional min-entropy of the channel's Choi-Jamiołkowski matrix [DW16, Theorem 2]. We can recover their result by setting $\varepsilon = 0$ in SDP (5.53). This will lead to $Y_{A_i B_o} = 0$ and thus $J_{\widetilde{\mathcal{N}}} = J_{\mathcal{N}}$. Then we have

$$S_{\mathrm{NS},0}^{(1)}(\mathcal{N}) = \frac{1}{2} \log \min \left\{ \operatorname{Tr} V_{B_o} \mid J_{\mathcal{N}} \leq \mathbb{1}_{A_i} \otimes V_{B_o} \right\} + \delta. \tag{5.55}$$

The right hand side is the conditional min-entropy of $J_{\mathcal{N}}$ which is additive (cf. [Tom12]). Then we can obtain that

$$S_{\mathrm{NS},0}(\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} S_{\mathrm{NS},0}^{(1)}\left(\mathcal{N}^{\otimes n}\right) = \frac{1}{2} \log \min \left\{ \operatorname{Tr} V_{B_o} \mid J_{\mathcal{N}} \leq \mathbb{1}_{A_i} \otimes V_{B_o} \right\}. \tag{5.56}$$

5.3.3 The channel's max-information and the channel's robustness

In this section, we introduce a novel entropy called the channel's smooth max-information and show that it has an operational interpretation regarding the quantum simulation cost of a channel. We further introduce the channel's robustness and illustrate its connection with the channel's max-information.

The channel's max-information

Recall that the max-relative entropy of $\rho \in \mathcal{S}_{\leq}(A)$ with respect to $\sigma \geq 0$ is defined as [Dat09b, Ren05]

$$D_{\max}(\rho\|\sigma) = \log \inf\{t > 0 \mid \rho \leq t \cdot \sigma\}. \tag{5.57}$$

The max-information that we will use in the following discussion (same as ${}^2I_{\max}$ introduced in Eq. (2.53)) is defined as

$$I_{\max}(A:B)_\rho := \inf_{\sigma_B \in \mathcal{S}_=(B)} D_{\max}(\rho_{AB}\|\rho_A \otimes \sigma_B). \tag{5.58}$$

Definition 5.19 *For any quantum channel $\mathcal{N}_{A'\to B}$, the channel's max-information of $\mathcal{N}$ is defined as*

$$I_{\max}(A:B)_{\mathcal{N}} := I_{\max}(A:B)_{\mathcal{N}_{A'\to B}(\Phi_{AA'})}, \tag{5.59}$$

where $\Phi_{AA'}$ is the maximally entangled state on AA'.

The following argument shows that for any full rank input state $\rho_{A'}$ with a purification $\phi_{AA'}$, it holds that

$$I_{\max}(A:B)_{\mathcal{N}} = I_{\max}(A:B)_{\mathcal{N}_{A'\to B}(\phi_{AA'})}. \tag{5.60}$$

From the definitions (5.57), (5.58) and (5.59), we have

$$I_{\max}(A:B)_{\mathcal{N}} = \log \inf\left\{ t>0 \;\middle|\; \mathcal{N}_{A'\to B}(\Phi_{AA'}) \le t\cdot \Phi_A \otimes \sigma_B,\ \sigma_B \in \mathcal{S}_=(B)\right\}. \tag{5.61}$$

Since we have

$$\begin{aligned}\mathcal{N}_{A'\to B}(\phi_{AA'}) &= |A|\cdot \mathcal{N}_{A'\to B}\left(\sqrt{\phi_A}\Phi_{AA'}\sqrt{\phi_A}\right) &(5.62)\\ &= |A|\cdot \sqrt{\phi_A}\mathcal{N}_{A'\to B}(\Phi_{AA'})\sqrt{\phi_A}, &(5.63)\end{aligned}$$

then the first condition in (5.61) is equivalent to $\mathcal{N}_{A'\to B}(\phi_{AA'}) \le t\cdot\phi_A\otimes\sigma_B$ and it implies Eq. (5.60).

From Eqs. (5.55) and (5.61), we can write the one-shot zero-error quantum simulation cost as the channel's max-information,

$$S^{(1)}_{\mathrm{NS},0}(\mathcal{N}) = \frac{1}{2} I_{\max}(A:B)_{\mathcal{N}} + \delta. \tag{5.64}$$

In the following, we show this relation beyond the zero-error case.

Definition 5.20 *For any quantum channel $\mathcal{N}_{A'\to B}$, the channel's smooth max-information is defined as*

$$I^{\varepsilon}_{\max}(A:B)_{\mathcal{N}} := \inf_{\substack{\frac{1}{2}\|\widetilde{\mathcal{N}}-\mathcal{N}\|_\diamond\le\varepsilon \\ \widetilde{\mathcal{N}}\in \mathrm{CPTP}(A':B)}} I_{\max}(A:B)_{\widetilde{\mathcal{N}}}, \tag{5.65}$$

where $\mathrm{CPTP}(A':B)$ *denotes the set of all the* CPTP *maps from A' to B.*

We show that the one-shot ε-error quantum simulation cost is exactly given by the channel's smooth max-information. This provides the operational meaning of this new measure.

Theorem 5.21 *For any quantum channel $\mathcal{N}_{A'\to B}$ and error tolerance $\varepsilon \ge 0$, it holds that*

$$S^{(1)}_{\mathrm{NS},\varepsilon}(\mathcal{N}) = \frac{1}{2} I^{\varepsilon}_{\max}(A:B)_{\mathcal{N}} + \delta, \tag{5.66}$$

where $\delta\in[0,1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

Proof Note that the constraints $J_{\widetilde{\mathcal{N}}} \ge 0$, $\mathrm{Tr}_B J_{\widetilde{\mathcal{N}}} = \mathbb{1}_A$ in Eq. (5.54d) uniquely define a CPTP map $\widetilde{\mathcal{N}}$ according to the Choi-Jamiołkowski isomorphism. Applying the SDP (2.82) of

the diamond norm, we find

$$S^{(1)}_{\mathrm{NS},\varepsilon}(\mathcal{N}) = \delta + \frac{1}{2}\log\min \operatorname{Tr} V_B \tag{5.67a}$$

$$\text{s.t. } \frac{1}{2}\|\widetilde{\mathcal{N}} - \mathcal{N}\|_\diamond \leq \varepsilon, \tag{5.67b}$$

$$\widetilde{\mathcal{N}} \in \mathrm{CPTP}\left(A' : B\right), \tag{5.67c}$$

$$J_{\widetilde{\mathcal{N}}} \leq \mathbb{1}_A \otimes V_B. \tag{5.67d}$$

From Eqs. (5.55) and (5.64), we know that

$$I_{\max}(A:B)_{\mathcal{N}} = \log\min\left\{\operatorname{Tr} V_B \mid J_{\mathcal{N}} \leq \mathbb{1}_A \otimes V_B\right\}. \tag{5.68}$$

Combining SDPs (5.67) and (5.68), we obtain the desired result. ■

In the setting of the entanglement-assisted one-shot capacity of quantum channels, Matthews and Wehner gave a converse bound in terms of the channel's hypothesis testing relative entropy [MW14]. Moreover, it was recently shown that the activated NS-assisted one-shot capacity is exactly given by the channel's hypothesis testing relative entropy [WFT17] – generalizing the corresponding classical results [PPV10, Mat12]. This suggests that the operational min- and max-type one-shot analogs of the channel's mutual information are the channel's hypothesis testing relative entropy and the channel's smooth max-information, respectively.

Remark 5.22 From this result, we can operationally verify that the data-processing inequality holds for the channel's smooth max-information, i.e., $I^{\varepsilon}_{\max}(A_0:B_0)_{\mathcal{T}\circ\mathcal{N}\circ\mathcal{F}} \leq I^{\varepsilon}_{\max}(A_1:B_1)_{\mathcal{N}}$ holds for any CPTP maps $\mathcal{N}_{A_1'\to B_1}$, $\mathcal{F}_{A_0'\to A_1'}$ and $\mathcal{T}_{B_1\to B_0}$. This follows from the fact that we need less resources to simulate a quantum channel with higher noise. Or we can directly prove it in the following proposition.

The channel's robustness and log-robustness

We introduce the log-robustness of a quantum channel and explain its relation with channel's max-information. As the name suggests, the robustness quantifies the minimal mixing required to make the given resource useless. In terms of quantum channels, a natural class of useless channels is the *constant channel*, i.e., it sends any input state to a fixed state. Denote the class of constant channels as

$$\mathcal{G} := \left\{\mathcal{N} \in \mathrm{CPTP}(A:B) \;\middle|\; \exists\, \sigma \text{ s.t. } \mathcal{N}(\rho) = \sigma, \forall\, \rho\right\}. \tag{5.69}$$

Definition 5.23 *The robustness of a quantum channel $\mathcal{N} : \mathcal{L}(A) \to \mathcal{L}(B)$ is defined as*

$$\mathcal{R}_g(\mathcal{N}) := \inf\left\{t \geq 0 \;\middle|\; \exists\, \mathcal{M} \in \mathrm{CPTP}(\mathrm{A}:\mathrm{B}) \text{ s.t. } \frac{\mathcal{N} + \mathrm{t}\mathcal{M}}{1+\mathrm{t}} \in \mathcal{G}\right\}. \tag{5.70}$$

Definition 5.24 *The log-robustness of a quantum channel $\mathcal{N}:\mathcal{L}(A)\to\mathcal{L}(B)$ is defined as*

$$\mathcal{LR}_g(\mathcal{N}):=\log\left(1+\mathcal{R}_g(\mathcal{N})\right). \tag{5.71}$$

Then its corresponding smoothed version is defined as

$$\mathcal{LR}_g^{\varepsilon}(\mathcal{N}):=\inf_{\substack{\frac{1}{2}\|\widetilde{\mathcal{N}}-\mathcal{N}\|_{\diamond}\leq\varepsilon\\ \widetilde{\mathcal{N}}\in\mathrm{CPTP}(A:B)}}\mathcal{LR}_g\big(\widetilde{\mathcal{N}}\big). \tag{5.72}$$

Robustness has a very nice geometric interpretation as shown in Figure 5.4. For any given quantum channel $\mathcal{N}$, we need to find another channel $\mathcal{M}$ such that a convex combination of them $\mathcal{F}=(\mathcal{N}+t\mathcal{M})/(1+t)$ is a constant channel. We hope to minimize the weight of $\mathcal{M}$ in this combination, i.e., to make $\mathcal{F}$ as close as possible to the channel $\mathcal{N}$.

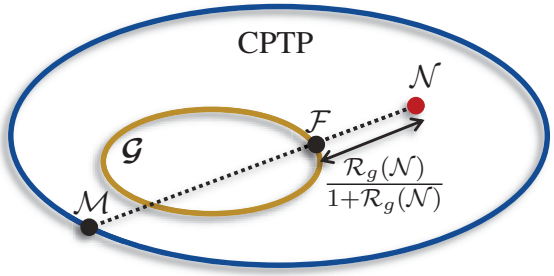


Figure 5.4 : Geometric interpretation of the channel's robustness defined in Eq. (5.70).

The following result shows the relations between the max-information, log-robustness and max-relative entropy of a quantum channel.

Proposition 5.25 *For any quantum channel $\mathcal{N}$, it holds*

$$I_{\max}^{\varepsilon}(A:B)_{\mathcal{N}}=\min_{\mathcal{F}\in\boldsymbol{\mathcal{G}}}D_{\max}^{\varepsilon}(\mathcal{N}\|\mathcal{F})=\mathcal{LR}_g^{\varepsilon}(\mathcal{N}). \tag{5.73}$$

Proof We only need to show the case when $\varepsilon=0$ and the general result follows by our unifying way of channel smoothing. The first equality follows by the noticing that the Choi-Jamiołkowski operator of every constant channel $\mathcal{F}(\rho_A)=\sigma_B$ is given by $\mathbb{1}_A\otimes\sigma_B$. As for the second equality, note that $\frac{\mathcal{N}+t\mathcal{M}}{1+t}\in\boldsymbol{\mathcal{G}}$ if and only if there exists $\mathcal{F}\in\boldsymbol{\mathcal{G}}$ such that $\mathcal{N}+t\mathcal{M}=(1+t)\mathcal{F}$. Suppose $\mathcal{F}(\rho_A)=\sigma_B$ for any $\rho_A\in\mathcal{S}_{=}(A)$. Then $\mathcal{N}+t\mathcal{M}=(1+t)\mathcal{F}$ holds for certain $\mathcal{M}\in\mathrm{CPTP(A:B)}$ if and only if $J_{\mathcal{N}}\leq(1+t)\mathbb{1}_A\otimes\sigma_B$. Then the second equality holds by definitions. ■

Together with the previous result $S_{\mathrm{NS},\varepsilon}^{(1)}(\mathcal{N})=\frac{1}{2}I_{\max}^{\varepsilon}(A:B)_{\mathcal{N}}+\delta$, we know that the quantum simulation cost can be characterized by the minimum "distance" to the class of constant channels, and it also provides an operational interpretation for the channel's log-robustness.

5.3.4 Asymptotic equipartition property and quantum reverse Shannon theorem

In the following, we prove the asymptotic equipartition property (AEP) of the channel's smooth max-information and explore its close relation to the quantum reverse Shannon theorem. In the framework of quantum channel simulation, the quantum capacity is given by the optimal rate of using $\mathcal{N}$ to simulate the qubit noiseless channel id_2, while the channel simulation cost is given by the optimal rate of using id_2 to simulate the channel $\mathcal{N}$. Thus, it operationally holds that

$$Q_E\left(\mathcal{N}\right) \leq Q_{\mathrm{NS}}\left(\mathcal{N}\right) \leq S_{\mathrm{NS}}\left(\mathcal{N}\right) \leq S_E\left(\mathcal{N}\right), \tag{5.74}$$

where the above four notations represent entanglement-assisted quantum capacity, NS-assisted quantum capacity, NS-assisted quantum simulation cost and entanglement-assisted quantum simulation cost, respectively. The QRST [BDH+14, BCR11] shows that the quantum simulation cost is equal to its quantum capacity under entanglement-assistance, i.e., $Q_E\left(\mathcal{N}\right) = S_E\left(\mathcal{N}\right)$. The QRST under NS-assistance means that $Q_{\mathrm{NS}}\left(\mathcal{N}\right) = S_{\mathrm{NS}}\left(\mathcal{N}\right)$.

The AEP of the channel's smooth max-information is the claim that

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}I^{\varepsilon}_{\max}\left(A:B\right)_{\mathcal{N}^{\otimes n}} = I\left(A:B\right)_{\mathcal{N}}, \tag{5.75}$$

where $I\left(A:B\right)_{\mathcal{N}} := \max_{\rho_A\in\mathcal{S}_{=}(A)} I\left(A:B\right)_{\mathcal{N}_{A'\to B}(\phi_{AA'})}$ is the mutual information of the quantum channel, $\phi_{AA'}$ is a purification of ρ_A and

$$I\left(A:B\right)_{\rho} = \min_{\sigma_B\in\mathcal{S}_{=}(B)} D\left(\rho_{AB}\|\rho_A\otimes\sigma_B\right) \tag{5.76}$$

is the mutual information of a quantum state. Based on the operational interpretation of the channel's max-information (cf. Theorem 5.21) and the known result $Q_E\left(\mathcal{N}\right) = \frac{1}{2}I\left(A:B\right)_{\mathcal{N}}$ [BSST02], we have

$$\text{AEP (5.75)} \iff Q_E\left(\mathcal{N}\right) = S_{\mathrm{NS}}\left(\mathcal{N}\right). \tag{5.77}$$

Thus, the QRST implies the AEP for the channel's smooth max-information. We now directly prove the AEP in Theorem 5.26. This proof then also implies the QRST in the presence of non-signalling correlations.

In the following, we will utilize various smooth entropies. Recall that the smooth max-information of a quantum state is defined as

$$I^{\varepsilon}_{\max}\left(A:B\right)_{\rho} := \min_{\widehat{\rho}\approx^{\varepsilon}\rho} I_{\max}\left(A:B\right)_{\widehat{\rho}}. \tag{5.78}$$

We also employ the following variation

$$\widehat{I}^{\varepsilon}_{\max}\left(A:B\right)_{\rho} := \min_{\substack{\widehat{\rho}\approx^{\varepsilon}\rho\\ \widehat{\rho}_A=\rho_A}} I_{\max}\left(A:B\right)_{\widehat{\rho}}, \tag{5.79}$$

which has the relation [ABJT18]

$$\widehat{I}^{\varepsilon}_{\max}(A:B)_{\rho} \leq I^{\varepsilon/6}_{\max}(A:B)_{\rho} + g(\varepsilon) \quad \text{with } g(\varepsilon) = O(\log(1/\varepsilon)). \tag{5.80}$$

Theorem 5.26 *For any quantum channel $\mathcal{N}_{A'\to B}$ we have the AEP for the channel's smooth max-information:*

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}I^{\varepsilon}_{\max}(A:B)_{\mathcal{N}^{\otimes n}} = I(A:B)_{\mathcal{N}}. \tag{5.81}$$

Proof The proof strategy is as follows. We first use the post-selection technique to show that the channel's smooth max-information is upper bounded by the quantity in Eq. (5.79). By Eq. (5.80) we can then use the properties of the smooth max-information developed in [BCR11] to show one direction of the proof. The other direction can be proved via the continuity of the mutual information of quantum states.

Consider n uses of the channel $\mathcal{N}$ and let $\omega^n_{RAA'}$ be the purification of the de Finetti state $\omega^n_{AA'} = \int \sigma^{\otimes n}_{AA'} d(\sigma_{AA'})$ with pure states $\sigma_{AA'} = |\sigma\rangle\langle\sigma|_{AA'}$ and $d(\cdot)$ the measure on the normalized pure states induced by the Haar measure. Furthermore we can assume without loss of generality that $|R| \leq (n+1)^{|A|^2-1}$. Note that $\omega^n_{A'}$ is a full rank state. We have the following inequality chain

$$I^{\varepsilon}_{\max}(A:B)_{\mathcal{N}^{\otimes n}} = \min_{\substack{\frac{1}{2}\|\widetilde{\mathcal{N}}^n - \mathcal{N}^{\otimes n}\|_{\diamond}\leq\varepsilon \\ \widetilde{\mathcal{N}}^n\in\mathrm{CPTP}(A'^n:B^n)}} I_{\max}(RA:B)_{\widetilde{\mathcal{N}}^n(\omega^n_{RAA'})}, \tag{5.82}$$

$$\leq \min_{\substack{\frac{1}{2}\|(\widetilde{\mathcal{N}}^n - \mathcal{N}^{\otimes n})(\omega^n_{RAA'})\|_1\leq\varepsilon_1 \\ \widetilde{\mathcal{N}}^n\in\mathrm{CPTP}(A'^n:B^n)}} I_{\max}(RA:B)_{\widetilde{\mathcal{N}}^n(\omega^n_{RAA'})}, \tag{5.83}$$

$$\leq \min_{\substack{\widetilde{\mathcal{N}}^n(\omega^n_{RAA'})\approx^{\varepsilon_2}\mathcal{N}^{\otimes n}(\omega^n_{RAA'}) \\ \widetilde{\mathcal{N}}^n\in\mathrm{CPTP}(A'^n:B^n)}} I_{\max}(RA:B)_{\widetilde{\mathcal{N}}^n(\omega^n_{RAA'})}, \tag{5.84}$$

where $\varepsilon_1 = \varepsilon(n+1)^{-(|A'|^2-1)}$, $\varepsilon_2 = \varepsilon_1$. In the first equality, we choose $\omega^n_{RAA'}$ as the input state of the channel's max-information. In the first inequality, we use the post-selection technique (cf. [BCR11, Prop. D.4]). This relaxes the diamond norm to the trace norm. In the second inequality, we replace the trace norm with the purified distance since $\frac{1}{2}\|\rho-\sigma\|_1 \leq P(\rho,\sigma)$.

From Lemma 5.28 below, we know that for any pure state $\phi_{AA'}$, the following two sets are the same,

$$\left\{\widetilde{\mathcal{N}}^n(\omega^n_{RAA'}) \approx^{\varepsilon_2} \mathcal{N}^{\otimes n}(\omega^n_{RAA'}) \;\middle|\; \widetilde{\mathcal{N}}^n \in \mathrm{CPTP}(A'^n:B^n)\right\}$$
$$= \left\{\sigma^n_{RAB} \approx^{\varepsilon_2} \mathcal{N}^{\otimes n}(\omega^n_{RAA'}) \;\middle|\; \sigma^n_{RA} = \omega^n_{RA}\right\}. \tag{5.85}$$

Combining Eqs. (5.84) and (5.85), we have

$$I_{\max}^{\varepsilon}\left(A:B\right)_{\mathcal{N}^{\otimes n}} \leq \widehat{I}_{\max}^{\varepsilon_2}\left(RA:B\right)_{\mathcal{N}^{\otimes n}\left(\omega_{RAA'}^{n}\right)}. \tag{5.86}$$

From Eq. (5.80), denote $\varepsilon_3 = \varepsilon_2/6$, we have

$$I_{\max}^{\varepsilon}\left(A:B\right)_{\mathcal{N}^{\otimes n}} \leq I_{\max}^{\varepsilon_3}\left(RA:B\right)_{\mathcal{N}^{\otimes n}\left(\omega_{RAA'}^{n}\right)} + g\left(\varepsilon_2\right). \tag{5.87}$$

Then we can use some known properties of the smooth max-information from [CBR14, BCR11], which leads to

$$\begin{aligned}
&I_{\max}^{\varepsilon_3}\left(RA:B\right)_{\mathcal{N}^{\otimes n}\left(\omega_{RAA'}^{n}\right)} \\
&\quad\leq I_{\max}^{\varepsilon_4}\left(B:RA\right)_{\mathcal{N}^{\otimes n}\left(\omega_{RAA'}^{n}\right)} + f\left(\varepsilon_4,\varepsilon_4\right) && (5.88)\\
&\quad\leq I_{\max}^{\varepsilon_4}\left(B:A\right)_{\mathcal{N}^{\otimes n}\left(\omega_{AA'}^{n}\right)} + 2\log|R| + f\left(\varepsilon_4,\varepsilon_4\right) && (5.89)\\
&\quad= I_{\max}^{\varepsilon_4}\left(B:A\right)_{\mathcal{N}^{\otimes n}\left(\sum_{i\in I} p_i\left(\sigma_{AA'}^{i}\right)^{\otimes n}\right)} + 2\log|R| + f\left(\varepsilon_4,\varepsilon_4\right) && (5.90)\\
&\quad\leq \max_{\sigma_{AA'}^{i}} I_{\max}^{\varepsilon_4}\left(B:A\right)_{\mathcal{N}^{\otimes n}\left(\left(\sigma_{AA'}^{i}\right)^{\otimes n}\right)} + \log|I| + 2\log|R| + f\left(\varepsilon_4,\varepsilon_4\right) && (5.91)\\
&\quad\leq \max_{\sigma_{AA'}} I_{\max}^{\varepsilon_4}\left(B:A\right)_{\mathcal{N}^{\otimes n}\left(\sigma_{AA'}^{\otimes n}\right)} + \log|I| + 2\log|R| + f\left(\varepsilon_4,\varepsilon_4\right), && (5.92)
\end{aligned}$$

where $\varepsilon_4 = \varepsilon_3/2$, $f\left(\varepsilon,\varepsilon'\right) := \log\left(\frac{1}{1-\sqrt{1-\varepsilon^2}} + \frac{1}{1-\varepsilon'}\right)$ and $|I| = (n+1)^{2|A||A'|-2}$. In the second line, we swap the system order according to [CBR14, Corollary 5]. In the third line, we get rid of purification system R according to [BCR11, Lemma B.12]. In the fourth line, we express the integral $\omega_{AA'}^{n} = \int \sigma_{AA'}^{\otimes n} d\left(\sigma_{AA'}\right)$ into convex combination of finite number of operators according to [BCR11, Corollary D.6]. In the fifth line, we use the quasi-convexity of the smooth max-information [BCR11, Lemma B.21]. In the last line, we relax the maximization to all pure states $\sigma_{AA'}$.

Combining Eqs. (5.87), (5.92) and the AEP for the smooth max-information from [BCR11, Lemma B.24], we get

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}I_{\max}^{\varepsilon}\left(A:B\right)_{\mathcal{N}^{\otimes n}} \leq \max_{\sigma_{AA'}}\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}I_{\max}^{\varepsilon_4}\left(B:A\right)_{\mathcal{N}^{\otimes n}\left(\sigma_{AA'}^{\otimes n}\right)} = I\left(A:B\right)_{\mathcal{N}}. \tag{5.93}$$

On the other hand, suppose the optimal solution of $I\left(A:B\right)_{\mathcal{N}}$ is taken at $\rho_{A'}$ with a purification $\phi_{AA'}$. Since we can always find a full rank state that is arbitrarily close to $\rho_{A'}$, thus it gives the mutual information arbitrarily close to $I\left(A:B\right)_{\mathcal{N}}$ due to the continuity. In the following, we can assume that $\rho_{A'}$ is of full rank without loss of generality and have the chain

of inequalities

$$
\begin{aligned}
& I_{\max}^{\varepsilon}\left(A:B\right)_{\mathcal{N}^{\otimes n}} \\
& = \min_{\substack{\frac{1}{2}\|\widetilde{\mathcal{N}}^n-\mathcal{N}^{\otimes n}\|_{\diamond}\leq\varepsilon \\ \widetilde{\mathcal{N}}^n\in \text{ CPTP}(A'^n:B^n)}} \min_{\sigma_B^n\in\mathcal{S}_{=}(B^{\otimes n})} D_{\max}\left(\widetilde{\mathcal{N}}_{A'\rightarrow B}^n\left(\phi_{AA'}^{\otimes n}\right)\|\phi_A^{\otimes n}\otimes\sigma_B^n\right) && (5.94) \\
& \geq \min_{\substack{\frac{1}{2}\|\widetilde{\mathcal{N}}^n-\mathcal{N}^{\otimes n}\|_{\diamond}\leq\varepsilon \\ \widetilde{\mathcal{N}}^n\in \text{ CPTP}(A'^n:B^n)}} \min_{\sigma_B^n\in\mathcal{S}_{=}(B^{\otimes n})} D\left(\widetilde{\mathcal{N}}_{A'\rightarrow B}^n\left(\phi_{AA'}^{\otimes n}\right)\|\phi_A^{\otimes n}\otimes\sigma_B^n\right) && (5.95) \\
& = \min_{\substack{\frac{1}{2}\|\widetilde{\mathcal{N}}^n-\mathcal{N}^{\otimes n}\|_{\diamond}\leq\varepsilon \\ \widetilde{\mathcal{N}}^n\in \text{ CPTP}(A'^n:B^n)}} I\left(A:B\right)_{\widetilde{\mathcal{N}}_{A'\rightarrow B}^n\left(\phi_{AA'}^{\otimes n}\right)} && (5.96) \\
& \geq I\left(A:B\right)_{\mathcal{N}_{A'\rightarrow B}^{\otimes n}\left(\phi_{AA'}^{\otimes n}\right)}-\left(8n\varepsilon\log|A|+2h_2\left(2\varepsilon\right)\right) && (5.97) \\
& = nI\left(A:B\right)_{\mathcal{N}_{A'\rightarrow B}\left(\phi_{AA'}\right)}-\left(8n\varepsilon\log|A|+2h_2\left(2\varepsilon\right)\right) && (5.98) \\
& = nI\left(A:B\right)_{\mathcal{N}}-\left(8n\varepsilon\log|A|+2h_2\left(2\varepsilon\right)\right), && (5.99)
\end{aligned}
$$

where $h_2\left(\cdot\right)$ is the binary entropy. In the third line, we use the fact that max-relative entropy is never smaller than the relative entropy [Dat09b]. The fourth line follows from the definition of the mutual information of a quantum state. The fifth line follows from the continuity of quantum mutual information in Lemma 5.29. The sixth line follows from the additivity of quantum mutual information. The last line follows from the assumption that $\phi_{AA'}$ is the optimizer of $I\left(A:B\right)_{\mathcal{N}}$. Finally, we have

$$
\lim_{\varepsilon\rightarrow 0}\lim_{n\rightarrow\infty}\frac{1}{n}I_{\max}^{\varepsilon}\left(A:B\right)_{\mathcal{N}^{\otimes n}}\geq I\left(A:B\right)_{\mathcal{N}}. \tag{5.100}
$$

Combining Eqs. (5.93) and (5.100), we conclude the claim. ■

Remark 5.27 Since the channel simulation cost has already been proved to be a strong converse rate [BDH+14], the AEP result should hold in a more general form:

$$
\lim_{n\rightarrow\infty}\frac{1}{n}I_{\max}^{\varepsilon}\left(A:B\right)_{\mathcal{N}^{\otimes n}}=I\left(A:B\right)_{\mathcal{N}}. \tag{5.101}
$$

But the continuity approach, specifically Eq. (5.99), in the above proof of Theorem 5.26 does not work well to obtain this more general result.

Lemma 5.28 *For any pure state $\phi_{AA'}$ and quantum state ρ_{AB} such that $\phi_A=\rho_A$, the following two sets are the same,*

$$
\left\{\mathcal{N}_{A'\rightarrow B}\left(\phi_{AA'}\right)\approx^{\varepsilon}\rho_{AB}\,|\,\mathcal{N}\in\text{ CPTP}\left(\text{A}':\text{B}\right)\right\}=\left\{\sigma_{AB}\approx^{\varepsilon}\rho_{AB}\,|\,\sigma_A=\rho_A\right\}. \tag{5.102}
$$

Proof Denote the l.h.s. and r.h.s. as $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. It is clear that $\mathcal{S}_1\subseteq\mathcal{S}_2$ and we now show the other direction. For any quantum state $\sigma_{AB}\in\mathcal{S}_2$, denote $\overline{\sigma}_{AB}=\sigma_A^{-1/2}\sigma_{AB}\sigma_A^{-1/2}$.

Then, we have $\overline{\sigma}_{AB} \geq 0$ and $\overline{\sigma}_A = \mathbb{1}_A$. From the Choi-Jamiołkowski isomorphism, we know that there exists a CPTP map $\mathcal{N}_{A'\to B}$ such that $\overline{\sigma}_{AB} = \mathcal{N}_{A'\to B}\left(\Phi_{AA'}\right)$, where $\Phi_{AA'}$ denotes the un-normalized maximally entangled state. Thus, we get $\sigma_{AB} = \mathcal{N}_{A'\to B}\left(\sigma_A^{1/2}\Phi_{AA'}\sigma_A^{1/2}\right)$. Denoting $\psi_{AA'} = \sigma_A^{1/2}\Phi_{AA'}\sigma_A^{1/2}$, we have that $\psi_{AA'}$ is a purification of σ_A and since $\sigma_A = \rho_A = \phi_A$ we get that $\phi_{AA'}$ is also a purification of σ_A. Due to Uhlmann's theorem [Uhl76], there exists a unitary U on the system A' such that $\psi_{AA'} = \mathcal{U}\left(\phi_{AA'}\right)$ with $\mathcal{U}\left(\cdot\right) = U \cdot U^\dagger$. Hence, we find $\sigma_{AB} = \mathcal{N} \circ \mathcal{U}\left(\phi_{AA'}\right) \in \mathcal{S}_1$. This completes the proof. ■

Lemma 5.29 *For any quantum states ρ_{AB} and σ_{AB} such that $\rho_A = \sigma_A$ and $\frac{1}{2}\|\rho - \sigma\|_1 \leq \varepsilon$, it holds that*

$$\left|I\left(A:B\right)_\rho - I\left(A:B\right)_\sigma\right| \leq 8\varepsilon \log|A| + 2h_2\left(2\varepsilon\right), \tag{5.103}$$

where $h_2\left(\cdot\right)$ is the binary entropy.

Proof Since $I\left(A:B\right)_\rho = H\left(A\right)_\rho - H\left(A|B\right)_\rho$, we have

$$\left|I\left(A:B\right)_\rho - I\left(A:B\right)_\sigma\right| = \left|H\left(A|B\right)_\rho - H\left(A|B\right)_\sigma\right| \leq 8\varepsilon \log|A| + 2h_2\left(2\varepsilon\right), \tag{5.104}$$

where $H\left(A\right)$ and $H\left(A|B\right)$ are von Neumann entropy and conditional entropy respectively. The second inequality follows from the Alicki-Fannes inequality [AF04].

■

5.3.5 Examples

In this section, we apply our results to some basic and important channels. For classical channels, the one-shot ε-error quantum simulation cost can be given by a linear program as shown in Eq. (5.105). Using the symmetry of the quantum depolarizing channel, we can also simplify its n-shot simulation cost as a linear program. Moreover, the zero-error simulation cost of some fundamental channels can be analytically solved.

Example 1. For any classical channel $\mathcal{N}\left(y|x\right)$, its simulation cost is given by a linear program,

$$S_{\mathrm{NS},\varepsilon}^{(1)}\left(\mathcal{N}\right) = \log\min\left\lceil\sqrt{\sum V_y}\right\rceil \tag{5.105a}$$

$$\text{s.t. } Y_{xy} \geq \widetilde{\mathcal{N}}\left(y|x\right) - \mathcal{N}\left(y|x\right), Y_{xy} \geq 0, \forall\, x, y, \tag{5.105b}$$

$$\widetilde{\mathcal{N}}\left(y|x\right) \geq 0, \forall\, x, y, \ \sum\nolimits_y \widetilde{\mathcal{N}}\left(y|x\right) = 1, \forall\, x, \tag{5.105c}$$

$$\widetilde{\mathcal{N}}\left(y|x\right) \leq V_y, \forall\, x, y, \ \sum\nolimits_y Y_{xy} \leq \varepsilon, \forall\, x. \tag{5.105d}$$

Example 2. The quantum depolarizing channel is given by $\mathcal{N}(\rho) = (1-p)\rho + p \cdot \frac{\mathbb{1}}{d}$ with dimension d. Its Choi matrix $J_{\mathcal{N}}$ commutes with any local unitary $U \otimes \overline{U}$ and $J_{\mathcal{N}}^{\otimes n}$ is invariant under any permutation of the tensor factors. Exploiting these symmetries, we can simplify the SDP (5.53) for $\mathcal{N}^{\otimes n}$ to a linear program (5.108). The results of a numerical implementation are shown in Figure 5.5. We can see that as the number of channel uses n increases, the average quantum simulation cost will approach its entanglement-assisted quantum capacity [BSST99], i.e., half of the quantum mutual information of the channel.

For the quantum depolarizing channel $\mathcal{N}(\rho) = (1-p)\rho + p \cdot \frac{\mathbb{1}}{d}$, its Choi-Jamiołkowski matrix is given by $J_{\mathcal{N}} = q_1 \Phi_d + q_2 \Phi_d^{\perp}$ where $q_1 = d(1-p) + \frac{p}{d}$, $q_2 = \frac{p}{d}$ and Φ_d is the maximally entangled state with dimension d, $\Phi_d^{\perp} = \mathbb{1} - \Phi_d$. Then we have

$$J_{\mathcal{N}}^{\otimes n} = \sum_{k=0}^{n} p_k P_k^n \left(\Phi_d, \Phi_d^{\perp}\right) \quad \text{with} \quad p_k = q_1^k q_2^{n-k}, \tag{5.106}$$

and $P_k^n\left(\Phi_d, \Phi_d^{\perp}\right)$ denotes the summation of n-fold tensor products of Φ_d and $\Phi_d^{\perp}$ with exactly k-fold of Φ_d. For example, $P_1^3\left(\Phi_d, \Phi_d^{\perp}\right) = \Phi_d^{\perp} \otimes \Phi_d^{\perp} \otimes \Phi_d + \Phi_d^{\perp} \otimes \Phi_d \otimes \Phi_d^{\perp} + \Phi_d \otimes \Phi_d^{\perp} \otimes \Phi_d^{\perp}$. Due to the symmetries of $J_{\mathcal{N}}^{\otimes n}$, we can take the optimal solution in SDP (5.53) as

$$J_{\widetilde{\mathcal{N}}^n} = \sum_{k=0}^{n} r_k P_k^n \left(\Phi_d, \Phi_d^{\perp}\right), \quad Y = \sum_{k=0}^{n} y_k P_k^n \left(\Phi_d, \Phi_d^{\perp}\right), \quad \text{and} \quad V = s\mathbb{1}. \tag{5.107}$$

Finally we have the LP as follows,

$$S_{\mathrm{NS},\varepsilon}^{(1)}\left(\mathcal{N}^{\otimes n}\right) = \log \min \left\lceil \sqrt{d^n \cdot s} \right\rceil \tag{5.108a}$$

$$\text{s.t. } y_k - r_k + p_k \geq 0, y_k \geq 0, 0 \leq r_k \leq s, \forall k \tag{5.108b}$$

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{d}\right)^k \left(d - \frac{1}{d}\right)^{n-k} r_k = 1, \tag{5.108c}$$

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{d}\right)^k \left(d - \frac{1}{d}\right)^{n-k} y_k \leq \varepsilon. \tag{5.108d}$$

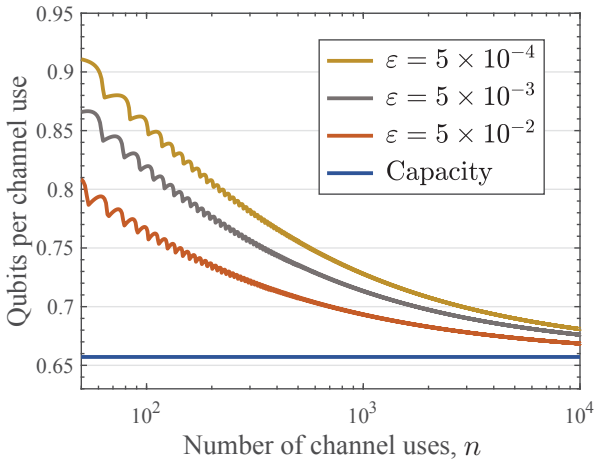


Figure 5.5 : Exact value by the linear program (5.108) of the average simulation cost for three different error tolerances $\varepsilon \in \{5\times10^{-4}, 5\times10^{-3}, 5\times10^{-2}\}$ and the qubit depolarizing channel with failure probability $p = 0.15$. The lowest line marks the entanglement-assisted quantum capacity of the channel (roughly 0.657 qubits per channel use).

Recall that the primal and dual SDPs of the zero-error simulation cost are given by [DW16]

$$\textbf{Primal:} \quad S_{\mathrm{NS},0}(\mathcal{N}) = \frac{1}{2}\log\min\left\{\operatorname{Tr} V_B \mid J_{\mathcal{N}} \leq \mathbb{1}_A \otimes V_B\right\}, \tag{5.109}$$

$$\textbf{Dual:} \quad S_{\mathrm{NS},0}(\mathcal{N}) = \frac{1}{2}\log\max\left\{\operatorname{Tr} J_{\mathcal{N}} X_{AB} \mid \operatorname{Tr}_A X_{AB} \leq \mathbb{1}_B, X_{AB} \geq 0\right\}. \tag{5.110}$$

We study some fundamental channels and show their analytical solutions by explicitly constructing feasible solutions in both primal and dual problems, respectively. Using the weak duality, we can argue that the feasible solutions we construct are optimal.

Example 3. The quantum depolarizing channel is $\mathcal{N}(\rho) = (1-p)\rho + p\cdot\frac{\mathbb{1}}{d}$ with dimension d. Taking

$$V_B = \left(d(1-p) + \frac{p}{d}\right)\mathbb{1}_B, \quad \text{and} \quad X_{AB} = \sum_{i,j=0}^{d-1} |ii\rangle\langle jj|, \tag{5.111}$$

in the primal and dual problems respectively, we can verify that they are feasible solutions. Thus, we have

$$\begin{aligned}
\frac{1}{2}\log\left(d^2(1-p)+p\right) &= \frac{1}{2}\log\operatorname{Tr} J_{\mathcal{N}} X_{AB} && (5.112)\\
&\leq S_{\mathrm{NS},0}(\mathcal{N}) && (5.113)\\
&\leq \frac{1}{2}\log\operatorname{Tr} V_B && (5.114)\\
&= \frac{1}{2}\log\left(d^2(1-p)+p\right). && (5.115)
\end{aligned}$$

We find that

$$S_{\mathrm{NS},0}\left(\mathcal{N}\right) = \frac{1}{2}\log\left(d^2\left(1-p\right)+p\right). \tag{5.116}$$

Example 4. The amplitude damping channel is $\mathcal{N}\left(\rho\right) = \sum_{i=0}^{1} E_i \rho E_i^\dagger$ with $E_0 = |0\rangle\langle 0| + \sqrt{1-r}|1\rangle\langle 1|$, $E_1 = \sqrt{r}|0\rangle\langle 1|$ and $0 \leq r \leq 1$. The optimal solutions are given by

$$V_B = \left(1+\sqrt{1-r}\right)|0\rangle\langle 0| + \left(\sqrt{1-r}+1-r\right)|1\rangle\langle 1| \tag{5.117}$$

$$X_{AB} = \left(|00\rangle + |11\rangle\right)\left(\langle 00| + \langle 11|\right). \tag{5.118}$$

We find that

$$S_{\mathrm{NS},0}\left(\mathcal{N}\right) = \frac{1}{2}\log\left(2\left(1+\sqrt{1-r}\right)-r\right). \tag{5.119}$$

Example 5. The dephasing channel is $\mathcal{N}\left(\rho\right) = \left(1-p\right)\rho + pZ\rho Z$ with $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$. The optimal solutions are given by

$$V_B = \left(|2p-1|+1\right)\mathbb{1}_B \quad \text{and} \quad X_{AB} = \left(|00\rangle + |11\rangle\right)\left(\langle 00| + \langle 11|\right). \tag{5.120}$$

We find that

$$S_{\mathrm{NS},0}\left(\mathcal{N}\right) = \frac{1}{2}\log\left(|4p-2|+2\right). \tag{5.121}$$

Example 6. The quantum erasure channel is $\mathcal{N}\left(\rho\right) = \left(1-p\right)\rho + p|e\rangle\langle e|$ with $|e\rangle$ orthogonal to the input Hilbert space. The optimal solutions are given by

$$V_B = d\left(1-p\right)\sum_{i,j=0}^{d-1}|i\rangle\langle i| + p|d\rangle\langle d|, \tag{5.122}$$

$$X_{AB} = \sum_{i,j=0}^{d-1}|ii\rangle\langle jj| + \frac{1}{d}\sum_{i=0}^{d-1}|i\rangle\langle i|\otimes|d\rangle\langle d|. \tag{5.123}$$

We find that

$$S_{\mathrm{NS},0}\left(\mathcal{N}\right) = \frac{1}{2}\log\left(d^2\left(1-p\right)+p\right). \tag{5.124}$$

5.4 Discussion

5.4.1 Summary of results

1. The zero-error coherence simulation cost under MIO is given by

$$S_{c,\mathrm{MIO}}^{(1),0}(\mathcal{N}) = \max_{\rho\in\mathcal{I}} C_{\max}(\mathcal{N}(\rho)) + \delta,$$
$$S_{c,\mathrm{MIO}}^{0}(\mathcal{N}) = \max_{\rho\in\mathcal{I}} C_{\max}(\mathcal{N}(\rho)),$$

   where $\delta \in [0,1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

2. We have introduced the channel's max-relative entropy and shown that the one-shot ε-error coherence simulation cost under MIO can be given by

$$S_{c,\mathrm{MIO}}^{(1),\varepsilon}(\mathcal{N}) = \min_{\mathcal{M}\in\mathrm{MIO}} D_{\max}^{\varepsilon}(\mathcal{N}||\mathcal{M}) + \delta,$$

   where $\delta \in [0,1]$ is the least constant such that the r.h.s. is the logarithm of an integer.

3. We have introduced the channel's max-information, the channel's log-robustness and shown that the one-shot ε-error quantum simulation cost is given by

$$S_{\mathrm{NS},\varepsilon}^{(1)}(\mathcal{N}) = \frac{1}{2} I_{\max}^{\varepsilon}(A:B)_{\mathcal{N}} + \delta = \frac{1}{2}\mathcal{LR}_g(\mathcal{N}) + \delta = \frac{1}{2}\min_{\mathcal{M}\in\boldsymbol{\mathcal{G}}} D_{\max}^{\varepsilon}(\mathcal{N}\|\mathcal{M}) + \delta,$$

   where $\delta \in [0,1]$ is the least constant such that the last three terms are the logarithm of an integer.

4. The AEP for the channel's smooth max-information holds

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n} I_{\max}^{\varepsilon}(A:B)_{\mathcal{N}^{\otimes n}} = I(A:B)_{\mathcal{N}}.$$

5. The AEP for the channel's smooth max-information implies the quantum reverse Shannon theorem under non-signalling assistance.

5.4.2 Outlook

The first part of this chapter has studied the minimal amount of coherent resources required to implement a given quantum channel. It shows that the coherence simulation cost can be

succinctly expressed by its minimum "distance" to the set of MIO operations, while the distance measure is characterized by the channel's max-relative entropy. However, we do not know if this "distance" type of characterization holds for other sets of free operations. Like any resource theory framework, another interesting question is to ask whether the maximal coherence generated from a channel is equal to the coherence consumed to implement this channel. Specifically, it asks whether a channel's coherence-generating capacity is equal to its coherence simulation cost. At this moment we only know that the reversibility holds for any classical-quantum channels, while the general result is still open.

In the second part of this chapter we have studied NS-assisted channel simulation via quantum noiseless channels and proved that it asymptotically converges to the entanglement-assisted capacity. Since the latter allows a single-letter characterization, it is natural to consider a second-order refinement thereof. A second-order expansion of an achievable rate was established in [DTW16] but no matching second-order converse bound is known. Our one-shot NS-assisted quantum simulation cost may provide some insights in this direction.

Bibliography

[Abe06] J. Aberg. Quantifying superposition. *arXiv:quant-ph/0612146*, 2006.

[ABG+07] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. Device-independent security of quantum cryptography against collective attacks. *Physical Review Letters*, 98(23):230501, 2007.

[ABJT18] A. Anshu, M. Berta, R. Jain, and M. Tomamichel. Partially smoothed information measures. *arXiv:1807.05630*, 2018.

[AD15] K. M. R. Audenaert and N. Datta. α-z-Rényi relative entropies. *Journal of Mathematical Physics*, 56(2):22202, 2015.

[ADVW02] K. Audenaert, B. De Moor, K. G. H. Vollbrecht, and R. F. Werner. Asymptotic relative entropy of entanglement for orthogonally invariant states. *Physical Review A*, 66(3):032310, 2002.

[AF04] R. Alicki and M. Fannes. Continuity of quantum conditional information. *Journal of Physics A: Mathematical and General*, 37(5):L55–L57, 2004.

[AJM+16] A. Anshu, R. Jain, P. Mukhopadhyay, A. Shayeghi, and P. Yao. New one shot quantum protocols with application to communication complexity. *IEEE Transactions on Information Theory*, 62(12):7566–7577, 2016.

[AJS18] A. Anshu, R. Jain, and A. Streltsov. Quantum state redistribution with local coherence. *arXiv:1804.04915*, 2018.

[AOP16] A. M. Alhambra, J. Oppenheim, and C. Perry. Fluctuating states: What is the probability of a thermodynamical transition? *Physical Review X*, 6:041016, 2016.

[APE03] K. Audenaert, M. B. Plenio, and J. Eisert. Entanglement cost under Positive-Partial-Transpose-Preserving operations. *Physical Review Letters*, 90(2):027901, 2003.

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *International Conference on Computers, Systems & Signal Processing, Bangalore, India, Dec 9-12, 1984*, pages 175–179, 1984.

[BBC+93] 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.

[BBCW13] M. Berta, F. G. S. L. Brandão, M. Christandl, and S. Wehner. Entanglement cost of quantum channels. *IEEE Transactions on Information Theory*, 59(10):6779–6795, 2013.

[BBP+96] 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. *Physical Review Letters*, 76:722–725, 1996.

[BBPS96] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Physical Review A*, 53:2046–2052, 1996.

[BCP14] T. Baumgratz, M. Cramer, and M. B. Plenio. Quantifying coherence. *Physical Review Letters*, 113:140401, 2014.

[BCR11] M. Berta, M. Christandl, and R. Renner. The quantum reverse Shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306(3):579, 2011.

[BCT16] M. Berta, M. Christandl, and D. Touchette. Smooth entropy bounds on one-shot quantum state redistribution. *IEEE Transactions on Information Theory*, 62(3):1425–1439, 2016.

[BD10a] F. Buscemi and N. Datta. Distilling entanglement from arbitrary resources. *Journal of Mathematical Physics*, 51(10):102201, 2010.

[BD10b] F. Buscemi and N. Datta. The quantum capacity of channels with arbitrarily correlated noise. *IEEE Transactions on Information Theory*, 56(3):1447–1460, 2010.

[BD11] F. G. S. L. Brandão and N. Datta. One-shot rates for entanglement manipulation under non-entangling maps. *IEEE Transactions on Information Theory*, 57(3):1754–1760, 2011.

[BD13] F. Buscemi and N. Datta. General theory of environment-assisted entanglement distillation. *IEEE Transactions on Information Theory*, 59(3):1940–1954, 2013.

[BDF+99] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters. Quantum nonlocality without entanglement. *Physical Review A*, 59(2):1070, 1999.

[BDH+14] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter. The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Transactions on Information Theory*, 60(5):2926–2959, 2014.

[BDL16] S. Beigi, N. Datta, and F. Leditzky. Decoding quantum information via the Petz recovery map. *Journal of Mathematics and Physics*, 57(8):082203, 2016.

[BDSW96] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824, 1996.

[Bel64] J. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[BG15] F. G. S. L. Brandão and G. Gour. Reversible framework for quantum resource theories. *Physical Review Letters*, 115:070503, 2015.

[BG17] F. Buscemi and G. Gour. Quantum relative Lorenz curves. *Physical Review A*, 95:012110, 2017.

[BGMW17] K. Ben Dana, M. García Díaz, M. Mejatty, and A. Winter. Resource theory of coherence: Beyond states. *Physical Review A*, 95(6):062327, 2017.

[BP08] F. G. S. L. Brandão and M. B. Plenio. Entanglement theory and the second law of thermodynamics. *Nature Physics*, 4:873–877, 2008.

[BP10] F. G. S. L. Brandão and M. B. Plenio. A generalization of quantum Stein's lemma. *Communications in Mathematical Physics*, 295(3):791–828, 2010.

[BPM+97] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger. Experimental quantum teleportation. *Nature*, 390(6660):575, 1997.

[BRG+16] F. G. S. L. Brandão, R. Ramanathan, A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, T. Szarek, and H. Wojewódka. Realistic noise-tolerant randomness amplification using finite number of devices. *Nature Communications*, 7:11345, 2016.

[BS10] G. Benenti and G. Strini. Computing the distance between quantum channels: usefulness of the fano representation. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 43(21):215508, 2010.

[BSF+17] K. Bu, U. Singh, S.-M. Fei, A. K. Pati, and J. Wu. Maximum relative entropy of coherence: An operational coherence measure. *Physical Review Letters*, 119(15):150405, 2017.

[BSST99] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters*, 83(15):3081, 1999.

[BSST02] 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 Transactions on Information Theory*, 48(10):2637–2655, 2002.

[BSW16] K. Bu, U. Singh, and J. Wu. Catalytic coherence transformations. *Physical Review A*, 93:042326, 2016.

[BW92] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881–2884, 1992.

[Cam07] E. T. Campbell. Distributed quantum-information processing with minimal local resources. *Physical Review A*, 76(4):040302, 2007.

[CBR14] N. Ciganovic, N. J. Beaudry, and R. Renner. Smooth max-information as one-shot generalization for mutual information. *IEEE Transactions on Information Theory*, 60(3):1573–1581, 2014.

[CDP08] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Transforming quantum operations: Quantum supermaps. *EPL (Europhysics Letters)*, 83(3):30004, 2008.

[CG16a] E. Chitambar and G. Gour. Comparison of incoherent operations and measures of coherence. *Physical Review A*, 94:052336, 2016.

[CG16b] E. Chitambar and G. Gour. Critical examination of incoherent operations and a physically consistent resource theory of quantum coherence. *Physical Review Letters*, 117:030401, 2016.

[CH16] E. Chitambar and M.-H. Hsieh. Relating the resource theories of entanglement and quantum coherence. *Physical Review Letters*, 117:020402, 2016.

[CH17] H.-C. Cheng and M.-H. Hsieh. Moderate deviation analysis for classical-quantum channels and quantum hypothesis testing. *arXiv:1701.03195*, 2017.

[Chi18] E. Chitambar. Dephasing-covariant operations enable asymptotic reversibility of quantum resources. *Physical Review A*, 97:050301, 2018.

[Cho75] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, 1975.

[CKR09] M. Christandl, R. König, and R. Renner. Postselection technique for quantum channels with applications to quantum cryptography. *Physical Review Letters*, 102(2):020504, 2009.

[CLMW11] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter. Zero-error channel capacity and simulation assisted by non-Local correlations. *IEEE Transactions on Information Theory*, 57(8):5509–5523, 2011.

[CML16] P. J. Coles, E. M. Metodiev, and N. Lütkenhaus. Numerical approach for unstructured quantum key distribution. *Nature Communications*, 7:11712, 2016.

[Col09] R. Colbeck. Quantum and relativistic protocols for secure multi-party computation. *arXiv:0911.3814*, 2009.

[CR12] R. Colbeck and R. Renner. Free randomness can be amplified. *Nature Physics*, 8(6):450, 2012.

[CSR+16] E. Chitambar, A. Streltsov, S. Rana, M. Bera, G. Adesso, and M. Lewenstein. Assisted distillation of quantum coherence. *Physical Review Letters*, 116:070402, 2016.

[CTSL06] L. Childress, J. M. Taylor, A. S. Sørensen, and M. D. Lukin. Fault-tolerant quantum communication based on solid-state photon emitters. *Physical Review Letters*, 96(7):070504, 2006.

[CTT17] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel. Moderate deviation analysis for classical communication over quantum channels. *Communications in Mathematical Ph*, 355(3):1283–1315, 2017.

[CW04] M. Christandl and A. Winter. Squashed entanglement: An additive entanglement measure. *Journal of Mathematics and Physics*, 45(3):829–840, 2004.

[Dat09a] N. Datta. Max-relative entropy of entanglement, alias log robustness. *International Journal of Quantum Information*, 7(02):475–491, 2009.

[Dat09b] N. Datta. Min- and Max-Relative Entropies and a New Entanglement Monotone. *IEEE Transactions on Information Theory*, 55(6):2816–2826, jun 2009.

[DBCZ99] W. Dür, H.-J. Briegel, J. I. Cirac, and P. Zoller. Quantum repeaters based on entanglement purification. *Physical Review A*, 59(1):169–181, 1999.

[DBG15] S. Du, Z. Bai, and Y. Guo. Conditions for coherence transformations under incoherent operations. *Physical Review A*, 91(5):052120, 2015.

[DDMW17] K. B. Dana, M. G. Díaz, M. Mejatty, and A. Winter. Resource theory of coherence: Beyond states. *Physical Review A*, 95(6):062327, 2017.

[DFR16] F. Dupuis, O. Fawzi, and R. Renner. Entropy accumulation. *arXiv:1607.01796*, 2016.

[DFW+18] M. G. Díaz, K. Fang, X. Wang, M. Rosati, M. Skotiniotis, J. Calsamiglia, and A. Winter. Using and reusing coherence to realize quantum processes. *arXiv:1805.04045*, 2018.

[DH13] N. Datta and M.-H. Hsieh. One-shot entanglement-assisted quantum and classical communication. *IEEE Transactions on Information Theory*, 59(3):1929–1939, 2013.

[DHW08] I. Devetak, A. W. Harrow, and A. J. Winter. A resource framework for quantum Shannon theory. *IEEE Transactions on Information Theory*, 54(10):4587–4618, 2008.

[DJ92] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A*, 439(1907):553–558, 1992.

[DKF+14] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner. Generalized entropies. In *XVIIth International Congress on Mathematical Physics*, pages 134–153. World Scientific, 2014.

[DL15] N. Datta and F. Leditzky. Second-order asymptotics for source coding, dense coding, and pure-state entanglement conversions. *IEEE Transactions on Information Theory*, 61(1):582–608, 2015.

[DMHB13] N. Datta, M. Mosonyi, M. Hsieh, and F. G. S. L. Brandão. A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels. *IEEE Transactions on Information Theory*, 59(12):8014–8026, 2013.

[DRRW13] N. Datta, J. M. Renes, R. Renner, and M. M. Wilde. One-shot lossy quantum data compression. *IEEE Transactions on Information Theory*, 59(12):8057–8076, 2013.

[DTW16] N. Datta, M. Tomamichel, and M. M. Wilde. On the second-order asymptotics for entanglement-assisted communication. *Quantum Information Processing*, pages 1–23, 2016.

[dVS17] J. I. de Vicente and A. Streltsov. Genuine quantum coherence. *Journal of Physics A: Mathematical and Theoretical*, 50:045301, 2017.

[DW04] I. Devetak and A. Winter. Distilling common randomness from bipartite quantum states. *IEEE Transactions on Information Theory*, 50(12):3183–3196, 2004.

[DW05] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 461, pages 207–235. The Royal Society, 2005.

[DW16] R. Duan and A. Winter. No-signalling-assisted zero-error capacity of quantum channels and an information theoretic interpretation of the Lovász number. *IEEE Transactions on Information Theory*, 62(2):891–914, 2016.

[EdMFD11] B. Escher, R. de Matos Filho, and L. Davidovich. General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology. *Nature Physics*, 7(5):406, 2011.

[Eke91] A. K. Ekert. Quantum cryptography based on bells theorem. *Physical Review Letters*, 67(6):661, 1991.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

[FD11] F. Fröwis and W. Dür. Stable macroscopic quantum superpositions. *Physical Review Letters*, 106(11):110402, 2011.

[FF18] H. Fawzi and O. Fawzi. Efficient optimization of the quantum relative entropy. *Journal of Physics A: Mathematical and Theoretical*, 51(15):154003, 2018.

[FSP18] H. Fawzi, J. Saunderson, and P. A. Parrilo. Semidefinite approximations of the matrix logarithm. *Foundations of Computational Mathematics*, 2018.

[FWL+18] K. Fang, X. Wang, L. Lami, B. Regula, and G. Adesso. Probabilistic Distillation of Quantum Coherence. *Physical Review Letters*, 121(7):070404, 2018.

[FWTD17] K. Fang, X. Wang, M. Tomamichel, and R. Duan. Non-asymptotic entanglement distillation. *arXiv:1706.06221*, 2017.

[GJC12] D. Gottesman, T. Jennewein, and S. Croke. Longer-baseline telescopes using quantum repeaters. *Physical Review Letters*, 109(7):1–5, 2012.

[Gla63] R. J. Glauber. Coherent and incoherent states of the radiation field. *Physical Review*, 131(6):2766, 1963.

[GLM11] V. Giovannetti, S. Lloyd, and L. Maccone. Advances in quantum metrology. *Nature Photonics*, 5(4):222, 2011.

[GMDLT+13] R. Gallego, L. Masanes, G. De La Torre, C. Dhara, L. Aolita, and A. Acín. Full randomness from arbitrarily deterministic events. *Nature Communications*, 4:2654, 2013.

[GMS09] G. Gour, I. Marvian, and R. W. Spekkens. Measuring the quality of a quantum reference frame: The relative entropy of frameness. *Physical Review A*, 80(1):012307, 2009.

[GS08] G. Gour and R. W. Spekkens. The resource theory of quantum reference frames: manipulations and monotones. *New Journal of Physics*, 10:033023, 2008.

[GZFG15] M. W. Girard, Y. Zinchenko, S. Friedland, and G. Gour. Erratum: Numerical estimation of the relative entropy of entanglement [Physical Review A 82 , 052336 (2010)]. *Physical Review A*, 91(2):029901, 2015.

[Hay09] M. Hayashi. Information spectrum approach to second-order coding rate in channel coding. *IEEE Transactions on Information Theory*, 55(11):4947–4966, 2009.

[Hay17] M. Hayashi. Role of hypothesis testing in quantum information. *arXiv:1709.07701*, 2017.

[HHH00] M. Horodecki, P. Horodecki, and R. Horodecki. Asymptotic manipulations of entanglement can exhibit genuine irreversibility. *Physical Review Letters*, 84:4260–4263, 2000.

[HHHH09] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81:865–942, 2009.

[Hil16] M. Hillery. Coherence as a resource in decision problems: The Deutsch-Jozsa algorithm and a variation. *Physical Review A*, 93(1):012111, 2016.

[HP91] F. Hiai and D. Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, 143(1):99–114, 1991.

[Hua14] Y. Huang. Computing quantum discord is NP-complete. *New journal of physics*, 16(3):33027, 2014.

[IP05] S. Ishizaka and M. B. Plenio. Multiparticle entanglement manipulation under Positive Partial Transpose Preserving operations. *Physical Review A*, 71:052303, 2005.

[Jam72] A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.

[JP99] D. Jonathan and M. B. Plenio. Entanglement-assisted local manipulation of pure quantum states. *Physical Review Letters*, 83:3566–3569, 1999.

[Kit97] A. Y. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[Kra71] K. Kraus. General state changes in quantum theory. *Annals of Physics*, 64(2):311–335, 1971.

[KW04] D. Kretschmann and R. F. Werner. Tema con variazioni: Quantum channel capacity. *New Journal of Physics*, 6, 2004.

[LDS17] F. Leditzky, N. Datta, and G. Smith. Useful states and entanglement distillation. *arXiv:1701.03081*, 2017.

[Li14] K. Li. Second-order asymptotics for quantum hypothesis testing. *The Annals of Statistics*, 42(1):171–189, 2014.

[LJR15] M. Lostaglio, D. Jennings, and T. Rudolph. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. *Nature Communications*, 6:6383, 2015.

[LM14] F. Levi and F. Mintert. A quantitative theory of coherent delocalization. *New Journal of Physics*, 16:033007, 2014.

[LM15a] D. Leung and W. Matthews. On the power of PPT-preserving and non-Signalling codes. *IEEE Transactions on Information Theory*, 61(8):4486–4499, 2015.

[LM15b] D. Leung and W. Matthews. On the power of PPT-preserving and non-signalling codes. *IEEE Transactions on Information Theory*, 61:4486–4499, 2015.

[LP01a] H.-K. Lo and S. Popescu. Concentrating entanglement by local actions: Beyond mean values. *Physical Review A*, 63(2):022301, 2001.

[LP01b] H.-K. Lo and S. Popescu. Concentrating entanglement by local actions: Beyond mean values. *Physical Review A*, 63:022301, 2001.

[Mat12] W. Matthews. A linear program for the finite block length converse of Polyanskiy–Poor–Verdú via nonsignaling codes. *IEEE Transactions on Information Theory*, 58(12):7036–7044, 2012.

[MK15] A. Mani and V. Karimipour. Cohering and decohering power of quantum channels. *Phys. Rev. A*, 92:032331, 2015.

[MLDS+13] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. On quantum Rényi entropies: a new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, 2013.

[MS16] I. Marvian and R. W. Spekkens. How to quantify coherence: Distinguishing speakable and unspeakable notions. *Physical Review A*, 94:052324, 2016.

[MW14] W. Matthews and S. Wehner. Finite blocklength converse bounds for quantum channels. *IEEE Transactions on Information Theory*, 60(11):7317–7329, 2014.

[MWKZ96] K. Mattle, H. Weinfurter, P. G. Kwiat, and A. Zeilinger. Dense coding in experimental quantum communication. *Physical Review Letters*, 76(25):4656, 1996.

[NC11] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[NFB14] N. H. Nickerson, J. F. Fitzsimons, and S. C. Benjamin. Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links. *Physical Review X*, 4(4):1–17, 2014.

[ON99] T. Ogawa and H. Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Transactions on Information Theory*, 45:2486–2489, 1999.

[ON00] T. Ogawa and H. Nagaoka. Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory*, 46(7):2428–2433, 2000.

[PAM+10] S. Pironio, A. Acín, S. Massar, A. B. de La Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, et al. Random numbers certified by Bells theorem. *Nature*, 464(7291):1021, 2010.

[Pau02] V. Paulsen. *Completely bounded maps and operator algebras*, volume 78. Cambridge University Press, 2002.

[Ple05] M. B. Plenio. Logarithmic negativity: A full entanglement monotone that is not convex. *Physical Review Letters*, 95(9):090503, 2005.

[PLOB17] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi. Fundamental limits of repeaterless quantum communications. *Nature Communications*, 8:15043, 2017.

[PPV10] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.

[Rai99a] E. M. Rains. Bound on distillable entanglement. *Physical Review A*, 60:179–184, 1999.

[Rai99b] E. M. Rains. Rigorous treatment of distillable entanglement. *Physical Review A*, 60:173–178, 1999.

[Rai01] E. M. Rains. A semidefinite program for distillable entanglement. *IEEE Transactions on Information Theory*, 47(7):2921–2933, 2001.

[RBG+17] W. Rosenfeld, D. Burchardt, R. Garthoff, K. Redeker, N. Ortegel, M. Rau, and H. Weinfurter. Event-ready Bell test using entangled atoms simultaneously closing detection and locality loopholes. *Physical review letters*, 119(1):010402, 2017.

[Reg18] B. Regula. Convex geometry of quantum resource quantification. *Journal of Physics A: Mathematical and Theoretical*, 51:045303, 2018.

[Ren05] R. Renner. Security of quantum key distribution. *PhD thesis, ETH Zurich*, 6(01):1–127, 2005.

[RFWA18] B. Regula, K. Fang, X. Wang, and G. Adesso. One-shot coherence distillation. *Physical Review Letters*, 121(1):010401, 2018.

[Roc70] R. T. Rockafellar. *Convex analysis*. Princeton University Press, Princeton, 1970.

[RR11] J. M. Renes and R. Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Transactions on Information Theory*, 57(11):7377–7385, 2011.

[RST+18] F. Rozpedek, T. Schiet, L. P. Thinh, D. Elkouss, A. C. Doherty, and S. Wehner. Optimising practical entanglement distillation. *arXiv:1803.10111*, 2018.

[SAP17] A. Streltsov, G. Adesso, and M. B. Plenio. Quantum coherence as a resource. *Reviews of Modern Physics*, 89:041003, 2017.

[Sch35] E. Schrödinger. Discussion of probability relations between separated systems. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 31, pages 555–563. Cambridge University Press, 1935.

[Sch36] E. Schrödinger. Probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(3):446452, 1936.

[SCR+16] A. Streltsov, E. Chitambar, S. Rana, M. Bera, A. Winter, and M. Lewenstein. Entanglement and coherence in quantum state merging. *Physical Review Letters*, 116:240405, 2016.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(4):623–656, 1948.

[Sha56] C. E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, 1956.

[Sho94] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994.

[Sio58] M. Sion. On general minimax theorems. *Pacific J. Math*, 8(1):171–176, 1958.

[Sud63] E. Sudarshan. Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams. *Physical Review Letters*, 10(7):277, 1963.

[Tan14] V. Y. F. Tan. Asymptotic estimates in information theory with non-vanishing error probabilities. *Foundations and Trends® in Communications and Information Theory*, 10(4):1–184, 2014.

[TBR16] M. Tomamichel, M. Berta, and J. M. Renes. Quantum coding with finite resources. *Nature Communications*, 7:11419, 2016.

[TCR09] M. Tomamichel, R. Colbeck, and R. Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009.

[TH13] M. Tomamichel and M. Hayashi. A hierarchy of information quantities for finite block length analysis of quantum tasks. *IEEE Transactions on Information Theory*, 59(11):7693–7710, 2013.

[Tom12] M. Tomamichel. A framework for non-asymptotic quantum information theory. *PhD thesis, ETH Zurich, arXiv:1203.2142*, 2012.

[Tom16] M. Tomamichel. *Quantum Information Processing with Finite Resources*, volume 5 of *SpringerBriefs in Mathematical Physics*. Springer International Publishing, Cham, 2016.

[TV00] B. M. Terhal and K. G. H. Vollbrecht. Entanglement of formation for isotropic states. *Physical Review Letters*, 85(12):2625, 2000.

[Uhl76] A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[VC01] G. Vidal and J. I. Cirac. Irreversibility in asymptotic manipulations of entanglement. *Physical Review Letters*, 86:5803–5806, 2001.

[VDC02] G. Vidal, W. Dr, and J. I. Cirac. Entanglement cost of bipartite mixed states. *Physical Review Letters*, 89:027901, 2002.

[Vid99] G. Vidal. Entanglement of pure states for a single copy. *Physical Review Letters*, 83:1046–1049, 1999.

[VP98] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Physical Review A*, 57(3):1619–1633, mar 1998.

[VPRK97] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight. Quantifying entanglement. *Physical Review Letters*, 78(12):2275, 1997.

[VW01] K. G. H. Vollbrecht and R. F. Werner. Entanglement measures under symmetry. *Physical Review A*, 64(6):62307, 2001.

[VW02] G. Vidal and R. F. Werner. Computable measure of entanglement. *Physical Review A*, 65:032314, 2002.

[Wat09] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(1):217–238, 2009.

[Wat11] J. Watrous. John Watrous's lecture notes: Theory of quantum information. *Lecture Notes*, (Fall), 2011.

[Wat17] J. Watrous. *The theory of quantum information*. Cambridge University Press., 2017.

[WD16a] X. Wang and R. Duan. Improved semidefinite programming upper bound on distillable entanglement. *Physical Review A*, 94:050301, 2016.

[WD16b] X. Wang and R. Duan. On the quantum no-signalling assisted zero-error classical simulation cost of non-commutative bipartite graphs. In *IEEE ISIT*, pages 2244–2248, 2016.

[WD17a] X. Wang and R. Duan. Irreversibility of asymptotic entanglement manipulation under quantum operations completely preserving positivity of partial transpose. *Physical Review Letters*, 119:180506, 2017.

[WD17b] X. Wang and R. Duan. Nonadditivity of Rains' bound for distillable entanglement. *Physical Review A*, 95(6):062322, 2017.

[WFD17] X. Wang, K. Fang, and R. Duan. Semidefinite programming converse bounds for quantum communication. *arXiv:1709.00200*, 2017.

[WFT17] X. Wang, K. Fang, and M. Tomamichel. On converse bounds for classical communication over quantum channels. *arXiv:1709.05258*, 2017.

[WGCE16] S. Waeldchen, J. Gertis, E. T. Campbell, and J. Eisert. Renormalizing entanglement distillation. *Physical Review Letters*, 116:020502, 2016.

[Wil17] M. M. Wilde. *Quantum information theory*. Cambridge University Press, 2 edition, 2017.

[Win99] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45:2481–2485, 1999.

[Wol78] J. Wolfowitz. *Coding theorems of information theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge. Springer-Verlag, Berlin Heidelberg, 3 edition, 1978.

[WR12] L. Wang and R. Renner. One-shot classical-quantum capacity and hypothesis testing. *Physical Review Letters*, 108(20):200501, 2012.

[WTB16] 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, 2016.

[WXD18] X. Wang, W. Xie, and R. Duan. Semidefinite programming strong converse bounds for classical capacity. *IEEE Transactions on Information Theory*, 64(1):640–653, 2018.

[WY16] A. Winter and D. Yang. Operational resource theory of coherence. *Physical Review Letters*, 116:120404, 2016.

[ZFG10] Y. Zinchenko, S. Friedland, and G. Gour. Numerical estimation of the relative entropy of entanglement. *Physical Review A*, 82(5):052336, 2010.

[ZLY+18] Q. Zhao, Y. Liu, X. Yuan, E. Chitambar, and X. Ma. One-shot coherence dilution. *Physical Review Letters*, 120:070403, 2018.

Appendix A

Distance characterization of resource theory

In this chapter we summarize some other distance characterizations appeared in different resource theory. We will see that the distillation and dilution processes are closely related with the hypothesis testing relative entropy and the max-relative entropy respectively while their asymptotics are characterized by the relative entropy distance.

Entanglement theory

The relative entropy of entanglement is given by [VPRK97, VP98]:

$$E_{\text{r,SEP}}\left(\rho_{AB}\right) = \min_{\sigma\in\text{SEP(A:B)}} D\left(\rho_{AB}\|\sigma_{AB}\right). \tag{A.1}$$

The PPT-relative entropy of entanglement is given by [Rai99a]:

$$E_{\text{r,PPT}}\left(\rho_{AB}\right) = \min_{\sigma\in\text{PPT(A:B)}} D\left(\rho_{AB}\|\sigma_{AB}\right). \tag{A.2}$$

The Rains bound is given by [Rai01, ADVW02]:

$$R\left(\rho_{AB}\right) = \min_{\sigma\in\text{PPT}'\text{(A:B)}} D\left(\rho_{AB}\|\sigma_{AB}\right). \tag{A.3}$$

These are bounds for LOCC and PPT-assisted distillable entanglement, i.e.,

$$E_{\text{D,LOCC}}\left(\rho_{AB}\right) \leq E_{\text{D,PPT}}\left(\rho_{AB}\right) \leq R\left(\rho_{AB}\right) \leq E_{\text{r,PPT}}\left(\rho_{AB}\right) \leq E_{\text{r,SEP}}\left(\rho_{AB}\right). \tag{A.4}$$

The one-shot distillable entanglement under separable-preserving operations is bounded by [BD11]

$$\left\lfloor \min_{\sigma\in\text{SEP}(A:B)} D_H^{\varepsilon}\left(\rho_{AB}\|\sigma_{AB}\right) \right\rfloor \leq E_{\text{D,SEPP}}^{(1),\varepsilon}\left(\rho_{AB}\right) \leq \min_{\sigma\in\text{SEP}(A:B)} D_H^{\varepsilon}\left(\rho_{AB}\|\sigma_{AB}\right). \tag{A.5}$$

The one-shot entanglement cost under separable-preserving operations is bounded by [BD11]

$$\min_{\sigma\in\text{SEP}(A:B)} D_{\max}^{\varepsilon}\left(\rho_{AB}\|\sigma_{AB}\right) \leq E_{\text{D,SEPP}}^{(1),\varepsilon}\left(\rho_{AB}\right) \leq \min_{\sigma\in\text{SEP}(A:B)} D_{\max}^{\varepsilon}\left(\rho_{AB}\|\sigma_{AB}\right) + 1. \tag{A.6}$$

Coherence theory

The asymptotic distillable coherence [WY16, RFWA18] and coherence cost [ZLY+18, Chi18] under MIO and DIO are given by

$$C_{d,\mathrm{DIO}}(\rho) = C_{d,\mathrm{MIO}}(\rho) = C_{c,\mathrm{MIO}}(\rho) = C_{c,\mathrm{DIO}}(\rho) = C_r(\rho) := \min_{\sigma\in\mathcal{I}} D(\rho\|\sigma). \tag{A.7}$$

The one-shot coherence cost under MIO is bounded by

$$\min_{\sigma\in\mathcal{I}} D_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma) \leq C_{c,\mathrm{MIO}}(\rho) \leq \min_{\sigma\in\mathcal{I}} D_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma) + 1. \tag{A.8}$$

The one-shot coherence cost under DIO is bounded by

$$\min_{\sigma\in\overline{A}_\rho} D_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma) \leq C_{c,\mathrm{DIO}}(\rho) \leq \min_{\sigma\in\overline{A}_\rho} D_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma) + 1, \tag{A.9}$$

where $A_\rho := \left\{ \frac{1}{t}\left((1+t)\Delta(\rho) - \rho\right) \mid t > 0,\ (1+t)\Delta(\rho) - \rho \geq 0 \right\}$ and $\overline{A}_\rho$ is the closure.

Channel capacity

There are two sets of useless operations in terms of channel capacity. The set of constant channels:

$$\boldsymbol{\mathcal{G}} := \{ \mathcal{M} \in \mathrm{CPTP}(A:B) \mid \exists\, \sigma \in \mathcal{S}_=(B) \text{ s.t. } \mathcal{M}(\rho) = \sigma, \forall \rho \in \mathcal{S}_=(A) \}, \tag{A.10}$$

and the set of constant-bounded subchannels:

$$\boldsymbol{\mathcal{V}} := \{\mathcal{M} \in \mathrm{CP}(A:B) : \exists\, \sigma \in \mathcal{S}_=(B) \text{ s.t. } \mathcal{M}(\rho) \leq \sigma, \forall \rho \in \mathcal{S}_=(A)\}. \tag{A.11}$$

Definition A.1 *For any relative divergence $\boldsymbol{D}(\cdot\|\cdot)$, define its corresponding channel's divergence as*

$$\boldsymbol{D}(\mathcal{N}\|\mathcal{M}) := \max_{\rho_{A'}} \boldsymbol{D}\left(\mathcal{N}_{A'\to B}(\phi_{AA'})\,\|\,\mathcal{M}_{A'\to B}(\phi_{AA'})\right), \tag{A.12}$$

where $\phi_{AA'}$ is a purification of $\rho_{A'}$.

This is a natural generalization of divergence between two CP maps. If the relative entropy between two quantum states characterizes some kind of distance between these two states, then the definition above is the worst-case distance between two quantum channels.

Some known distance characterizations are given by:

- One-shot NS-assisted activated classical capacity of quantum channels [WFT17]:

$$C_{\mathrm{NS,a}}^{(1),\varepsilon}\left(\mathcal{N}\right)=R\left(\mathcal{N},\varepsilon\right):=\min_{\mathcal{M}\in\boldsymbol{\mathcal{G}}}D_{H}^{\varepsilon}\left(\mathcal{N}\|\mathcal{M}\right). \tag{A.13}$$

  where $R\left(\mathcal{N},\varepsilon\right)$ is the Matthews-Wehner converse bound [MW14]. This is also a converse bound for entanglement-assisted classical capacity and it it asymptotically tight,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\min_{\mathcal{M}^{n}\in\boldsymbol{\mathcal{G}}}D_{H}^{\varepsilon}\left(\mathcal{N}^{\otimes n}\|\mathcal{M}^{n}\right)=\min_{\mathcal{M}\in\boldsymbol{\mathcal{G}}}D\left(\mathcal{N}\|\mathcal{M}\right)=I\left(A:B\right)_{\mathcal{N}}=C_{E}\left(\mathcal{N}\right).$$

- Υ-information converse bound on one-shot unassisted classical capacity of quantum channels [WFT17]:

$$C^{(1),\varepsilon}\left(\mathcal{N}\right)\leq\min_{\mathcal{M}\in\boldsymbol{\mathcal{V}}}D_{H}^{\varepsilon}\left(\mathcal{N}\|\mathcal{M}\right). \tag{A.14}$$

- Υ-information converse bound on Holevo capacity and classical capacity [WFT17]:

$$\chi\left(\mathcal{N}\right)\leq\Upsilon\left(\mathcal{N}\right):=\min_{\mathcal{M}\in\boldsymbol{\mathcal{V}}}D\left(\mathcal{N}\|\mathcal{M}\right),\quad C\left(\mathcal{N}\right)\leq\Upsilon^{\infty}\left(\mathcal{N}\right). \tag{A.15}$$

Appendix B

Algorithm for the Rains bound

We provide an algorithm to numerically calculate the Rains bound with high accuracy. In particular, the calculation of upper and lower bounds for the Rains bound have near-machine precision while the final result of Rains bound itself is within error tolerance 10^{-6} by default. This algorithm closely follows the approach in [ZFG10, GZFG15] which intends to calculate the PPT-relative entropy of entanglement.

Note that the only difference between the Rains bound and the PPT-relative entropy of entanglement is the feasible set. Due to the similarity between these two quantities, we can have a similar algorithm for the Rains bound. For the sake of completeness, we will restate the main idea of this algorithm and clarify that our adjustment will work to calculate the Rains bound. In the following discussion, we will consider the natural logarithm for convenience.

The key idea for this algorithm is based on the cutting-plane method combined with semidefinite programming. Clearly, calculating the Rains bound is equivalent to the optimization problem

$$\min_{\sigma\in\mathrm{PPT}'} \left(-\operatorname{Tr}\rho\ln\sigma\right), \text{ with } \mathrm{PPT}' = \left\{\sigma \geq 0 \;\middle|\; \|\sigma^{T_B}\|_1 \leq 1\right\}. \tag{B.1}$$

If we relax the minimization over all quantum states, the optimal solution is taken at $\sigma = \rho$. Thus $-\operatorname{Tr}\rho\ln\rho$ provides a trivial lower bound on (B.1). Since the objective function is convex with respect to σ over the Rains set, its epigraph is supported by tangent hyperplanes at every interior point $\sigma^{(i)} \in \operatorname{int}\mathrm{PPT}'$. Thus we can construct a successively refined sequence of approximations to the epigraph of the objective function restricted to the interior of the Rains set.

Specifically, for an arbitrary positive definite operator X, we have a spectral decomposition $X = U_X \operatorname{diag}(\lambda_X) U_X^\dagger$ with unitary matrix U_X and diagonal matrix $\operatorname{diag}(\lambda_X)$ formed by the eigenvalues λ_X. Then we have the first-order expansion

$$\ln(X+\Delta) = \ln X + U_X\left(D(\lambda_X)\circ U_X^\dagger \Delta U_X\right)U_X^\dagger + O\left(\|\Delta\|^2\right), \tag{B.2}$$

where $\circ$ denotes the Hadamard product and $D(\lambda)$ is the Hermitian matrix given by

$$D(\lambda)_{i,j} = \begin{cases} \frac{\ln \lambda_i - \ln \lambda_j}{\lambda_i - \lambda_j}, & \lambda_i \neq \lambda_j, \\ \frac{1}{\lambda_i}, & \lambda_i = \lambda_j. \end{cases} \tag{B.3}$$

For any given set of feasible points $\{\sigma^{(i)}\}_{i=0}^{N} \subset \text{int PPT}'$, we have spectral decompositions $\sigma^{(i)} = U_{(i)} \text{diag}\left(\lambda^{(i)}\right) U_{(i)}^{\dagger}$. Then epi $(-\operatorname{Tr}\rho\ln\sigma)\,|_{\text{int PPT}'}$ is a subset of all $(\sigma, t) \in \text{int PPT}' \times \mathbb{R}$ satisfying

$$-\operatorname{Tr}\rho\left(\ln\sigma^{(i)} + U_{(i)}\left(D\left(\lambda^{(i)}\right) \circ U_{(i)}^{\dagger}\left(\sigma - \sigma^{(i)}\right)U_{(i)}\right)U_{(i)}^{\dagger}\right) \leq t,\ i = 0, \cdots, N. \tag{B.4}$$

Equivalently, we can introduce slack variables s_i on the l.h.s of Eq. (B.4) and have

$$\operatorname{Tr} E^{(i)}\sigma + t - s_i = -\operatorname{Tr}\rho\ln\sigma^{(i)} + \operatorname{Tr} E^{(i)}\sigma^{(i)}, s_i \geq 0,\ i = 0, \cdots, N, \tag{B.5}$$

where $E^{(i)} = U_{(i)}\left(D\left(\lambda^{(i)}\right) \circ U_{(i)}^{\dagger}\rho U_{(i)}\right)U_{(i)}^{\dagger}$. So the optimal value of optimization problem

$$\begin{aligned} &\min\ t && \text{(B.6a)} \\ &\text{s.t. } \operatorname{Tr} E^{(i)}\sigma + t - s_i = -\operatorname{Tr}\rho\ln\sigma^{(i)} + \operatorname{Tr} E^{(i)}\sigma^{(i)},\ i = 0, \cdots, N, && \text{(B.6b)} \\ &\quad s_i \geq 0, i = 0, \cdots, N, && \text{(B.6c)} \\ &\quad \sigma \in \text{PPT}' && \text{(B.6d)} \end{aligned}$$

provides a lower bound on (B.1). For any feasible point $\sigma^* \in \text{PPT}'$, $-\operatorname{Tr}\rho\ln\sigma^*$ provides an upper bound on (B.1). For each iteration of the algorithm, we add a interior point $\sigma^{(N+1)}$ of the Rains set to the set $\left\{\sigma^{(i)}\right\}_{i=0}^{N}$, which may lead to a tighter lower bound and update the feasible point σ^* if $\sigma^{(N+1)}$ provides a tighter upper bound. We use the variables $\overline{R}$ and $\underline{R}$ to store the upper and lower bounds. Since $\underline{R}$ and $\overline{R}$ are nondecreasing and nonincreasing, respectively, at each iteration, we can terminate the algorithm when $\underline{R}$ and $\overline{R}$ are close enough, say, less than given tolerance ε. The full algorithm is presented in Algorithm 1.

Algorithm 1 Rains bound algorithm

1: **Input:** bipartite state $\rho \in \mathcal{S}_{=}(AB)$ and dimensions of subsystem d_A, d_B
2: **Output:** Upper bound $\overline{R}$, lower bound $\underline{R}$
3: **if** $\rho \in \mathrm{PPT}'$ **then**
4: **return** $\underline{R} = \overline{R} = 0$
5: **else**
6: **initialize** $\varepsilon = 10^{-6}$, $N = 0$, $\sigma^* = \sigma^{(0)} = \frac{\mathbb{1}_{AB}}{d_A d_B}$, $\underline{R} = -\operatorname{Tr}\rho\ln\rho$, $\overline{R} = -\operatorname{Tr}\rho\ln\sigma^*$
7: **while** $\overline{R} - \underline{R} \geq \varepsilon$ **do**
8: **solve** SDP (B.6) with additional constraint $t \geq \underline{R}$.
9: store optimal solution $(\underline{t}, \underline{\sigma})$ and update lower bound $\underline{R} = \underline{t}$
10: **if** gap between upper and lower bounds is within given tolerance, $\overline{R} - \underline{R} \leq \varepsilon$ **then**
11: **return** $\underline{R}$, $\overline{R}$
12: **else**
13: add one more point $\sigma^{(N+1)}$, and **set** $N = N + 1$
14: **if** $-\operatorname{Tr}\rho\ln\sigma^{(N)} \leq -\operatorname{Tr}\rho\ln\sigma^*$ **then**
15: **update** feasible point $\sigma^* = \sigma^{(N)}$, and upper bound $\overline{R} = -\operatorname{Tr}\rho\ln\sigma^*$

Note that for the condition $\sigma \in \mathrm{PPT}'\left(\sigma \geq 0, \left\|\sigma^{T_B}\right\|_1 \leq 1\right)$, Lemma $B.1$ ensures that it can be expressed as semidefinite conditions.

Lemma B.1 $\sigma \in \mathrm{PPT}'$ *if and only if* $\sigma \geq 0$ *and there exist operators* σ_+, $\sigma_- \geq 0$ *such that* $\sigma^{T_B} = \sigma_+ - \sigma_-$ *and* $\operatorname{Tr}(\sigma_+ + \sigma_-) \leq 1$.

Proof If $\sigma \in \mathrm{PPT}'$, then $\sigma \geq 0$. Use the spectral decomposition $\sigma^{T_B} = \sigma_+ - \sigma_-$, where σ_+ and σ_- are positive operator with orthogonal support. Then $\left|\sigma^{T_B}\right| = \sigma_+ + \sigma_-$ and $\operatorname{Tr}(\sigma_+ + \sigma_-) = \left\|\sigma^{T_B}\right\|_1 \leq 1$. On the other hand, if there exist positive operators σ_+ and σ_- such that $\sigma^{T_B} = \sigma_+ - \sigma_-$ and $\operatorname{Tr}(\sigma_+ + \sigma_-) \leq 1$, then $\left\|\sigma^{T_B}\right\|_1 = \|\sigma_+ - \sigma_-\|_1 \leq \|\sigma_+\|_1 + \|\sigma_-\|_1 = \operatorname{Tr}(\sigma_+ + \sigma_-) \leq 1$. Thus $\sigma \in \mathrm{PPT}'$. ■

For given $\left\{\sigma^{(i)}\right\}_{i=0}^{N}$, the step 8 in Algorithm 1 is an SDP which can be explicitly given by

$$\min t \tag{B.7a}$$

$$\text{s.t. } \operatorname{Tr} E^{(i)}\sigma + t - s_i = -\operatorname{Tr}\rho\ln\sigma^{(i)} + \operatorname{Tr} E^{(i)}\sigma^{(i)}, i = 0, \cdots, N, \tag{B.7b}$$

$$t \geq \underline{R},\ s_i \geq 0, i = 0, \cdots, N, \tag{B.7c}$$

$$\sigma, \sigma_+, \sigma_- \geq 0,\ \sigma^{T_B} = \sigma_+ - \sigma_-,\ \operatorname{Tr}(\sigma_+ + \sigma_-) \leq 1. \tag{B.7d}$$

As for step 13, variable $\sigma^{(N+1)}$ can be given by

$$\sigma^{(N+1)} = \arg\min\left\{-\operatorname{Tr}\rho\ln\sigma : \sigma = \alpha Z + (1-\alpha)\underline{\sigma}, \alpha \in [0,1]\right\}, \tag{B.8}$$

where Z is some fixed reference point. This one-dimensional minimization can be efficiently performed using the standard derivative-based bisection scheme [ZFG10].

Using this algorithm, we can check that the Rains bound is not additive, which has been recently proved in [WD17b]. We also consider the states ρ_r in [WD17b]. Denote $\underline{R}_1$ the lower bound calculated by our algorithm for $R(\rho_r)$ and $\overline{R}_2$ the upper bound calculated by our algorithm for $R\left(\rho_r^{\otimes 2}\right)$. In Figure B.1, we can clearly observe that there is a strict gap between $\overline{R}_2$ and $2\underline{R}_1$, which implies $R\left(\rho_r^{\otimes 2}\right) \leq \overline{R}_2 < 2\underline{R}_1 \leq 2R(\rho_r)$. Since the lower and upper bounds derived from our algorithm only depend on the SDP in Eq. (B.7) and Eq. (B.8), both of which can be solved to a very high (near-machine) precision, while the maximal gap in the plot is approximately 10^{-2}. Thus our algorithm provides a direct numerical evidence (not involving any other entanglement measures) for the nonadditivity of the Rains bound.

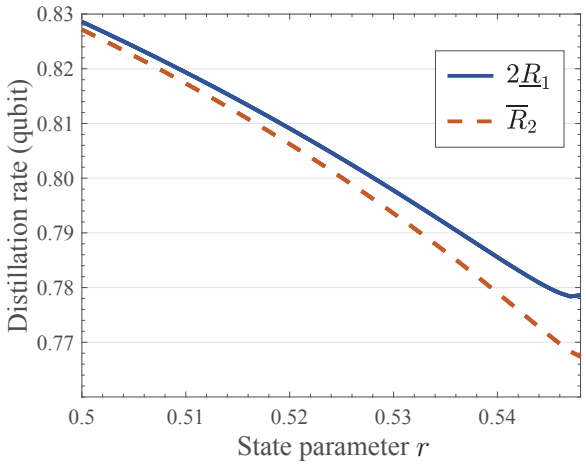


Figure B.1 : This figure shows the difference between the lower bound $2\underline{R}_1$ on $2R(\rho_r)$ and the upper bound $\overline{R}_2$ on $R\left(\rho_r^{\otimes 2}\right)$. The solid line depicts $2\underline{R}_1$ while the dashed line depicts $\overline{R}_2$. The state ρ_r is the same defined as in Ref. [WD17b]

Remark B.2 It is worth mentioning that there is another approach recently developed to efficiently calculate the Rains bound in [FF18, FSP18]. In these works, the authors make use of rational (Padé) approximations of the (matrix) logarithm function and then transform the rational functions to SDPs. Without the successive refinement, their algorithm can be much faster with relatively high accuracy. However, our algorithm is efficient enough in the case of low dimensions. We can obtain almost the same result as Figure B.1 via both methods.