Simulating quantum circuits with restricted quantum computers

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES (Dr. sc. ETH Zurich)

presented by

Christophe Piveteau

born on 23.10.1996

accepted on the recommendation of

Prof. Dr. Renato Renner, examiner Dr. Joseph M. Renes, co-examiner Prof. Dr. David Gross, co-examiner Prof. Dr. Aram Harrow, co-examiner

Acknowledgements

I would like to thank my doctoral advisor Renato Renner for granting me the outstanding opportunity to perform my doctoral studies in the quantum information theory group at ETH Zurich. The extraordinary and highly stimulating environment has been instrumental for my academic development and has sharpened my scientific skills over the last four years. I especially appreciate the freedom I was given to work on a wide variety of different topics of my choice.

I would like to convey special thanks to Joe Renes who has been a phenomenal supervisor. No matter the circumstances, he always found time for me, and he always supported me in every way necessary. I enjoyed our plentiful discussions about science and non-scientific topics, which have made my doctoral studies so much more enjoyable. Joe is an amazing researcher to collaborate with, and he is a great inspiration for the kind of scientist I aspire to become. Furthermore, I would like to express my deepest gratitude to David Sutter and Christopher Chubb, who both served as mentors to me. I am truly thankful for their guidance that has helped me navigate the scientific and non-scientific aspects of academic life.

I also want to thank my other collaborators Lukas Schmitt, Lukas Brenner, Thomas Dubach, Stefan Wörner, Kristan Temme, Sergey Bravyi and Jay Gambetta for the many fascinating research projects that I could partake in. My gratitude also extends to the present and past members of the quantum information theory group for the supportive and friendly atmosphere. I especially appreciate the numerous and highly productive discussions with Lukas Schmitt and David Sutter that were instrumental in the development of this thesis.

I also want to thank both of my sisters as well as my parents for their constant support and encouragement. Finally, and most importantly, I would like to thank my partner Anna for her unwavering support and the willingness to put up with the ups and downs that are an inevitable part the PhD journey. I am immensely grateful to have shared this period of my life with her.

Christophe Piveteau Zurich, January 2025

Abstract

It is one of the most fundamental objectives in quantum information science to understand the boundary between the computational power of classical and quantum computers. One possible avenue to explore this boundary is to identify classes of quantum circuits that can be efficiently simulated on a classical computer. In recent years, this idea has been extended one step further: Instead of simulating a general quantum circuit with a classical device, new schemes have emerged to simulate them on a *quantum* device that is restricted in some manner. As such, these techniques allow us to study how the restrictions impact the computational power of the device. One such technique is called *quasiprobability simulation* (QPS) and it estimates the result of a quantum circuit with a Monte Carlo procedure that randomly replaces circuit elements with ones that can be executed on the restricted quantum device.

The main focus of this thesis is dedicated to the QPS-based simulation of non-local quantum computation using local quantum operations. On the practical side, this enables the simulation of large quantum circuits using multiple smaller quantum devices — a procedure that is sometimes called *circuit knitting*. We uncover a rich mathematical formalism with many connections to the resource theory of entanglement. We characterize the optimal simulation overhead for a broad range of practically relevant nonlocal states and channels and we explicitly provide achieving protocols. Moreover, we also investigate the utility of classical communication between the local parties. Our results address both the single-shot and asymptotic regime.

Furthermore, this thesis also presents a comprehensive overview of recent developments of QPS. Besides the simulation of nonlocal computation, we also investigate the simulation of magic computation with a Clifford device, the simulation noise-free computation with a noisy device as well as the simulation of nonphysical operations. We frame QPS in a quantum resource theoretic framework, which highlights similarities that arise in the different instantiations of the technique. Furthermore, we study the importance of classical side information in the QPS procedure and how it impacts the overhead and expressibility of QPS.

Overall, this thesis provides a self-contained introduction for researchers interested in learning about circuit knitting and QPS.

Zusammenfassung

Es ist eines der grundlegendsten Ziele der Quanteninformationswissenschaft, die Grenze zwischen der Rechenleistung klassischer und Quantencomputer zu verstehen. Ein möglicher Ansatz, diese Grenze zu erforschen, besteht darin, Klassen von Quantenschaltkreisen zu identifizieren, die effizient auf einem klassischen Computer simuliert werden können. In den letzten Jahren wurde diese Idee um einen Schritt erweitert: Anstatt einen allgemeinen Quantenschaltkreis mit einem klassischen Gerät zu simulieren, sind neue Ansätze entstanden, um diese auf einem eingeschränkten Quantencomputer zu simulieren. Solche Techniken erlauben es uns, zu untersuchen, wie diese Einschränkungen die Rechenleistung des Geräts beeinflussen. Eine solche Technik nennt sich quasiprobability simulation (QPS), bei der das Ergebnis eines Quantenschaltkreises mithilfe eines Monte-Carlo-Verfahrens geschätzt wird, das Schaltelemente zufällig durch solche ersetzt, die auf dem eingeschränkten Quantencomputer ausführbar sind.

Der Hauptfokus dieser Arbeit liegt auf der QPS-basierten Simulation nichtlokaler Quantenrechnungen unter Verwendung lokaler Quantenoperationen. Auf der praktischen Seite ermöglicht dies die Simulation grosser Quantenschaltkreise mithilfe mehrerer kleinerer Quantencomputer – ein Verfahren, das manchmal als *circuit knitting* bezeichnet wird. Wir entdecken einen reichen mathematische Formalismus mit zahlreichen Verbindungen zur Ressourcentheorie der Verschränkung. Wir charakterisieren den optimalen Simulationsaufwand für eine breite Palette praktisch relevanter nichtlokaler Zustände und Kanäle und stellen explizite Protokolle vor, die diesen Simulationsaufwand erreichen. Darüber hinaus untersuchen wir den Nutzen klassischer Kommunikation zwischen den lokalen Parteien. Unsere Ergebnisse behandeln sowohl das single-shot als auch das asymptotische Regime.

Des Weiteren bietet diese Arbeit einen umfassenden Überblick über aktuelle Entwicklungen im Bereich der QPS. Neben der Simulation nichtlokaler Berechnungen untersuchen wir auch die Simulation von "magischen" Berechnungen mit einem Clifford-Computer, die Simulation rauschfreier Berechnungen mit einem fehlerbehafteten Gerät sowie die Simulation nichtphysikalischer Operationen. Wir rahmen QPS in einem ressourcentheoretischen Framework ein, das die Gemeinsamkeiten zwischen den verschiedenen Varianten dieser Technik hervorhebt. Ausserdem untersuchen wir die Bedeutung klassischer Zusatzinformationen im QPS-Verfahren und deren Einfluss auf den Simulationsaufwand und die Ausdruckskraft von QPS.

Insgesamt bietet diese Arbeit eine Einführung für Forschende, die sich für circuit knitting und QPS interessieren.

Contents

1	Intr	oduction	11
	1.1	A brief history of quasiprobability simulation	15
	1.2	Outline	
	1.3	Overview of contributions and published papers	16
2	Prel	iminaries	19
	2.1	Mathematical notation	19
	2.2	Elementary quantum gates	21
	2.3	Quantum channels and instruments	22
	2.4	Quantum resource theories	26
3	Qua	siprobability simulation framework	31
	3.1	Motivational example: non-Clifford simulation	31
	3.2	Quasiprobability simulation	33
	3.3	QPS with classical side information	38
	3.4	Basic properties of the quasiprobability extent	42
	3.5	The quasiprobability extent in quantum resource theories	51
	3.6	Convex programming for finite decomposition sets	59
	3.7	Further reading	61
	3.8	Contributions	62
4	Sim	ulating non-physical computation	65
	4.1	Utility of classical side information	66
	4.2	Characterization of quasiprobability extent	68
	4.3	Applications	71
	4.4	Further reading	79
	4.5	Contributions	80
5	Sim	ulating nonlocal computation	81
	5.1	Space-like cuts	86
		5.1.1 Decomposition sets	86
		5.1.2 Utility of classical side information	92
		5.1.3 Quasiprobability extent of states	94
		5.1.3.1 Pure states	
		5.1.3.2 Mixed states in the single-shot regime	101

			5.1.3.3 Mixed states in the asymptotic regime	. 105
		5.1.4	Quasiprobability extent of channels	. 114
			5.1.4.1 Clifford gates	
			5.1.4.2 Gates with unitary operator Schmidt decomposition .	
			5.1.4.3 State conversion bounds	
			5.1.4.4 PPTC lower bound	
			5.1.4.5 SDP hierarchy for γ_{SEPC}	
			5.1.4.6 Overview of one-shot bounds	
			5.1.4.7 Asymptotic characterization with entanglement cost .	
		5.1.5	Exploiting strict sub-multiplicativity in the black box setting .	
		5.1.6	Applications	
	5.2		like cuts	
		5.2.1	Decomposition sets	
		5.2.2	Utility of classical side information	
		5.2.3	Quasiprobability extent of the identity channel	
	г э	5.2.4	Going beyond the identity channel	
	5.3		er reading	
	5.4	Contri	ibutions	. 160
6	Simulating magic computation			163
	6.1		nposition sets	
	6.2	-	probability extent of states	
	6.3		probability extent of channels	
	6.4		er reading	
	6.5	Contri	ibutions	. 170
7	Sim	ulating	noise-free computation	171
	7.1		setup of probabilistic error cancellation	
	7.2		ing noise with randomized compilation	
	7.3		ining QEM and QEC	
	7.4		er reading	
	7.5	Contri	ibutions	. 180
8	Con	clusion	1	181
	8.1	Open	questions	. 184
Bi	bliog	raphy		187
A	Hoe	ffding	s inequality	205
В	Circ	uits for	r QPDs of unitaries	207
C	Cod	e switc	hing on Kitaev's surface code	209

Chapter 1

Introduction

Over the last few decades, the study of quantum information theory has produced remarkable insights into how the physical laws of quantum mechanics impact the fundamental limits of information storage, processing and transmission. The perhaps most prominent instance is the concept of quantum computing, which harnesses the quantum mechanical nature of quantum bits (qubits) to perform calculations in a fundamentally different way than a conventional classical computer. The construction of a large-scale fault-tolerant quantum computer is one of the greatest scientific and technological challenges of our time, with vast efforts across academia and industry currently dedicated to this goal.

It is a widely held belief that quantum computers can solve certain problems exponentially faster than their classical counterparts. Most famously, Peter Shor discovered a quantum algorithm in 1994 that efficiently computes the prime factorization of a number [146]. This task is believed to be so difficult for classical computers that many cryptographic protocols fundamentally rely on its hardness. Since Shor's discovery, many other tasks have been found with presumed polynomial or superpolynomial speedups.

To better understand the boundary between the computational power of classical and quantum computing, there are broadly two approaches that can be taken. On one hand, one can attempt to find evidence for the hardness of simulating certain classes of quantum algorithms. Efforts in this direction include results on boson sampling [1] and instantaneous quantum polynomial (IQP) circuits [27], [28]. On the converse side, one can identify classes of quantum circuits that can be efficiently simulated on classical computers. The prime example thereof is the famous Gottesman-Knill theorem, which states that any quantum circuit involving only Clifford operations can be efficiently classically simulated [2], [63]. This result can be extended to quantum circuits containing only a small (i.e. logarithmic in system size) number of non-Clifford gates [22], [120]. Other efficiently simulatable classes include, for example, circuits with low entanglement [165] and matchgate circuits [90], [155], [160]. These various insights provide us with an understanding of certain features that must be present in a quantum algorithm to be out of reach for classical computers.

In this thesis, we will consider an extension of this idea that takes it one step further. Instead of simulating a quantum computer with a classical computer, we will consider the task of simulating it with some notion of a "restricted" quantum computer. The utility of this line of work is twofold. On one hand, it provides us with a more fine-grained understanding of the separation of classical and quantum computers by studying intermediate computational models that lie between the two. On the other hand, these simulation algorithms can potentially be of practical interest from a technological point of view. Near-term quantum devices will likely suffer from severe limitations, such as high noise levels and a limited number of (logical) qubits, for the foreseeable future. Understanding how these restrictions affect their computational power is of significant interest. More concretely, these simulation algorithms can identify classes of quantum circuits that can already be simulated on near-term devices. Quantum circuits in this class, which also lie outside the reach of classical computers, constitute potential candidates for demonstrating a quantum advantage.

Our main workhorse in this thesis will be a technique which we refer to as *quasiprobability simulation* (QPS). Its roots lie in the simulation of near-Clifford quantum circuits with classical computers, but in recent years, it has been increasingly realized that it can also be extended to simulate quantum circuits with quantum computers. Consider a restricted quantum computer that can only realize some limited non-universal "instruction set" S, i.e., the circuits that our device can execute may only contain operations that lie in S. To simulate a circuit that contains a non-supported operation $\mathcal{E} \notin S$, we first need to find a so-called *quasiprobability decomposition* (QPD)

$$\mathcal{E} = \sum_{i=1}^{m} a_i \mathcal{F}_i \quad \text{for some} \quad a \in \mathbb{R}^m, \mathcal{F}_i \in S.$$
 (1.1)

Given this QPD, we can simulate \mathcal{E} by probabilistically replacing it with the operations \mathcal{F}_i and appropriately post-processing the final measurement outcome of the circuit, in a fashion reminiscent of quantum Monte Carlo algorithms. This allows for an unbiased estimation of the expectation value of an observable at the end of the circuit, albeit with an increased sampling overhead. The increased number of required shots is characterized by the 1-norm $\|a\|_1 = \sum_i |a_i|$ of the QPD. For a circuit with m instances of the unsupported gate \mathcal{E} , the number of circuit executions is increased by $\|a\|_1^{2m}$. Clearly, when performing QPS, one should use the best possible QPD that exhibits the smallest possible overhead. This smallest achievable 1-norm is such an important quantity, that we endow it with its own symbol $\gamma_S(\mathcal{E})$, and we call it the quasiprobability extent of \mathcal{E} w.r.t. S.

In this thesis, we will explore four different notions of a restricted quantum computer, which essentially amount to choosing different decomposition sets S. In each setting, one of the central objectives will be to characterize the corresponding quasiprobability extent γ_S .

Simulating nonlocal computation with local operations In this instance of QPS, we consider the simulation of operations \mathcal{E} , which act nonlocally across two parties, using only local operations (and possibly classical communication). This can be used to simulate nonlocal entangling gates across two isolated quantum

computers that cannot exchange any quantum information. Put differently, we will investigate the overhead for simulating large-scale quantum computation using smaller computers of bounded size. The sampling overhead is exponential in the amount of nonlocal interaction between the individual parties.

On a more practical note, this setting has potential technological implications. Near-term quantum computers are likely to suffer from a small number of available (logical) qubits. If a desired quantum circuit is almost local, we can simulate it using two or more smaller quantum devices in a process sometimes called *circuit knitting*. This can be practically useful, provided that the smaller local quantum computations are out of reach of classical simulation.

Simulating noise-free computation with a noisy quantum computer. The most significant obstacle in experimentally realizing large-scale robust quantum computers is noise, which leads to erroneous results in the computation. Due to decoherence from interactions with the environment, quantum information is inherently more prone to noise than its classical counterpart. The universally accepted approach to remedy effects of noise and decoherence in quantum computing is the use of *quantum error correction* (QEC) [32], [145], [149]. The celebrated threshold theorem guarantees that errors in a quantum computation can be suppressed efficiently to any level for arbitrarily long circuits, provided that the physical noise rate lies below some threshold value [3], [96], [144]. The low noise thresholds and the qubit overhead needed to implement error correction protocols are extremely challenging.

As an alternative approach to remove noise on an imperfect computer that doesn't fulfill the stringent requirements of QEC, one can utilize QPS to simulate ideal noise-free gates¹. In this case, the decomposition set *S* contains noisy operations that the device can physically execute. This is an instance of a *quantum error mitigation* technique, a family of recently proposed methods to remove the effect of noise on near-term devices in a hardware-friendly manner. QPS-based error mitigation has been a central component for some of the most impressive hardware demonstrations in recent times [94].

Simulating magic computation with a Clifford quantum computer In this setting, we consider the decomposition set *S* which only contains Clifford operations. Note that such a quantum computer is essentially equally powerful to a classical computer, due to the Gottesman-Knill theorem. Clifford operations are not universal and must be complemented with additional non-classical (also called "magic") states or channels to achieve universality.

Non-Clifford operations, such as the T gate, can be simulated using QPS. This results in a classical simulation algorithm for Clifford+T circuits with a runtime that is polynomial in the circuit size, but exponential in the number of T gates.

Simulating non-physical computation Finally, we will also consider QPS w.r.t. the "maximal" decomposition set *S* which contains all physically implementable

¹In fact, one could also view QEC itself as a simulation algorithm of noise-free computation using a noisy quantum computer.

(i.e. CPTP) maps. We will characterize the quasiprobability extent $\gamma_{\text{CPTP}}(\mathcal{E})$ of non-physical maps \mathcal{E} that may not be completely positive or trace-preserving. In a sense, this can be understood as simulating a computer that is even more powerful than a full-blown ideal quantum computer. Interestingly, this can have certain practical applications, which we will briefly explore.

The four settings above amount to choosing different decomposition sets S, and for each we will strive to evalute the quasiprobability extent for some target operation \mathcal{E} that lies outside S. This setup is very reminiscent of a *quantum resource theory* (QRT) which aims to characterize a certain quantum phenomenon by dividing the set of all quantum states and channels into "free" and "resourceful" ones. If a QRT captures some quantum phenomenon that is difficult to realize for a quantum computer, then it is very natural to pick S to be the free channels of that theory. In fact, the simulation of nonlocal and magic computation that we outlined above precisely correspond to the QRTs of entanglement and magic. Arguably, the simulation of non-physical operations can also be attributed to the trivial resource theory where all states and channels are free. In this thesis, we will explore this connection between QPS and QRTs. We will observe how certain techniques and properties involving the quasiprobability extent can be traced back to more general features of the underlying QRT.

One might naturally assume that the decomposition set *S* should always be chosen to be a subset of the physical quantum channels, i.e., only contain CPTP maps. However, this cannot capture the possibility of having intermediate measurements that produce classical side information which is then included in the QPS post-processing. For instance, it is possible to use completely positive trace-non-increasing maps as part of *S* by weighting a sample with zero if a certain measurement outcome occurs. More generally, it is even possible to simulate non-completely positive maps using negative weights. In much of the QPS literature, this trick is not utilized, or at least not to its full extent. In many instances, the implications of using classical side information were previously poorly understood. We aim to rectify this shortcoming in this thesis by introducing a general framework that captures the utilization of classical side information in its full generality. Particular emphasis is placed on understanding in which settings the side information provides an advantage in terms of the sampling overhead of QPS.

Another aspect that we will investigate in this thesis is the asymptotic cost of performing QPS. As we will see, the quasiprobability extent is sub-multiplicative $\gamma_S(\mathcal{E}\otimes\mathcal{F})\leq\gamma_S(\mathcal{E})\gamma_S(\mathcal{F})$ and in many instances even *strictly* sub-multiplicative. This has direct practical consequences, as simulating multiple parallel operations together can be strictly cheaper than simulating them individually. As such, we will be interested in the regularized extent $\overline{\gamma}_S(\mathcal{E}):=\lim_{n\to\infty}\gamma_S(\mathcal{E}^{\otimes n})^{1/n}$. Interestingly, an even lower asymptotic cost can be achieved by considering some *approximate* notion of QPS with a vanishing asymptotic error $\gamma_S^\infty(\mathcal{E}):=\lim_{n\to\infty}\lim_{n\to\infty}\inf_{\eta_S^\infty(\mathcal{E}^{\otimes n})^{1/n}$. Here, γ_S^ε is the *smoothed* quasiprobability extent which captures the minimal sampling overhead if we allow for an error of at most ε in the QPS. Evaluating $\overline{\gamma}_S$ is generally very difficult, and γ_S^∞ even more so. Nevertheless, we will remarkably find some explicit examples where $\overline{\gamma}_S$ is strictly larger than γ_S^∞ , meaning that approximate QPS is asymptotically more

efficient.

We briefly highlight that the largest focus of this thesis will be on the QPS of non-local operations. This theory exhibits a rich mathematical structure and we establish many connections to previously known results in the study of entanglement. We investigate the importance of classical communication between the smaller quantum devices for the task of circuit knitting and explicitly characterize the quasiprobability extent for a wide range of nonlocal states and channels, both in the single-shot and asymptotic regime.

1.1 A brief history of quasiprobability simulation

The notion of quasiprobability distribution has a long history in quantum physics. The Wigner function, introduced by Eugene Wigner in 1932, is a widely used representation of quantum states of continuous-variable systems. It can be understood as a distribution on the phase space of the system, though it can also exhibit negative values (hence the *quasi* in quasiprobability). The negativity in the Wigner function is strongly linked to non-classical properties, such as nonlocality [91]. As such, it is a natural question to ask which states are "classical" in the sense that they exhibit a non-negative Wigner function. This problem is solved by Hudson's theorem, which states that a pure state has a non-negative Wigner function if and only if it is a Gaussian state.

Following the rise of quantum information theory, considerable interest has been directed towards studying the finite-dimensional counterparts of the Wigner function. In particular, odd-dimensional qudit systems exhibit a well-behaved Wigner function which is non-negative for a pure state if and only if it is a stabilizer state [69]. Stabilizer states are thus endowed with some notion of being "classical", analogous to Gaussian states. This insight can be translated to the setting of computation, as it was soon realized that any odd-dimensional qudit quantum circuit can be efficiently classically simulated, as long as its elements (initial states, gates and measurements) exhibit non-negative Wigner functions [107], [161]. This can be considered a generalization of the Gottesman-Knill theorem [2], [63].

In a seminal work in Ref. [120], Pashayan, Wallman and Bartlett extended this result and demonstrated a classical simulation algorithm for quantum circuits where the runtime is governed by the amount of negativity in the Wigner function of the circuit elements². This can be considered the first instance of QPS, though it relies on the Wigner representation of the circuit elements (or more generally a representation w.r.t. to some fixed frame and dual frame), which has no well-behaved analog for qubits. Only later work by Howard and Campbell applied QPS in the qubit setting [86] by representing magic states as a QPD of stabilizer states. As such, they defined and investigated the quasiprobability extent of states for the first time. Shortly thereafter, Bennink *et al.* proposed a generalization to simulate non-Clifford qubit channel with stabilizer channels [14], providing the first qubit instance of QPS of channels (rather than states).

²This negativity is related to the magic monotone called *mana*, introduced in Ref. [162].

Until that point, all discussed references exclusively focused on *classical* simulation of quantum circuits. In Ref. [154], Temme, Bravyi and Gambetta came to the pivotal insight that the QPS technique can also be used to simulate quantum circuits using a quantum computer itself. More precisely, they considered using QPS to simulate the noise-free execution of a circuit with an imperfect noisy device. Since then, other such applications have been proposed and explored, such as the simulation of non-local computation [111], [129] and unphysical computation [88], [136]. They will be introduced in extended details throughout the thesis.

1.2 Outline

The preliminaries in Chapter 2 can be read as needed and are mostly intended to specify the notation and definitions used throughout the thesis. In Chapter 3, we formally introduce QPS and the general framework for including classical side information. We also study various properties of the quasiprobability extent, especially in the context of general QRTs. While Chapter 3 lays the crucial foundation for this thesis, the subsequent Chapters 4 to 7 can in principle be read independently and in any order, as they treat four separate applications of QPS. We summarize some important insights from this thesis as well as open questions in Chapter 8.

In the author's opinion, the most important contributions of this thesis lie in Chapter 5.

1.3 Overview of contributions and published papers

To provide a cohesive treatment of the subject, this thesis contains not solely contributions by the author, but also some results from other researchers. Every main chapter in this thesis includes an "overview of contribution" section which precisely outlines which parts are due to the author which are not. The current section serves as an additional brief overview of where the main contributions are located. We also highlight that this thesis contains a large amount of novel contributions which have not been published before, since they were either deemed not interesting enough to warrant writing a paper or because they will be included in an upcoming publication.

In Chapter 3, the main novelty lies in the establishment of a general framework to describe QPS with classical side information. We briefly summarize the contributions in the subsequent four chapters, in decreasing order of importance.

- Almost the entirety of Chapter 5 is novel work by the author, with a large part being previously unpublished.
- QPDs of nonphysical operations, which are treated in Chapter 4, were originally
 introduced by the author. However, the chapter also contains more advanced
 results which were subsequently found by other researchers.
- The general discussion of QEM in Chapter 7 summarizes various techniques in the field that were not introduced by the author. However, the Section 7.3 is dedicated to the author's publication in Ref. [130] discussing the interplay between QEM and QEC.

• Finally, Chapter 6 contains almost no contributions by the authors and is mostly included for completeness.

This thesis includes the content of following publications by the author.

- [129]: **C. Piveteau** and D. Sutter, *Circuit knitting with classical communication, IEEE Transactions on Information Theory*, vol. 70, no. 4, pp. 2734–2745, 2024. DOI: 10.1109/TIT.2023.3310797
- [29]: L. Brenner, **C. Piveteau**, and D. Sutter, *Optimal wire cutting with classical communication*, 2023. arXiv: 2302.03366 [quant-ph]. Available: https://arxiv.org/abs/2302.03366
- [140]: L. Schmitt, **C. Piveteau**, and D. Sutter, *Cutting circuits with multiple two-qubit unitaries*, 2024. arXiv: 2312.11638 [quant-ph]. Available: https://arxiv.org/abs/2312.11638
- [130]: C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, Error mitigation for universal gates on encoded qubits, Phys. Rev. Lett., vol. 127, p. 200505, 20 Nov. 2021. DOI: 10.1103/PhysRevLett.127.200505. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.200505

The following publication is also about QPS and its application to QEM.

• [131]: **C. Piveteau**, D. Sutter, and S. Woerner, *Quasiprobability decompositions with reduced sampling overhead, npj Quantum Information*, vol. 8, no. 1, Feb. 2022, ISSN: 2056-6387. DOI: 10.1038/s41534-022-00517-3. Available: http://dx.doi.org/10.1038/s41534-022-00517-3

As the main result is somewhat heuristic in nature and doesn't fit nicely with the rest of the thesis, we refrain from introducing it in detail. However, some of its technical results will be included where relevant.

Finally, we note three further publications by the author in the domain of quantum error correction. They are not covered in this thesis in any form.

- [128]: **C. Piveteau** and J. M. Renes, *Quantum message-passing algorithm for optimal and efficient decoding*, *Quantum*, vol. 6, p. 784, Aug. 2022, ISSN: 2521-327X. DOI: 10.22331/q-2022-08-23-784. Available: https://doi.org/10.22331/q-2022-08-23-784
- [125]: H. D. Pfister, **C. Piveteau**, J. M. Renes, and N. Rengaswamy, *Belief propagation for classical and quantum systems: Overview and recent results, IEEE BITS the Information Theory Magazine*, vol. 2, no. 3, pp. 20–32, 2022. DOI: 10.1109/MBITS. 2023.3285848
- [127]: C. Piveteau, C. T. Chubb, and J. M. Renes, Tensor-network decoding beyond 2d, PRX Quantum, vol. 5, p. 040303, 4 Oct. 2024. DOI: 10.1103/PRXQuantum. 5.040303. Available: https://link.aps.org/doi/10.1103/PRXQuantum.5.040303

Chapter 2

Preliminaries

In this chapter, we briefly summarize the notation and conventions that we use throughout this manuscript. Furthermore, we review some crucial prerequisites from basic quantum theory, quantum computing and quantum resource theory. An experienced reader may directly skip to Chapter 3.

2.1 Mathematical notation

Throughout this manuscript, we will always restrict ourselves to finite-dimensional quantum systems. A list of used mathematical symbols is depicted in Table 2.1.

		1
1-6	ene	rai
\sim	-11	LUI

$\mathbb{C}, \mathbb{R}, \mathbb{N}$	complex, real and natural numbers where N starts with 1
A, B, C, \dots	typical names for finite-dimensional quantum systems
$\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C, \dots$	Hilbert spaces corresponding to quantum systems A, B, C, \ldots
AB	joint quantum system corresponding to $\mathcal{H}_A \otimes \mathcal{H}_B$
$\dim(A)$	dimension of Hilbert space \mathcal{H}_A
$\operatorname{span}_{\mathbb{R}}(X)$, $\operatorname{span}_{\mathbb{C}}(X)$	span of set <i>X</i> over the real and complex number fields
sgn(x)	sign function $sgn(x) = 1$ for $x \ge 0$ and $sgn(x) = -1$ for $x < 0$
$\log(x)$	logarithm of <i>x</i> to the base 2
ln(x)	natural logarithm of x
Re[z], $Im[z]$	real and imaginary parts of complex number $z \in \mathbb{C}$
conv(X)	convex hull of the set <i>X</i>
aconv(X)	absolute convex hull of the set X
$\operatorname{diag}(\lambda)$	diagonal $n \times n$ matrix with entries $\lambda \in \mathbb{C}^n$

Operators

o p crattors	
L(A,B)	set of linear operators from <i>A</i> to <i>B</i>
H(A)	set of Hermitian operators on A
P(A)	set of positive semi-definite operators on A
U(A)	set of unitary operators on A

D(A)	set of density operators on A
ρ, σ, τ	typical names for density operators
$T(A \rightarrow B)$	set of linear superoperators $L(L(A), L(B))$
$\mathcal{E},\mathcal{F},\mathcal{G}$	typical names for superoperators
$HP(A \rightarrow B)$	set of Hermitian-preserving superoperators from <i>A</i> to <i>B</i>
$CP(A \rightarrow B)$	set of completely positive superoperators from <i>A</i> to <i>B</i>
$TP(A \rightarrow B)$	set of trace-preserving superoperators from <i>A</i> to <i>B</i>
$HPTP(A \rightarrow B)$	set of Hermitian-preserving trace-preserving superoperators
,	from A to B
$CPTP(A \rightarrow B)$	set of quantum channels from <i>A</i> to <i>B</i>
$CPTN(A \rightarrow B)$	set of completely positive trace-non-increasing maps from A
	to B
$QI(A \rightarrow B)$	set of quantum instruments from A to B
POVM(A)	set of POVMs on A
\mathcal{I},\mathcal{J}	typical names for quantum instruments
$SC(A, B \rightarrow C, D)$	set of superchannels from $CPTP(A \rightarrow B)$ to $CPTP(C \rightarrow D)$
$\mathbb{1}_A$	identity map from <i>A</i> to <i>A</i>
id_A	identity channel on A (map from $L(A)$ to $L(A)$)
$J_{\mathcal{E}_{\underline{\cdot}}}$	Choi operator of superoperator $\mathcal{E} \in T(A \to B)$
O^{\dagger}	conjugate transpose of operator O
O	absolute value $\sqrt{O^{\dagger}O}$ of operator O
$O_1 \otimes O_2$	tensor product of operators O_1 and O_2
$\operatorname{tr}[O]$	trace of operator O
$\operatorname{tr}_A[O]$	partial trace of operator <i>O</i> over subsystem <i>A</i>
T_A	transpose map on system A
Ad [O]	linear superoperator $X \mapsto OXO^{\dagger}$ induced by operator O
$\langle O_1, O_2 \rangle$	Hilbert-Schmidt inner product $\frac{1}{\dim(A)} \operatorname{tr}[O_1^{\dagger}O_2]$ for $O_i \in L(A)$
$O_1 \succcurlyeq O_2$	Loewner order on Hermitian operators $O_1, O_2 \in H(A)$

Norms

$ x _p$	p -norm $(\sum_i x_i ^p)^{1/p}$ of vector $x \in \mathbb{C}^n$
$ O _p$	Schatten <i>p</i> -norm of operator $O \in L(A, B)$
$\ O\ _{\infty}^{'}$	spectral norm of operator $O \in L(A, B)$
$\ \mathcal{E}\ _{\Diamond}$	diamond norm of superoperator $\mathcal{E} \in T(A \to B)$

Quantum resource theories

$\mathfrak{F}(A)$	set of free states on A
$\mathfrak{O}(A \to B)$	set of free channels from <i>A</i> to <i>B</i>
$\mathfrak{I}(A \to B)$	set of free quantum instruments from A to B
$\mathfrak{S}(A, B \to C, D)$	set of free superchannels from $CPTP(A \rightarrow B)$ to $CPTP(C \rightarrow D)$

Entanglement theory

$\overline{\mathrm{SEP}(A;B)}$	set of separable states across A and B
PPT(A; B)	set of states across A and B with positive partial transpose
$LO(A; B \rightarrow A'; B')$	set of local operations
$LOCC(A; B \rightarrow A'; B')$	set of local operations with classical communication
$SEPC(A; B \rightarrow A'; B')$	set of separable channels
$PPTC(A; B \rightarrow A'; B')$	set of PPT channels
$EBC(A \rightarrow A')$	set of entanglement breaking channels

Abbreviations

11001010110	
PPT	positive partial transpose
SDP	semidefinite program
QPS	quasiprobability simulation
QPD	quasiprobability decomposition
QRT	quantum resource theory
QEM	quantum error mitigation
QEC	quantum error correction

Table 2.1: Summary of symbols used throughout this thesis.

For the above sets of operators, whenever the input and output systems are identical, we only specify one system as the argument. For instance, the set of endomorphisms is L(A) := L(A,A) and similarly $CPTP(A) := CPTP(A \to A)$, $QI(A) := QI(A \to A)$ and so forth.

Whenever we write the composition $\mathcal{E}_2 \circ \mathcal{E}_1$ of two superoperators which do not have matching image and support spaces, the operations are implicitly meant to be tensored with the identity channel. For instance, if $\mathcal{E}_1 \in T(A \to BC)$ and $\mathcal{E}_2 \in T(BD \to E)$, then

$$\mathcal{E}_2 \circ \mathcal{E}_1 := (\mathcal{E}_2 \otimes \mathrm{id}_C)(\mathcal{E}_1 \otimes \mathrm{id}_D) \in \mathrm{T}(AD \to CE). \tag{2.1}$$

2.2 Elementary quantum gates

We assume basic familiarity with the quantum circuit formalism and refer the reader to [116] for a detailed introduction. Here, we briefly summarize the most important gates that we will encounter.

Each of the three Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{2.2}$$

induces a corresponding single-qubit gate. The *n*-fold tensor products of the Pauli matrices and the identity matrix generate the *n*-qubit *Pauli group*

$$P_n := \{i^c Q_1 \otimes Q_2 \otimes \cdots \otimes Q_n | c \in \{0, 1, 2, 3\}, Q_i \in \{1, X, Y, Z\}\}. \tag{2.3}$$

Note that the phase i^c is of no importance when considering an n-qubit Pauli operator as a gate. Based on the Pauli group, we can define the *Clifford hierarchy*, an increasing sequence of sets of n-qubit gates recursively defined as

$$CH_n^{(1)} := P_n, \quad CH_n^{(k)} := \{ U \in U(\mathbb{C}^{2^n}) | \forall Q \in P_n : UQU^{\dagger} \in CH_n^{(k-1)} \}.$$
 (2.4)

The second level of the hierarchy $CH_n^{(2)}$ is called the *n*-qubit *Clifford group* and plays a central role in the theory of quantum computation. Important single-qubit Clifford gates include the Hadamard and phase gates

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \tag{2.5}$$

and all two-qubit Clifford gates are equivalent to either the CNOT, SWAP or iSWAP gate

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad SWAP = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad iSWAP = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$
(2.6)

up to local Clifford unitaries. The Clifford unitaries are generated by the H, S and CNOT gates. The well-known Gottesman-Knill theorem [2], [63] states that there exists a polynomial-time classical simulation algorithm for circuits consisting only of Clifford gates as well as preparations and measurements of qubits in the computational basis. Unsurprisingly, the Clifford gates are not universal, i.e., it is not possible to approximate every unitary arbitrarily well using only Clifford gates. To reach universality, we need at least one gate from $CH_n^{(3)}$, as the third level does not exhibit a group structure as the first two levels do. Two typical examples of third-level gates that can each achieve universality when paired with the Clifford gates are the T gate and the Toffoli (controlled-controlled-not) gates

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} \quad \text{Toffoli} = |00\rangle\langle 00| \otimes \mathbb{1} + |01\rangle\langle 01| \otimes \mathbb{1} + |10\rangle\langle 10| \otimes \mathbb{1} + |11\rangle\langle 11| \otimes X.$$
(2.7)

2.3 Quantum channels and instruments

A state of a quantum system A can most generally be described by a density operator $\rho \in \mathrm{D}(A)$, i.e., a non-negative operator with unit trace $\mathrm{tr}[\rho] = 1$. One of the fundamental postulates of quantum mechanics states that the time evolution of an isolated system is described by $\rho \mapsto U\rho U^\dagger$ for some unitary operator $U \in \mathrm{U}(A)$. When the system can also interact with the environment, then the evolution can take a more general form of a quantum channel.

Definition 2.1: A superoperator $\mathcal{E} \in T(A \to B)$ is called *trace-preserving* if

$$\forall \omega \in L(A) : tr[\mathcal{E}(\omega)] = tr[\omega]$$
 (2.8)

and it is called *completely positive* if $\mathcal{E} \otimes \mathrm{id}_E \in \mathrm{T}(AE \to BE)$ is positive for any system E. We say \mathcal{E} is a *quantum channel* if it is both completely positive and trace-preserving.

The sets of completely positive superoperators, trace-preserving superoperators and quantum channels from A to B are denoted by $CP(A \rightarrow B)$, $TP(A \rightarrow B)$ and $CPTP(A \rightarrow B)$. Quantum channels precisely represent the most general class of superoperators which always map density operators to density operators, even when only applied to a subsystem.

Note that there exist superoperators which are positive, but not completely positive. A canonical example is the transpose operator $T_A \in T(A)$ which is defined as

$$T\left(\sum_{i,j}\alpha_{i,j}|b_i\rangle\langle b_j|\right) := \sum_{i,j}\alpha_{i,j}|b_j\rangle\langle b_i|$$
 (2.9)

with respect to some basis $\{|b_i\rangle\}_i$ of A.

Whenever the input Hilbert space is of unit dimension $\mathcal{H}_A = \mathbb{C}$, the quantum channels in $\operatorname{CPTP}(A \to B)$ can be understood as a deterministic preparation of some state on B. Therefore, we can identify the set of quantum channels $\operatorname{CPTP}(A \to B)$ with the set of quantum states $\operatorname{D}(B)$.

There exists a variety of representations of quantum channels. We summarize the ones that will be relevant throughout this manuscript.

Proposition 2.2: Choi-Jamiołkowski representation

The map from $T(A \rightarrow B)$ to L(AB) defined by

$$\mathcal{E} \mapsto J_{\mathcal{E}} := (\mathrm{id}_{A} \otimes \mathcal{E})(|\Psi\rangle\langle\Psi|_{AA'}) \tag{2.10}$$

is an isomorphism, where A' is a copy of A, $\mathcal E$ is understood to be acting on A' and $|\Psi\rangle:=\dim(A)^{-1/2}\sum_{i=1}^{\dim(A)}|i\rangle\otimes|i\rangle$ is the maximally entangled state w.r.t. some fixed orthonormal basis.

A major advantage of the Choi representation is that it directly captures some important properties of the superoperator \mathcal{E} . For instance,

$$\mathcal{E}$$
 is Hermitian-preserving $\iff J_{\mathcal{E}}$ is Hermitian, (2.11)

$$\mathcal{E}$$
 is completely positive $\iff J_{\mathcal{E}} \succcurlyeq 0$ (2.12)

and

$$\mathcal{E}$$
 is trace-preserving $\iff \operatorname{tr}_B[J_{\mathcal{E}}] = \frac{\mathbb{1}_A}{\dim(A)}$. (2.13)

Another important representation is provided by the Stinespring dilation theorem. This result states that any quantum channel can be considered to be a unitary evolution by sufficiently enlarging the space.

Proposition 2.3: Stinespring representation

Let $\mathcal{E} \in \operatorname{CPTP}(A \to B)$. Then there exists a system E, and an isometry $V \in \operatorname{L}(A,BE)$ such that

$$\forall \omega \in L(A) : \mathcal{E}(\omega) = \operatorname{tr}_{E}[V\omega V^{\dagger}]. \tag{2.14}$$

Next, we introduce the representation of completely positive linear operators based on Kraus operators.

Proposition 2.4: Operator-sum representation

Let $\mathcal{E} \in \operatorname{CP}(A \to B)$. Then there exists a finite family of operators $E_i \in \operatorname{L}(A,B)$, $i=1,\ldots,r$ such that

$$\forall \omega \in L(A) : \mathcal{E}(\omega) = \sum_{i=1}^{r} E_i \omega E_i^{\dagger}.$$
 (2.15)

The map \mathcal{E} is trace-preserving if and only if $\sum_{i=1}^{r} E_i^{\dagger} E_i = \mathbb{1}_A$.

Sometimes it can be useful to study completely positive maps that are trace-non-increasing, i.e., $\operatorname{tr}[\mathcal{E}(\rho)] \leq \operatorname{tr}[\rho]$. These are characterized by their Choi operator fulfilling $\operatorname{tr}_B[J_{\mathcal{E}}] \preccurlyeq \frac{\mathbb{1}_A}{\dim(A)}$, respectively by their Kraus operators fulfilling $\sum_{i=1}^r E_i^{\dagger} E_i \preccurlyeq \mathbb{1}_A$.

Finally, we mention one last representations that is applicable to n-qubit superoperators $\mathcal{E} \in \mathrm{T}(A)$ for $\mathcal{H}_A = (\mathbb{C}^2)^{\otimes n}$. The Pauli transfer matrix $M \in \mathbb{C}^{4^n \times 4^n}$ of \mathcal{E} is defined as

$$M_{ij} := \langle Q_i, \mathcal{E}(Q_j) \rangle = \frac{1}{2^n} \operatorname{tr} \left[Q_i \mathcal{E}(Q_j) \right]$$
 (2.16)

where $\{Q_1, Q_2, \dots, Q_{4^n}\}$ are the *n*-qubit Pauli strings $\{1, X, Y, Z\}^{\otimes n}$ in some fixed order. As such, the Pauli transfer matrix of a superoperator can be understood as its matrix representation in terms of the Pauli strings, which form an orthonormal basis of the space of $2^n \times 2^n$ matrices.

In quantum information theory we are often interested in measurements, i.e., processes which produce classical information from a quantum state. The most general measurement process can be modelled by a quantum channel $\mathcal{E}_{A \to BX} \in \operatorname{CPTP}(A \to BX)$ which maps any state $\rho_A \in \operatorname{D}(A)$ to a so-called *classical-quantum* state of the form

$$\sum_{i} p_{i} \sigma_{B}^{(i)} \otimes |i\rangle\langle i|_{X} \tag{2.17}$$

where p_i is a probability distribution $p_i \ge 0$, $\sum_i p_i = 1$, $|i\rangle$ is an orthonormal basis of X and $\sigma_B^{(i)} \in D(B)$. As such, a generalized measurement can always be written in the form

$$\mathcal{E}_{A \to BX}(\rho) = \sum_{i} p_i(\rho) \sigma^{(i)}(\rho) \otimes |i\rangle\langle i|$$
 (2.18)

where both $p_i(\rho)$ and $\sigma^{(i)}(\rho)$ are functions of ρ . If we define $\mathcal{E}_i(\rho) := p_i(\rho)\sigma^{(i)}(\rho)$ we can equivalently write

 $\mathcal{E}_{A \to BX}(\rho) = \sum_{i} \mathcal{E}_{i}(\rho) \otimes |i\rangle\langle i| . \qquad (2.19)$

Since $\mathcal{E} \in \operatorname{CP}(A \to BX)$, the maps \mathcal{E}_i must also be completely positive. Furthermore, $\sum_i \mathcal{E}_i = \operatorname{tr}_X \circ \mathcal{E}_{A \to BX}$ which is trace-preserving. Any generalized quantum measurement is fully characterized by a family $(\mathcal{E}_i)_i$ that fulfills these two properties.

Definition 2.5: A quantum instrument from system A to system B is a countable family $\mathcal{I} = (\mathcal{E}_i)_i$ of non-zero operators $\mathcal{E}_i \in \operatorname{CP}(A \to B)$ such that $\sum_i \mathcal{E}_i \in \operatorname{TP}(A \to B)$. We denote the set of such quantum instruments by $\operatorname{QI}(A \to B)$.

A quantum instrument is therefore a mathematical abstraction of a general quantum measurement process. The defined objects are sometimes also called *discrete* quantum instruments due to their countable nature. In this manuscript, we will never consider non-discrete instruments.

We say that a quantum instrument \mathcal{I}' is a coarse-graining of another instrument \mathcal{I} if it corresponds to the same physical measurement process, except that we merge some of the classical measurement outcomes (i.e., we consider them to be the same outcome). More mathematically, this is captured as follows.

Definition 2.6: Let $\mathcal{I} = (\mathcal{E}_i)_{i \in \Theta}$ and $\mathcal{I}' = (\mathcal{F}_j)_{j \in \Theta'}$ be quantum instruments from A to B. We say that \mathcal{I}' is a coarse-graining of \mathcal{I} if there exists a map $f: \Theta \to \Theta'$ such that

$$\mathcal{F}_j = \sum_{i:f(i)=j} \mathcal{E}_i. \tag{2.20}$$

For a quantum instrument $\mathcal{I} = (\mathcal{E}_i)_i \in \mathrm{QI}(A \to B)$, we define its associated marginal channel as

$$\operatorname{chan}[\mathcal{I}] := \sum_{i} \mathcal{E}_{i} \in \operatorname{CPTP}(A \to B). \tag{2.21}$$

This channel essentially corresponds to $\operatorname{tr}_X \circ \mathcal{E}_{A \to BX}$ from before and can also be understood as the "fully coarse-grained" instrument of \mathcal{I} . Similarly, if we consider a set $Q \subset \operatorname{QI}(A \to B)$ of quantum instruments, we define

$$\operatorname{chan}[Q] := \left\{ \operatorname{chan}[\mathcal{I}] \middle| \mathcal{I} \in Q \right\} \subset \operatorname{CPTP}(A \to B). \tag{2.22}$$

Oftentimes, one is purely interested in the distribution of the outcomes of a measurement process and not in the precise post-measurement state. Put differently, the only relevant quantity in this case is the quantum-to-classical channel

$$(\operatorname{tr}_{B} \circ \mathcal{E}_{A \to BX})(\rho) = \sum_{i} \operatorname{tr}[\mathcal{E}_{i}(\rho)] |i\rangle\langle i|$$
 (2.23)

$$= \sum_{i} \operatorname{tr}\left[\sum_{j} E_{j}^{(i)} \rho(E_{j}^{(i)})^{\dagger}\right] |i\rangle\langle i|$$
 (2.24)

$$= \sum_{i} \operatorname{tr}[M_{i}\rho] |i\rangle\langle i| \qquad (2.25)$$

where we inserted the operator-sum representation $\mathcal{E}_i(\rho) = \sum_j E_j^{(i)} \rho(E_j^{(i)})^{\dagger}$ and defined the positive operators $M_i := \sum_j (E_j^{(i)})^{\dagger} E_j^{(i)} \in \mathrm{P}(A)$ which fulfill $\sum_i M_i = \mathbb{1}_A$. So to compute the statistics of the measurement outcomes, is suffices to only specify the operators M_i .

Definition 2.7: A positive operator-valued measure (POVM) on the system A is a countable family $(M_i)_i$ of positive operators $M_i \in P(A)$ that fulfills $\sum_i M_i = \mathbb{1}_A$. We denote the set of POVMs on A by POVM(A).

2.4 Quantum resource theories

The framework of quantum resource theories (QRT) is used to study and characterize various phenomena in quantum mechanics, such as entanglement, coherence, magic and many more. The idea is to consider a restricted setting where operations can only be accomplished if they do not incur the quantum phenomena in question. As such, every QRT has a characteristic set of "free channels", which constitutes the CPTP maps which can be realized under the restrictions of the resource theory. Similarly, states are categorized into "free states" and "resource states". Some of the typical goals of a QRT include the quantification of the amount of resource contained in a state as well as the convertibility between states under free operations.

For a detailed introduction to the framework of QRTs, we refer the reader to Ref. [43], [65], [66]. Note that there are some technical variations in how resource theories are defined in literature. We will mostly follow the conventions from [65].

Mathematically speaking, a QRT is in essence an object that specifies which quantum channels between two systems *A* and *B* are considered "free".

Definition 2.8: A *quantum resource theory* is a mapping $\mathfrak O$ that takes two quantum systems A,B to a set of quantum channels $\mathfrak O(A\to B)\subset \operatorname{CPTP}(A\to B)$ that fulfills following condition:

- 1. *Doing nothing is free:* For any system A, $id_A \in \mathfrak{O}(A \to A)$.
- 2. Composition is free: For any $\mathcal{E} \in \mathfrak{O}(A \to B)$ and $\mathcal{F} \in \mathfrak{O}(B \to C)$ we also have $\mathcal{F} \circ \mathcal{E} \in \mathfrak{O}(A \to C)$.
- 3. *Discarding is free:* For any system A, $\operatorname{tr} \in \mathfrak{D}(A \to \mathbb{C})$.

Every QRT has an associated set of free states which do not contain any resource.

Definition 2.9: The free states of a QRT $\mathfrak D$ on a system A are $\mathfrak F(A) \coloneqq \mathfrak D(\mathbb C \to A)$.

The following two properties are direct consequences of properties 2 and 3 of a QRT.

Lemma 2.10 (Golden rule of quantum resource theories): Consider a QRT \mathfrak{D} . If $\mathcal{E} \in \mathfrak{D}(A \to B)$ and $\rho \in \mathfrak{F}(A)$, then $\mathcal{E}(\rho) \in \mathfrak{F}(B)$.

Lemma 2.11: Consider a QRT $\mathfrak D$ and let $\rho \in \mathfrak F(B)$. For any physical system A the map $\mathcal E \in \operatorname{CPTP}(A \to B)$ defined by $\mathcal E(\sigma) \coloneqq \operatorname{tr}[\sigma]\rho$ is in $\mathfrak D(A \to B)$.

Most QRTs studied in practice also fulfill additional properties when considered on multipartite systems.

Definition 2.12: We say a QRT \mathfrak{D} has a tensor product structure if

- 1. Completely free operations: For any $\mathcal{E} \in \mathfrak{O}(A \to B)$ and any system C, one has $\mathcal{E} \otimes \mathrm{id}_C \in \mathfrak{O}(AC \to BC)$.
- 2. Freedom of relabelling: Let $A = A_1 A_2 \dots A_n$ and $B = B_1 B_2 \dots B_n$ be n-fold copies of identical systems and let $P_{\pi}^A \in \text{CPTP}(A), P_{\pi^{-1}}^B \in \text{CPTP}(B)$ be permutation channels corresponding to a permutation π of n elements. Then $\forall \mathcal{E} \in \mathfrak{O}(A \to B)$ we have $P_{\pi^{-1}}^B \circ \mathcal{E} \circ P_{\pi}^A \in \mathfrak{O}(A \to B)$.

Every QRT naturally induces a set of free quantum instruments.

Definition 2.13: The set of free instruments $\Im(A \to B) \subset \operatorname{QI}(A \to B)$ associated to a QRT $\mathfrak O$ is defined to be all quantum instruments $(\mathcal E_i)_i$ which either consist of a single element in $\mathfrak O(A \to B)$ or for which the map

$$\rho_A \mapsto \sum_{i} \mathcal{E}_i(\rho_A)_B \otimes |i\rangle\langle i|_X \tag{2.26}$$

is a free channel in $\mathfrak{O}(A \to BX)$ for some system X with orthonormal basis $\{|i\rangle | i=1,2,\dots\}$.

A QRT with tensor product structure is fully characterized by its set of free instruments.

Lemma 2.14: For a QRT $\mathfrak D$ with tensor product structure, one has

$$\operatorname{chan}[\mathfrak{I}(A \to B)] = \mathfrak{O}(A \to B). \tag{2.27}$$

Recall the definition of chan in Equation (2.22).

Proof. For any $\mathcal{E} \in \mathfrak{D}(A \to B)$, the quantum instrument with a single component (\mathcal{E}) is an element of $\mathfrak{I}(A \to B)$. Thus, $\mathcal{E} = \operatorname{chan}[(\mathcal{E})] \in \operatorname{chan}[\mathfrak{I}(A \to B)]$. Conversely, let $\mathcal{I} = (\mathcal{E}_i)_i$ be a free instrument in $\mathfrak{I}(A \to B)$. There exists a free channel $\tilde{\mathcal{E}}(\rho) = \sum_i \mathcal{E}_i(\rho) \otimes |i\rangle\langle i|_X$, $\tilde{\mathcal{E}} \in \mathfrak{D}(A \to BX)$. Therefore, $\operatorname{chan}[\mathcal{I}] = \sum_i \mathcal{E}_i = (\operatorname{id}_B \otimes tr_X) \circ \tilde{\mathcal{E}} \in \mathfrak{D}(A \to B)$.

Consider a general quantum measurement \mathcal{E} of the form

$$\rho \mapsto \sum_{i} \mathcal{E}_{i}(\rho) \otimes |i\rangle\langle i|_{X}$$
 (2.28)

which happens to be a free channel in some QRT \mathfrak{O} , i.e., $\mathcal{E} \in \mathfrak{O}(A \to BX)$. The golden rule of QRTs implies that for any free state $\rho \in \mathfrak{F}(A)$ the output $\mathcal{E}(\rho)$ is also free. However, it doesn't rule out the possibility that a post-selected quantum state $\frac{1}{\operatorname{tr}[\mathcal{E}_i(\rho)]}\mathcal{E}_i(\rho)$ could possibly be non-free. Put differently, \mathcal{E} can in general create resource states in a probabilistic fashion.

It turns out that this is not possible for the vast majority of QRTs that are studied in literature, i.e., measurement maps that probabilistically produce resource states are never free in these theories. This insight is captured by following axiom.

Definition 2.15 (Axiom of free instruments): We say a QRT $\mathfrak O$ fulfills the *axiom of free instruments* if for all free states $\rho \in \mathfrak F(A)$ and for all free instruments $\mathcal I = (\mathcal E_i)_i \in \mathfrak I(A \to B)$ one has

$$\forall i: \mathcal{E}_i(\rho) / \operatorname{tr}[\mathcal{E}_i(\rho)] \in \mathfrak{F}(B). \tag{2.29}$$

Now we turn our attention to measuring the amount of resource contained in a state or channel. Functions that quantify a resource are typically called *resource measures* or *resource monotones*. The exact meanings of these two terms in literature can vary, and in the following we present the definitions that we will use throughout this manuscript.

Definition 2.16: A *resource measure* for a QRT $\mathfrak O$ is a map $M: \bigcup_{\mathcal H} D(\mathcal H) \to \mathbb R_{\geq 0}$ fulfilling the two properties:

- 1. Weak monotonicity: $\forall \mathcal{E} \in \mathfrak{O}(A \to B), \forall \rho \in D(A) : M(\mathcal{E}(\rho)) \leq M(\rho)$.
- 2. *Normalization:* M(1) = 0 where 1 is the only state of the Hilbert space \mathbb{C} .

Here, $\cup_{\mathcal{H}}$ denotes the union over all Hilbert spaces. Note that any resource measure M fulfills $\rho \in \mathfrak{F}(A) \Rightarrow M(\rho) = 0$ for any $\rho \in D(A)$.

Definition 2.17: A resource measure M for a QRT $\mathfrak O$ is called a *resource monotone* if it additionally fulfills

1. Strong monotonicity: For any $\rho \in D(A)$ and $(\mathcal{E}_i)_i \in \mathfrak{I}(A \to B)$

$$M(\rho) \ge \sum_{i} \operatorname{tr}[\mathcal{E}_{i}(\rho)] M\left(\frac{\mathcal{E}_{i}(\rho)}{\operatorname{tr}[\mathcal{E}_{i}(\rho)]}\right)$$
 (2.30)

2. *Convexity:* For any finite ensemble of states $(p_i, \rho_i)_i$ with $\rho_i \in D(A)$

$$M\left(\sum_{i} p_{i} \rho_{i}\right) \leq \sum_{i} p_{i} M(\rho_{i}).$$
 (2.31)

Beyond the above two definitions, we list some further properties that a desirable measure or monotone might fulfill.

- Faithfulness: For $\rho \in D(A)$: $\rho \in \mathfrak{F}(A) \Leftrightarrow M(\rho) = 0$
- Sub-additivity: $\forall \rho \in D(A), \sigma \in D(B) : M(\rho \otimes \sigma) \leq M(\rho) + M(\sigma)$.
- Additivity: $\forall \rho \in D(A), \sigma \in D(B) : M(\rho \otimes \sigma) = M(\rho) + M(\sigma).$

While (*static*) resource theories content themselves with the study of non-free states, the broader field of *dynamic resource theories* goes beyond that by also quantifying the resources of general non-free channels. In that case, we additionally require some notion of free transformations that map free channels back to free channels.

We call a *superchannel* Θ a linear map from $T(A \to B)$ to $T(C \to D)$ that maps quantum channels back to quantum channels, i.e., $\Theta(\operatorname{CPTP}(A \to B)) \subset \operatorname{CPTP}(C \to D)$. We denote the set of all such superchannels $SC(A, B \to C, D)$. It was shown in Ref. [41] that any superchannel can be written in the form

$$\Theta(\mathcal{E}_{A \to B}) = \mathcal{F}_{BE \to D}^{\text{post}} \circ (\mathcal{E}_{A \to B} \otimes \text{id}_{E}) \circ \mathcal{F}_{C \to AE}^{\text{pre}}$$
(2.32)

for some ancillary system E and pre- and post-processing maps $\mathcal{F}^{\text{pre}}_{C \to AE} \in \text{CPTP}(C \to AE)$ and $\mathcal{F}^{\text{post}}_{BE \to D} \in \text{CPTP}(BE \to D)$.

One possible choice of free superchannels would be all such maps Θ where the pre- and post-processing can be realized using free channels. In order to remain more general, we will instead choose to characterize the set of free superchannels in a more axiomatic fashion (following the approach in Ref. [66]), more akin to how we introduced the free channels of a QRT.

Definition 2.18: A *dynamical quantum resource theory* is a mapping \mathfrak{S} that takes four quantum systems A, B, C, D to a set of superchannels $\mathfrak{S}(A, B \to C, D) \subset SC(A, B \to C, D)$ and fulfills following conditions

- 1. \mathfrak{S} induces a QRT \mathfrak{O} with tensor product structure $\mathfrak{O}(A \to B) := \mathfrak{S}(\mathbb{C}, A \to \mathbb{C}, B)$.
- 2. $\mathbb{1}_{\mathsf{T}(A\to B)} \in \mathfrak{S}(A,B\to A,B)$ for all systems A,B.
- 3. If $\Theta_1 \in \mathfrak{S}(A, B \to C, D)$ and $\Theta_2 \in \mathfrak{S}(C, D \to E, F)$ then $\Theta_2 \circ \Theta_1 \in \mathfrak{S}(A, B \to E, F)$.
- 4. $\forall \Theta \in \mathfrak{S}(A, B \to C, D)$ and any systems $E, F: \Theta \otimes \mathbb{1}_{T(E \to F)} \in \mathfrak{S}(AE, BF \to CE, DF)$.

The golden rule can straightforwardly be translated to dynamical QRTs.

Lemma 2.19: Consider a dynamical QRT S. Then

$$\forall \mathcal{E} \in \mathfrak{D}(A \to B), \Theta \in \mathfrak{S}(A, B \to C, D) : \Theta(\mathcal{E}) \in \mathfrak{D}(C \to D). \tag{2.33}$$

Finally, we also introduce the analog notion of a resource measure for a dynamical QRT.

Definition 2.20: A *dynamical resource measure* for a dynamical QRT \mathfrak{S} is a map $M: \bigcup_{(\mathcal{H}_A, \mathcal{H}_B)} \text{CPTP}(A \to B) \to \mathbb{R}_{\geq 0}$ fulfilling the following two properties:

- 1. Monotonicity: $\forall \Theta \in \mathfrak{S}(A, B \to C, D), \forall \mathcal{E} \in \mathfrak{D}(A \to B) : M(\Theta(\mathcal{E})) \leq M(\mathcal{E}).$
- 2. *Normalization:* $M(id_A) = 0$ for all systems A.

Here, $\cup_{(\mathcal{H}_A,\mathcal{H}_B)}$ denotes the union over all pairs of two Hilbert spaces. Note that these two properties again imply that $M(\mathcal{E}) = 0$ for any free channel \mathcal{E} .

Chapter 3

Quasiprobability simulation framework

3.1 Motivational example: non-Clifford simulation

We now explore a first application of the quasiprobability simulation (QPS) technique as an illustrative example. Consider the following setup: suppose we have access to a quantum computer that is only capable of executing Clifford operations. Specifically, this means the computer can take the description of some quantum circuit containing only Clifford gates along with qubit preparations and measurements in the computational basis, execute the circuit, and generate samples from its output distribution. Such a computer is not universal, i.e., it cannot approximate any unitary with arbitrary precision. In fact, the celebrated Gottesman-Knill theorem states that any such circuit can be efficiently simulated classically [2], [63].

Now, the objective is to use this restricted quantum computer to simulate a general quantum circuit that may include both Clifford and non-Clifford gates. For simplicity, we will assume that the only non-Clifford operations are T gates which are enough to elevate the Clifford group to a universal gate set. We first illustrate the QPS technique on the simplest case which is a circuit containing only a single T gate, as depicted on the left of Figure 3.1. The state at the end of the *n*-qubit circuit is

$$\omega := (\operatorname{Ad} [U_2] \circ \operatorname{Ad} [T] \circ \operatorname{Ad} [U_1]) (|0\rangle\langle 0|^{\otimes n})$$
(3.1)

where U_1 and U_2 are the Clifford unitaries representing all Clifford gates that occur

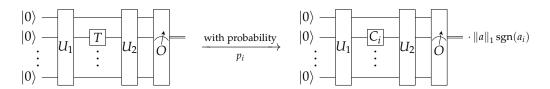


Figure 3.1: QPS of a Clifford circuit with one T gate. The T gate is probabilistically replaced with a Clifford gate C_i and the measurement outcome of the circuits is correspondingly multiplied by the factor $||a||_1 \operatorname{sgn}(a_i)$.

before and after the T gate. Recall that $\operatorname{Ad}[U](\rho) := U\rho U^{\dagger}$ is the channel induced by the unitary U. Let us assume that we are interested in estimating the expectation value $\operatorname{tr}[O\omega]$ of some observable O. If we know a decomposition of the T gate channel in terms of Clifford gates, i.e.,

$$Ad [T] = \sum_{i=1}^{m} a_i Ad [C_i]$$
(3.2)

for some coefficients $a \in \mathbb{R}^m$ and one-qubit Clifford gates C_i , then we can simply rewrite the expectation value as

$$\operatorname{tr}[O\omega] = \sum_{i=1}^{m} a_i \operatorname{tr}\left[O\left(\operatorname{Ad}\left[U_2\right] \circ \operatorname{Ad}\left[C_i\right] \circ \operatorname{Ad}\left[U_1\right]\right) \left(|0\rangle\langle 0|\right)^{\otimes n}\right]$$
(3.3)

$$= \sum_{i=1}^{m} p_i \operatorname{tr} \left[O\left(\operatorname{Ad}\left[U_2\right] \circ \operatorname{Ad}\left[C_i\right] \circ \operatorname{Ad}\left[U_1\right]\right) \left(|0\rangle\langle 0|\right)^{\otimes n} \right] \cdot \operatorname{sgn}(a_i) \|a\|_1 \qquad (3.4)$$

where $p_i := |a_i| / \|a\|_1$ is a probability distribution $p_i \ge 0$ and $\sum_i p_i = 1$. As depicted in Figure 3.1, this probabilistic mixture provides us with an explicit Monte Carlo method to evaluate $\operatorname{tr}[O\omega]$ while only having access to a restricted Clifford quantum computer. For every execution of the circuit, the T gate is randomly replaced by C_i according to the probability p_i and one keeps track of the sampled value of i. At the end of the circuit, the measurement outcome of the observable is weighted by $\|a\|_1 \operatorname{sgn}(a_i)$. Repeating this sampling procedure many times and then averaging the (weighted) measurement outcomes, Equation (3.4) tells us that this procedure will exactly retrieve $\operatorname{tr}[O\omega]$ in the limit of infinitely many shots.

Note that decompositions as in Equation (3.2) do exist, for example

$$Ad[T] = \frac{1}{2}id + \frac{1}{\sqrt{2}}Ad[S] - \frac{(\sqrt{2} - 1)}{2}Ad[Z].$$
 (3.5)

While the procedure outlined above allows us to retrieve the expectation value of the observable, it will typically require more shots to reach the same confidence compared to natively running the original circuit on a quantum computer that does support T gates. The reason for this is that the circuit output is scaled up by a factor of $\pm \|a\|_1$ (note that generally $\|a\|_1$ will be ≥ 1), so the variance increases by $\|a\|_1^2$. The number of shots required to reach a certain confidence thus also roughly increases with $\|a\|_1^2$. A rigorous statement on this sampling overhead will be made in Section 3.2.

To finish this exposition, let us consider how this technique can be applied to a circuit that contains more than one T gate. The main idea here is that the quasiprobability decompositions of the individual T gates can be combined into one large quasiprobability decomposition of the total circuit. For instance, if we have 3 T gates, then the channel describing the total circuit is¹

$$Ad [U_4] \circ Ad [T] \circ Ad [U_3] \circ Ad [T] \circ Ad [U_2] \circ Ad [T] \circ Ad [U_1]$$

$$(3.6)$$

¹Note that our notation here is ambiguous about which qubits the T gates act on. They can in principle act on different qubits.

where the U_i are Clifford unitaries. Inserting the decomposition from Equation (3.2), the channel of the total circuit can be decomposed into Clifford channels

$$\sum_{i_1,i_2,i_3} p_{i_1} p_{i_2} p_{i_3} \cdot \operatorname{Ad} \left[U_4 C_{i_3} U_3 C_{i_2} U_2 C_{i_1} U_1 \right] \cdot \operatorname{sgn}(a_1 a_2 a_3) \|a\|_1^3.$$
 (3.7)

The Monte Carlo sampling procedure in this case works as follows: For every instance of the T gate in the circuit, we separately and independently sample a random Clifford gate C_i to replace it. The sampled indices $(i_1, i_2 \text{ and } i_3 \text{ in the example})$ are stored and at the end of the circuit we weight the measurement outcome correspondingly by either $+ ||a||_1^3$ or $- ||a||_1^3$.

Notice how the 1-norms of the individual decompositions combine in a multiplicative fashion, so the 1-norm of the total circuit QPD is given by $||a||_1^t$ where t is the number of T gates. The sampling overhead of the technique thus scales as $\mathcal{O}(||a||_1^{2t})$.

Using the above QPS technique on top of a Gottesman-Knill simulator for Clifford circuits, one obtains a classical algorithm to simulate Clifford+T circuits with a runtime that remains efficient for a small (logarithmic) number of T gates. It is not too surprising that QPS entails an exponential sampling overhead — a polynomial overhead would have implied that quantum computers could be simulated efficiently with classical computers. We also note that there exist asymptotically faster (but still exponential) classical simulation algorithms for Clifford+T circuits. Though they usually utilize strong simulation of Clifford circuits and as such cannot be used to simulate Clifford+T circuits on a Clifford *quantum* computer. A more in-depth discussion will be presented in Chapter 6.

3.2 Quasiprobability simulation

Consider a quantum computation described by some quantum circuit C followed by the measurement of an observable $O \in H(A)$, which is assumed to be normalized $\|O\|_{\infty} \le 1$ without loss of generality. Let us denote by $\omega \in H(A)$ the state that C prepares. Our ultimate goal is to estimate the expectation value $\operatorname{tr}[O\omega]$.

Furthermore, assume that we have access to a restricted quantum computer that cannot directly execute the circuit C, because it contains operations that it cannot physically implement². Instead, assume that we know a certain family of circuits C_1, \ldots, C_m that the computer can execute, and let us denote by σ_i the state prepared by C_i . These circuits are carefully chosen such that they provide a quasiprobability decomposition (QPD) of the desired target state

$$\omega = \sum_{i=1}^{m} a_i \sigma_i \tag{3.8}$$

for some $a \in \mathbb{R}^m$. Algorithm 1 provides an unbiased estimate \hat{O} of $tr[O\omega]$ with an error that can be made arbitrarily small by increasing the number of circuit executions $n_s \in \mathbb{N}$ (also called number of *shots*).

 $^{^{2}}$ Note however, that the restricted quantum computer must be able to realize the measurement of the observable O.

Algorithm 1 Quasiprobability simulation algorithm

```
Input: descriptions of O and C_i, a \in \mathbb{R}^m, number of shots n_s

Output: estimate \hat{O} of \operatorname{tr}[O\omega]

r \leftarrow \operatorname{empty} list

repeat n_s times

i \leftarrow \operatorname{sample} random number in \{1, \ldots, m\} with probability p_i = |a_i| / \|a\|_1

o \leftarrow \operatorname{execute} circuit C_i and measure O

o' \leftarrow o \cdot \operatorname{sgn}(a_i) \|a\|_1

append o' to r

end

return \operatorname{mean}(r)
```

Proposition 3.1: QPS sampling overhead

Consider the setup described above for Algorithm 1. Let $\epsilon > 0$ and $\delta \in (0,1]$. If $n_s \ge 2\left(\frac{\|a\|_1}{\epsilon}\right)^2 \ln\left(\frac{2}{\delta}\right)$ then $|\hat{O} - \text{tr}[O\omega]| \le \epsilon$ with probability at least $1 - \delta$.

This is a direct consequence of Hoeffding's inequality (see Appendix A).

Proof. Denote by X_1, \ldots, X_{n_s} the i.i.d. random variables describing the weighted outcome of the n_s different shots. The output of the QPS algorithms is thus $\hat{O} = \frac{1}{n_s} \sum_j X_j$. First, notice that

$$\mathbb{E}[\hat{O}] = \mathbb{E}[X_j] = \sum_i p_i \operatorname{tr}[O\sigma_i] \operatorname{sgn}(a_i) \|a\|_1 = \operatorname{tr}[O\omega].$$
 (3.9)

Since the $|X_i|$ are bounded by $||a||_1$, Hoeffding's inequality (Theorem A.1) states

$$\mathbb{P}\left[\left|\hat{O} - \text{tr}[O\omega]\right| \ge \epsilon\right] \le 2\exp\left(-\frac{n_s \epsilon^2}{2\|a\|_1^2}\right). \tag{3.10}$$

The right-hand side is $\leq \delta$ if and only if

$$n_s \ge \frac{2 \|a\|_1^2}{\epsilon^2} \ln\left(\frac{2}{\delta}\right). \tag{3.11}$$

Remark 3.2 (Tightness of Proposition 3.1): Notice that Proposition 3.1 only provides us with an upper bound for the required number of shots to achieve a certain confidence on the result. Indeed, for some circuits and decompositions, one might get lucky and the required number of shots could be much smaller. Consider for instance the extreme example where the observable *O* is zero-clearly, a single execution of the circuit is sufficient to retrieve the desired expectation value with certainty.

Still, there is a notion in which the statement is tight, because there do exist instances for which the number of shots described in Proposition 3.1 is required, at least up to constant factors. More concretely, consider an observable O with eigenvalues ± 1 as well as a state ω such that $\mathrm{tr}[O\omega]=0$. In this case, the QPS estimator outputs one of the values $+\|a\|_1$ or $-\|a\|_1$ every shot, with probability 1/2 each. By Lemma A.2, the required number of shots to estimate the true expectation value, up to error ε and with probability at least $1-\delta$, must be at least $\frac{1}{4}\left(\frac{\|a\|_1}{\varepsilon}\right)^2\ln\left(\frac{2}{15\delta}\right)$.

We note that lower sampling overheads can sometimes be achieved when additional structure about the problem is given. For example, in Ref. [74] a quadratic improvement is demonstrated for the task of circuit knitting under the condition that *O* fulfills some sparsity condition.

To minimize the sampling overhead, one should use a QPD of the circuit C which exhibits the smallest possible 1-norm $\|a\|_1$. Finding this optimal QPD is very difficult and generally intractable. Instead, the standard approach consists of constructing the overall circuit QPD from decompositions of the individual components in the circuit³. We illustrated this in the example from Section 3.1 where QPDs of individual T gates were combined. To describe the technique more generally, let us write the circuit C as a sequence of n superoperators, which could e.g. represent gates or state preparations⁴

$$\omega = \mathcal{E}_n \circ \mathcal{E}_{n-1} \circ \cdots \circ \mathcal{E}_2 \circ \mathcal{E}_1. \tag{3.12}$$

Each circuit element \mathcal{E}_i is then decomposed into operations $\mathcal{E}_i = \sum_j a_j^{(i)} \mathcal{F}_j^{(i)}$ that our quantum computer can physically execute. This in turn provides us with a valid QPD of the total circuit

$$\omega = \sum_{j_{1}, j_{2}, \dots, j_{n}} a_{j_{1}}^{(1)} a_{j_{2}}^{(2)} \cdots a_{j_{n}}^{(n)} \left(\mathcal{F}_{j_{n}}^{(n)} \circ \mathcal{F}_{j_{n-1}}^{(n-1)} \circ \cdots \circ \mathcal{F}_{j_{2}}^{(2)} \circ \mathcal{F}_{j_{1}}^{(1)} \right)$$

$$= \sum_{j_{1}, j_{2}, \dots, j_{n}} p_{j_{1}}^{(1)} p_{j_{2}}^{(2)} \cdots p_{j_{n}}^{(n)} \left(\mathcal{F}_{j_{n}}^{(n)} \circ \mathcal{F}_{j_{n-1}}^{(n-1)} \circ \cdots \circ \mathcal{F}_{j_{2}}^{(2)} \circ \mathcal{F}_{j_{1}}^{(1)} \right)$$

$$\cdot \left\| a^{(1)} \right\|_{1} \left\| a^{(2)} \right\|_{1} \cdots \left\| a^{(n)} \right\|_{1} \operatorname{sgn} \left(\prod_{k} a_{j_{k}}^{(k)} \right)$$

$$(3.13)$$

where
$$p_j^{(i)} := |a_j^{(i)}| / ||a^{(i)}||_1$$
.

Note that Proposition 3.1 only makes a statement about the number of circuit executions in QPS. There are a few other considerations that also impact the overall runtime of the technique, such as the size of the circuits C_i and the complexity of sampling from the probability distribution. If we construct the circuit QPD from individual circuit elements, as previously described, then both of these issues can be

³Interestingly, we will encounter some instances where we will go beyond this standard approach later in the thesis.

⁴Recall that states can be seen as channels (superoperators) with trivial input space.

addressed. First, the complexity of the circuits C_i is then identical to the complexity of C (assuming that the operations $\mathcal{F}_j^{(i)}$ can be implemented with constant overhead). Furthermore, the probability distribution factorizes into separate distributions $(p_j^{(i)})_j$ for every circuit element, which can be independently (and hence efficiently) sampled. This justifies our focus on the number of circuit executions to measure the complexity of the QPS method.

As illustrated in Equation (3.13), the 1-norm of the QPD we constructed for our circuit is given by the product of the 1-norms of the decompositions $\mathcal{E}_i = \sum_j a_j^{(i)} \mathcal{F}_j^{(i)}$ of the individual circuit elements. It is therefore advantageous to use QPDs of these circuit elements that have the smallest possible 1-norm. Finding such optimal QPDs for small circuit elements is more tractable than for the entire circuit itself (albeit often still very difficult). Given a fixed set $S \subset \mathrm{HP}(A \to B)$ that represents the operations that our restricted computer can perform, we define the *quasiprobability* extent of $\mathcal{E} \in \mathrm{HP}(A \to B)$ to be the smallest achievable 1-norm.

Definition 3.3: For a decomposition set $S \subset HP(A \to B)$ and a target operation $\mathcal{E} \in HP(A \to B)$, the quasiprobability extent of \mathcal{E} w.r.t. S is defined as

$$\gamma_S(\mathcal{E}) := \inf \left\{ \sum_{i=1}^m |a_i| \,\middle|\, \mathcal{E} = \sum_{i=1}^m a_i \mathcal{F}_i \text{ where } m \ge 1, \mathcal{F}_i \in S \text{ and } a_i \in \mathbb{R} \right\}.$$
(3.15)

If no valid decomposition exists, γ_S is defined to be infinite.

Example 3.4. If we pick S to be all Clifford operations and $\mathcal{E} = \operatorname{Ad}[T]$ to be the T gate channel, then the QPD in Equation (3.5) happens to be optimal, i.e., $\gamma_S(\operatorname{Ad}[T]) = \sqrt{2}$. A proof of this fact will be presented in Chapter 6.

One aspect of Definition 3.3 that immediately stands out is that we allow for both \mathcal{E} and the elements of S to possibly be non-CPTP maps. Indeed, the quasiprobability simulation algorithm works with arbitrary linear superoperators and can be used to estimate expectation values of circuits C that contain nonphysical operations in them. We will encounter multiple settings throughout this thesis where simulating nonphysical target operations \mathcal{E} can be very useful. Regarding the nonphysicalness of the decomposition set S, we will elaborate in Section 3.3 that more general Hermitian-preserving maps can capture processes that include intermediate measurements and post-processing, such as completely positive trace-non-increasing maps.

Remark 3.5 (real versus complex coefficients): The QPS technique can straightforwardly be extended to also work with QPDs that include complex coefficients a_i instead of real ones. The main advantage of QPDs with complex coefficients is that they are more expressive: For a Hermitian-preserving decomposition set $S \subset \operatorname{HP}(A \to B)$, $\operatorname{span}_{\mathbb{R}}(S)$ is itself a subset of $\operatorname{HP}(A \to B)$. In comparison, $\operatorname{span}_{\mathbb{C}}(S)$ can generally contain all of $\operatorname{T}(A,B)$.

In this work, we refrain from discussing complex QPDs due to two observations:

1. There is no physically meaningful situation known to us where we would

be interested in a target operation $\mathcal{E} \in T(A \to B) \setminus HP(A \to B)$.

2. For a target operation $\mathcal{E} \in \mathrm{HP}(A \to B)$, the quasiprobability extent remains unaffected by the choice of allowing complex coefficients.

To see why the second point holds, consider a QPD $\mathcal{E} = \sum_j a_j \mathcal{F}_j$ for $\mathcal{E}, \mathcal{F}_j \in \operatorname{HP}(A \to B)$ and $a_j \in \mathbb{C}$. By considering the QPD in the Choi representation $J_{\mathcal{E}} = \sum_j \operatorname{Re}[a_j]J_{\mathcal{F}_j} + i\sum_j \operatorname{Im}[a_j]J_{\mathcal{F}_j}$, we see that the skew-Hermitian component $i\sum_j \operatorname{Im}[a_j]J_{\mathcal{F}_j}$ must be zero as $J_{\mathcal{E}}$ is Hermitian. Therefore, $\mathcal{E} = \sum_i \operatorname{Re}[a_i]\mathcal{F}_i$ is also a valid QPD with at most the same 1-norm.

Note that QPS can also be used to estimate the outcome probabilities of a circuit, since they can be phrased as the expectation value of some observable. For instance, if at the end of a circuit n qubits are measured in the computational basis yielding an outcome $x \in \{0,1\}^n$, then the probability p(x) of obtaining x is given by the expectation value of $O = |x\rangle\langle x|$. Unfortunately, when the distribution p(x) is very flat (i.e. all probabilities are exponentially small in n) this is only of little use, as QPS only efficiently provides estimates of expectation values up to an absolute error ϵ , which would need to be exponentially small in n to properly assess the structure of the distribution. This is in stark contrast to some classical simulation algorithms which can efficiently estimate probability amplitudes up to a multiplicative error, which can then be used to sample from the output distribution using the chain rule for conditional probabilities⁵

There are still cases where the output distribution (or at least important aspects thereof) can be fully captured with QPS. For instance, when the output distribution is sufficiently sparse, it becomes possible to accurately sample from it using QPS [119]. Similarly, when n is very small (asymptotically speaking, i.e. constant or logarithmic in the problem size), then estimates up to an absolute error are sufficient to retrieve the distribution. The extreme example thereof are decision problems, where n=1. Therefore, QPS is applicable to circuits corresponding to languages in the complexity class BQP. More practically speaking, many applications of quantum computing are anyway primarily concerned with evaluating expectation values. This is the case for variational algorithms or for estimating physical observables in the context of Hamiltonian simulation of physical systems. Furthermore, other quantum algorithms, like Shors's factoring algorithm and quantum phase estimation, can be decomposed in a series of decision problems [152, Appendix H].

In general, there are many notions of what it means to "simulate" a quantum circuit, and the differences between them are often subtle. We refer readers that are further interested in this matter to the comprehensive treatment in Ref. [119]. In the language of these authors, QPS constitutes a "poly-box". Under certain complexity theoretic assumptions, it can be proven that poly-boxes provide a strictly weaker simulation compared to " ϵ -simulators", which arguably constitute the operationally

⁵The idea here is to sample one bit at a time in a sequential fashion, where each bit is sampled according to its (binary) marginal that is conditioned on the values of the previous bits. This technique is sometimes called the "qubit-by-qubit" algorithm in the context of stabilizer rank simulators [21], [23].

most meaningful notion of simulation⁶.

We end this section by briefly highlighting a special instance where the input space of a circuit element $\mathcal{E} \in \mathrm{HP}(A \to B)$ is trivial $A = \mathbb{C}$. In this case, \mathcal{E} is a state preparation channel and can be directly identified with a state $\rho \in \mathrm{H}(B)$. Put differently, QPS cannot only be used to simulate certain gates in a circuit, but also to simulate the preparation of states that are used in the circuit. The quasiproability extent of a state $\rho \in \mathrm{H}(B)$ w.r.t. some decomposition set $S \subset \mathrm{H}(B)$ is analogously defined as

$$\gamma_S(\rho) := \inf \left\{ \sum_{i=1}^m |a_i| \middle| \rho = \sum_{i=1}^m a_i \tau_i \text{ where } m \ge 1, \tau_i \in S \text{ and } a_i \in \mathbb{R} \right\}.$$
(3.16)

Example 3.6. We briefly illustrate the concept of state QPDs with the Clifford+T setting from Section 3.1. It is a widely known trick that a T gate can be realized using a Clifford circuit that consumes a single-qubit magic state

$$|H\rangle := \frac{1}{\sqrt{2}} \left(|0\rangle + e^{i\pi/4} |1\rangle \right)$$
 (3.17)

as depicted in Figure 3.2. Using this gadget, the action of an arbitrary Clifford+T circuit C with m T gates can be rewritten as

$$\mathcal{C}(|0\rangle\langle 0|^{\otimes n}) = \mathcal{C}'(|0\rangle\langle 0|^{\otimes n} \otimes |H\rangle\langle H|^{m}) \tag{3.18}$$

for some appropriately chosen Clifford circuit \mathcal{C}' .

Using a QPD of the magic state

$$|H\rangle\langle H| = \frac{1}{2}|+\rangle\langle +|+\frac{1}{\sqrt{2}}|i+\rangle\langle i+|-\frac{(\sqrt{2}-1)}{2}|-\rangle\langle -|$$
 (3.19)

into stabilizer states (i.e. states which are reachable from a computational basis state using Clifford gates), we can write the original circuit \mathcal{C} as a quasiprobabilistic mixture of Clifford circuits and hence apply QPS.

In fact, this state-based approach for Clifford+T simulation is historically more commonly considered in literature compared to the channel-based approach [79], [86]. Note also that the QPD in Equation (3.19) has the same 1-norm of $\sqrt{2}$ as the channel-based one in Equation (3.5).

3.3 QPS with classical side information

The optimal simulation overhead for some target operation $\mathcal{E} \in T(A \to B)$ strongly depends on the decomposition set S with respect to which we consider its QPD. At first, one might expect that S always has to be chosen as some subset of CPTP($A \to B$), since these are precisely the physically realizable maps. In this section, we will

⁶It is statistically impossible to distinguish an agent that can execute a quantum circuit from another agent which only possesses an ϵ -simulator thereof.

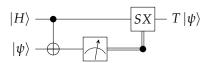


Figure 3.2: The magic state injection circuit realizes a T gate using a magic state and Clifford operations. The *X* and *S* gate on the upper qubit are only performed if the measurement outcome is 1.

see that it can be very useful to allow *S* to also contain certain trace-non-increasing maps and even certain non-completely positive maps. This should be understood as a mathematical method to capture the idea that intermediate measurements can produce classical side information during the circuit execution, as will be explained below. We will first consider two explanatory examples before explaining the trick in full generality.

As an example, suppose we want to simulate the one-qubit trace-non-increasing superoperator $\mathcal{E}=\mathrm{Ad}\left[|0\rangle\langle 0|\right]$ which maps $\rho\mapsto\langle 0|\rho|0\rangle\cdot|0\rangle\langle 0|$. Clearly, no valid QPD of $\mathrm{Ad}\left[|0\rangle\langle 0|\right]$ w.r.t. a decomposition set $S\subset\mathrm{CPTP}(A\to B)$ can exist, because any linear combination of trace-preserving maps is itself proportional to a trace-preserving map. Yet still, there is a way to simulate $\mathrm{Ad}\left[|0\rangle\langle 0|\right]$ using CPTP maps. To see this, notice that the action of $\mathrm{Ad}\left[|0\rangle\langle 0|\right]$ can be written as

$$\operatorname{Ad}\left[\left|0\right\rangle\left\langle 0\right|\right]\left(\rho\right) = \left(\mathbb{1} \otimes \left\langle 0\right|_{Y}\right) \cdot \mathcal{M}(\rho) \cdot \left(\mathbb{1} \otimes \left|0\right\rangle_{Y}\right). \tag{3.20}$$

where the CPTP map ${\mathcal M}$ denotes a measurement in the computational basis

$$\mathcal{M}(\rho) := \sum_{i=0,1} \operatorname{Ad} \left[|i\rangle\langle i| \right] (\rho) \otimes |i\rangle\langle i|_{X}$$
 (3.21)

that stores the outcome in a classical ancilla register X. Consider now that we want to estimate the expectation value of an observable O on Ad $[|0\rangle\langle 0|]$ (ρ_{in}) for some input state ρ_{in}

$$\operatorname{tr}\left[O \cdot \operatorname{Ad}\left[|0\rangle\langle 0|\right](\rho_{\mathrm{in}})\right] = \operatorname{tr}\left[\left(O \otimes |0\rangle\langle 0|_{X}\right) \cdot \mathcal{M}(\rho_{\mathrm{in}})\right]. \tag{3.22}$$

Notice how we replaced Ad $[|0\rangle\langle 0|]$ with \mathcal{M} by extending the observable to also act on the classical register X. Put differently, our simulation procedure operates as follows: replace Ad $[|0\rangle\langle 0|]$ by a computational basis measurement, and in case the measurement result corresponds to $|1\rangle\langle 1|$, then weight the final measurement outcome by zero. Note that this is not post-selection, as the "discarded" runs are still taken into account when averaging over multiple runs, they just have value zero.

It is illustrative to briefly consider a second example. This time, we want to simulate a target operation $\mathcal{E}=\operatorname{Ad}[|0\rangle\langle 0|]-\operatorname{Ad}[|1\rangle\langle 1|]$ which is not even completely positive. Here again, if we only care about the expectation value of some final observable O, we can simulate \mathcal{E} by replacing it with a computational basis measurement

$$\operatorname{tr}\left[O\cdot\left(\operatorname{Ad}\left[|0\rangle\langle 0|\right]-\operatorname{Ad}\left[|1\rangle\langle 1|\right]\right)(\rho_{\operatorname{in}})\right]=\operatorname{tr}\left[\left(O\otimes\left(|0\rangle\langle 0|_{X}-|1\rangle\langle 1|_{X}\right)\right)\cdot\mathcal{M}(\rho_{\operatorname{in}})\right]. \tag{3.23}$$

and by weighting the circuit outcome by either +1 or -1 depending on the intermediate measurement outcome.

The main insights of the previous two examples is that for the purpose of estimating some expectation value one can use classical side information provided by intermediate measurements to simulate non-CPTP maps by weighting the final observable measurement by some factor depending on the intermediate measurement outcome. The most general operation that produces classical information is described by a quantum instrument. Given a quantum instrument $(\mathcal{G}_i)_i$ that our computer can physically realize, this technique allows us to effectively perform any operation $\sum_i \beta_i \mathcal{G}_i$ for some weighting factors $\beta_i \in \mathbb{R}$.

For completeness, we state the modified quasiprobability simulation algorithm with classical side information. Consider a given target circuit C preparing a state $\omega \in H(A)$. Furthermore, consider that we have a family of circuits that can be executed on the restricted quantum computer C_1, \ldots, C_m where C_i prepares the classical-quantum state $\sigma_i = \sum_j \tau_j^{(i)} \otimes |j\rangle\langle j|$. These circuits shall form a valid QPD in the sense that $\omega = \sum_i a_i (\sum_j \beta_j^{(i)} \tau_j)$ for some $a \in \mathbb{R}^m$ and $\beta_j^{(i)} \in \mathbb{R}$. Algorithm 2 produces an estimate \hat{O} of $\operatorname{tr}[O\omega]$.

Algorithm 2 QPS algorithm with classical side information

```
Input: descriptions of O and C_i, a \in \mathbb{R}^m, \beta_j^{(i)} \in \mathbb{R}, number of shots n_s

Output: estimate \hat{O} of \text{tr}[O\omega]

r \leftarrow \text{empty list}

repeat n_s times

i \leftarrow \text{sample random number in } \{1, \ldots, m\} with probability p_i = |a_i| / \|a\|_1

execute Circuit C_i to prepare CQ state \sum_j \tau_j^{(i)} \otimes |j\rangle\langle j|_X

j \leftarrow \text{value of classical register } X

o \leftarrow \text{measurement outcome of observable } O \text{ on } \tau_j^{(i)}

o' \leftarrow o \cdot \text{sgn}(a_i) \|a\|_1 \beta_j^{(i)}

append o' to r

end

return \text{mean}(r)
```

It is easy to check that \hat{O} is an unbiased estimator for the true expectation value as

$$\operatorname{tr}[O \cdot \omega] = \sum_{i} a_{i} \operatorname{tr} \left[O \cdot \left(\sum_{j} \beta_{j}^{(i)} \tau_{j}^{(i)} \right) \right]$$

$$\sum_{i} \operatorname{tr} \left[\left(O \odot \sum_{j} \beta_{j}^{(i)} |I_{i}\rangle\langle I_{i}\rangle \right) \left(\sum_{j} \tau_{j}^{(i)} \odot |I_{i}\rangle\langle I_{i}\rangle \right) \right] \| \mathbf{a} \|_{L^{2}(\Omega)}$$
(3.24)

$$= \sum_{i} p_{i} \operatorname{tr} \left[\left(O \otimes \sum_{k} \beta_{k}^{(i)} |k\rangle\langle k| \right) \cdot \left(\sum_{j} \tau_{j}^{(i)} \otimes |j\rangle\langle j| \right) \right] \|a\|_{1} \operatorname{sgn}(a_{i}). \quad (3.25)$$

In order for Proposition 3.1 to still apply, we must also make sure that we don't scale up the outcome of the estimator to a magnitude larger than $||a||_1$. For this reason, we restrict ourselves coefficients of magnitude $|\beta_i^{(i)}| \le 1$.

In summary, given a quantum computer that can realize some set of quantum instruments *Q*, the associated decomposition set which captures the capabilities of that computer is given as follows.

Definition 3.7: Let $Q \subset QI(A \to B)$ be a set of quantum instruments. We define

$$\operatorname{chan}^*[Q] := \left\{ \sum_i \beta_i \mathcal{E}_i | (\mathcal{E}_i)_i \in Q, \beta_i \in [-1, 1] \right\}. \tag{3.26}$$

The notation is meant to mirror the definition in Equation (2.22).

The two introductory examples in this section have shown that classical side information can enable the simulation of circuits containing nonphysical superoperators that would be out of reach otherwise, i.e.,

$$\gamma_{\operatorname{chan}^*[\operatorname{QI}]}(\operatorname{Ad}[|0\rangle\langle 0|]) = \gamma_{\operatorname{chan}^*[\operatorname{QI}]}(\operatorname{Ad}[|0\rangle\langle 0|] - \operatorname{Ad}[|1\rangle\langle 1|]) \le 1 \tag{3.27}$$

whereas

$$\gamma_{\text{CPTP}}(\text{Ad}[|0\rangle\langle 0|]) = \gamma_{\text{CPTP}}(\text{Ad}[|0\rangle\langle 0|] - \text{Ad}[|1\rangle\langle 1|]) = \infty. \tag{3.28}$$

An arguably even more interesting question is, whether the utilization of classical side information can also help to lower the quasiprobability extent to simulate certain physical CPTP maps. This is one of the key questions that we will investigate in this thesis, and interestingly, the answer will be different depending on the considered setting.

We finish this chapter by providing a simple mathematical characterization of $chan^*[Q]$. For this, we introduce following notions for sets of quantum instruments.

Definition 3.8: For a set of quantum instruments $Q \subset QI(A \to B)$ we say that

- *Q* is *coarse-grainable* if $\forall \mathcal{I} \in Q$, any coarse-graining of \mathcal{I} is also in Q.
- Q is trivially fine-grainable if $\forall (\mathcal{E}_i)_i \in Q$ we have $((1-p_i)\mathcal{E}_i)_i \cup (p_i\mathcal{E}_i)_i \in Q$ for any choice of probabilities $p_i \in [0,1]$.
- Q is *closed under mixture* if for any two instruments $(\mathcal{E}_i)_i$ and $(\mathcal{F}_j)_j$ in Q, their mixture $(p\mathcal{E}_i)_i \cup ((1-p)\mathcal{F}_i)_j$ is also in Q, for any $p \in [0,1]$.

These properties are extremely natural when describing the set of quantum instruments Q that a hypothetical quantum computer can implement, and all three properties should hold in most physically relevant settings. Indeed, Q being coarse-grainable simply means that one can discard classical information and treat different classical outcomes as the same. Trivial fine-grainability essentially means that one can generate a random bit with the probability depending on the observed measurement outcome. Similarly, when Q is closed under mixture, then randomly choosing between two quantum instruments is itself also a quantum instrument in Q.

At first glance, the trick of using classical side information might seem to add a lot of complexity to the process of finding QPDs, as elements in the decomposition set can now consist of a possibly infinite sum of quantum instrument maps. Fortunately,

for a sufficiently nice underlying set of quantum instruments Q, it turns out that we can restrict ourselves to only considering the two-element instruments in Q and coefficients $\beta_i = \pm 1$.

Lemma 3.9: If $Q \subset QI(A \to B)$ is coarse-grainable and trivially fine-grainable, then

$$chan^*[Q] = \{ \mathcal{E}^+ - \mathcal{E}^- | (\mathcal{E}^+, \mathcal{E}^-) \in Q \}.$$
 (3.29)

Proof. Consider an element $\mathcal{E} = \sum_i \beta_i \mathcal{E}_i \in \text{chan}^*[Q]$ where $(\mathcal{E}_i)_i \in Q$, $|\beta_i| \leq 1$. By the trivially fine-grainable property of Q, we can also express \mathcal{E} as a weighted sum of another instrument in Q using only the coefficients +1, -1 and 0:

$$\mathcal{E} = \sum_{i} \operatorname{sgn}(\beta_i) \left(|\beta_i| \, \mathcal{E}_i \right) + \sum_{i} 0 \cdot \left((1 - |\beta_i|) \mathcal{E}_i \right) \tag{3.30}$$

We now coarse-grain this instrument into a three-element instrument $(\mathcal{E}^+,\mathcal{E}^-,\mathcal{E}^0)$

$$\mathcal{E}^{+} := \sum_{i:\beta_{i} > 0} |\beta_{i}| \,\mathcal{E}_{i} \tag{3.31}$$

$$\mathcal{E}^{-} := \sum_{i:\beta_{i}<0} |\beta_{i}| \,\mathcal{E}_{i} \tag{3.32}$$

$$\mathcal{E}^0 := \sum_{i} (1 - |\beta_i|) \mathcal{E}_i \tag{3.33}$$

which is still in Q by the coarse-grainability property. We thus have

$$\mathcal{E} = \mathcal{E}^+ - \mathcal{E}^- + 0 \cdot \mathcal{E}^0. \tag{3.34}$$

Finally, we use the fine-grainability and coarse-grainability properties once more to argue that $(\mathcal{E}^+,\mathcal{E}^-,\frac{1}{2}\mathcal{E}^0,\frac{1}{2}\mathcal{E}^0)$ and thus $(\mathcal{E}^++\frac{1}{2}\mathcal{E}^0,\mathcal{E}^-+\frac{1}{2}\mathcal{E}^0)$ are quantum instruments in Q. The desired statement thus follows from

$$\mathcal{E} = (\mathcal{E}^+ + \frac{1}{2}\mathcal{E}^0) - (\mathcal{E}^- + \frac{1}{2}\mathcal{E}^0). \tag{3.35}$$

3.4 Basic properties of the quasiprobability extent

In this section, we explore some central mathematical properties of the quasiprobability extent. They will be repeatedly used in later chapters, so we collect them here for convenient reference. The first one is straightforward, but provides an important tool for finding lower and upper bounds.

Lemma 3.10: If
$$S \subset S' \subset HP(A \to B)$$
 then $\forall \mathcal{E} \in HP(A \to B) : \gamma_S(\mathcal{E}) \geq \gamma_{S'}(\mathcal{E})$.

Proof. This follows directly from the definition of γ , as any QPD of \mathcal{E} w.r.t. S is also a valid QPD w.r.t. S'.

Convex and absolutely convex decomposition sets In most practical instances, the decomposition set S is convex. Clearly, convexity is a very physical property: if our computer has the capabilities to physically realize the operations \mathcal{F}_1 and \mathcal{F}_2 , then it will likely be able to also realize $(1-p)\mathcal{F}_1+p\mathcal{F}_2$ for some $p\in[0,1]$ by randomly choosing one of the two with the according probability. Convex decomposition sets have the useful property that we can without loss of generality only consider QPDs with at most two elements.

Lemma 3.11: For a convex decomposition set $S \subset HP(A \to B)$ and $\mathcal{E} \in HP(A \to B)$ one has

$$\gamma_S(\mathcal{E}) = \inf \left\{ a^+ + a^- \middle| \mathcal{E} = a^+ \mathcal{F}^+ - a^- \mathcal{F}^- \text{ where } \mathcal{F}^\pm \in S \text{ and } a^\pm \ge 0 \right\}.$$
 (3.36)

Proof. Any QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ with $\mathcal{F}_i \in S$ induces a two-element QPD with the same 1-norm by grouping together all positive and negative terms

$$\mathcal{E} = a^+ \mathcal{F}^+ - a^- \mathcal{F}^- \tag{3.37}$$

where

$$a^{+} := \sum_{i:a_{i} \ge 0} a_{i}, \qquad a^{-} := -\sum_{i:a_{i} < 0} a_{i}$$
 (3.38)

and

$$\mathcal{F}^{+} := \frac{\sum_{i:a_{i} \geq 0} a_{i} \mathcal{F}_{i}}{a^{+}}, \qquad \mathcal{F}^{-} := -\frac{\sum_{i:a_{i} < 0} a_{i} \mathcal{F}_{i}}{a^{-}}, \tag{3.39}$$

where $\mathcal{F}^{\pm} \in S$ due to convexity and $a^+ + a^- = \sum_i |a_i|$.

In Section 3.3 we saw that it can be meaningful to think of the set of quantum instruments Q that a computer can realize, in order to capture the notion of classical side information. When Q is closed under mixture, then the induced decomposition set has an even stronger property than convexity.

Definition 3.12: A subset *X* of a real vector space *V* is called *absolutely convex* if it is convex and balanced (i.e. $\forall x \in X, \forall \lambda \in [-1,1] : \lambda x \in X$).

It is an easy exercise to verify that absolute convexity is equivalent to being closed under absolutely convex mixtures: $\sum \lambda_i x_i$ for $x_i \in X$ and $\sum_i |\lambda_i| \le 1$.

Lemma 3.13: If $Q \subset QI(A \to B)$ is closed under mixture, then chan[Q] is convex and chan*[Q] is absolutely convex.

Proof. We start with the convexity of chan[Q]. For any two $\mathcal{E} = \text{chan}[\mathcal{I}]$, $\mathcal{F} = \text{chan}[\mathcal{I}']$ with $\mathcal{I} = (\mathcal{E}_i)_i$, $\mathcal{I}' = (\mathcal{F}_j)_j$, we also have $(1-p)\mathcal{E} + p\mathcal{F} = \text{chan}[\mathcal{I}]$ where $\mathcal{J} := ((1-p)\mathcal{E}_i)_i \cup (p\mathcal{F}_j)_j$ is also in Q. The convexity of chan*[Q] follows from a similar argument. It remains to show that chan*[Q] is balanced:

$$\lambda \cdot \sum_{i} \beta_{i} \mathcal{E}_{i} = \sum_{i} (\lambda \beta_{i}) \mathcal{E}_{i} \in \text{chan}^{*}[Q]$$
(3.40)

for
$$|\lambda| \leq 1$$
, $|\beta_i| \leq 1$ and $(\mathcal{E}_i)_i \in Q$.

When a decomposition set is absolutely convex, we can even assume without loss of generality that our QPDs only have a single term in them.

Lemma 3.14: For an absolutely convex decomposition set $S \subset HP(A \to B)$ and $\mathcal{E} \in HP(A \to B)$

$$\gamma_S(\mathcal{E}) = \inf\{\kappa \ge 0 : \mathcal{E} = \kappa \mathcal{F} \text{ for some } \mathcal{F} \in S\}.$$
 (3.41)

Proof. Similar to the proof in Lemma 3.11, we only need to show that any QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ with $\mathcal{F}_i \in S$ induces another QPD with only one element and the same 1-norm. Define $\kappa := \sum_i |a_i|$ and $\mathcal{F} := \sum_i \frac{a_i}{\kappa} \mathcal{F}_i$. Clearly, $\mathcal{E} = \kappa \mathcal{F}$ and $\mathcal{F} \in S$ by the absolute convexity.

Besides the previous physical motivation, there is another reason why in many instances one can assume without loss of generality that the decomposition set *S* is either convex or absolutely convex. If we only care about computing the quasiprobability extent, we can replace *S* by its convex hull or its absolute convex hull.

Lemma 3.15: For any $S \subset HP(A \to B)$ and $\mathcal{E} \in HP(A \to B)$ one has

$$\gamma_S(\mathcal{E}) = \gamma_{\text{conv}(S)}(\mathcal{E}) = \gamma_{\text{aconv}(S)}(\mathcal{E})$$
 (3.42)

where the *convex hull* and *absolute convex hull* of *S* are defined as

$$\operatorname{conv}(S) := \left\{ \sum_{i=1}^{n} \lambda_i x_i \middle| n \in \mathbb{N}, x_i \in S, \lambda_i \in [0, 1], \sum_{i=1}^{m} \lambda_i = 1 \right\}$$
(3.43)

$$\operatorname{aconv}(S) := \left\{ \sum_{i=1}^{n} \lambda_i x_i \middle| n \in \mathbb{N}, x_i \in S, \lambda_i \in [-1, 1], \sum_{i=1}^{m} |\lambda_i| \le 1 \right\}. \tag{3.44}$$

Note that this statement does not rule out the possibility that classical side information can reduce the quasiprobability extent, as aconv(chan[Q]) is generally a strict subset of $chan^*[Q]$ for some $Q \subset QI(A \to B)$.

Proof. By Lemma 3.10 we know $\gamma_S(\mathcal{E}) \geq \gamma_{\operatorname{conv}(S)}(\mathcal{E}) \geq \gamma_{\operatorname{aconv}(S)}(\mathcal{E})$ so it suffices to show that $\gamma_{\operatorname{aconv}(S)}(\mathcal{E}) \geq \gamma_S(\mathcal{E})$. We argue this by showing that any QPD of \mathcal{E} w.r.t. aconv(S) induces a QPD of \mathcal{E} w.r.t. S with at most the same 1-norm of its coefficients. Let $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ be such a QPD with $\mathcal{F}_i = \sum_j \lambda_j^{(i)} \mathcal{G}_j^{(i)}$ where $\mathcal{G}_j^{(i)} \in S$. We thus have

$$\mathcal{E} = \sum_{i,j} (a_i \lambda_j^{(i)}) \mathcal{G}_j^{(i)}$$
(3.45)

which is a valid QPD with 1-norm

$$\sum_{i,j} \left| a_i \lambda_j^{(i)} \right| = \sum_{i} |a_i| \sum_{i} \left| \lambda_j^{(i)} \right| \le \sum_{i} |a_i|. \tag{3.46}$$

Relation to robustness Motivated by a class of resource measures studied in many QRTs, we define the *robustness* (sometimes also called *total robustness* to differentiate it from related measures) of a given operation \mathcal{E} w.r.t. a decomposition set S as follows.

Definition 3.16: The robustness of $\mathcal{E} \in \mathrm{HP}(A \to B)$ w.r.t. $S \subset \mathrm{HP}(A \to B)$ is defined as

$$R_S(\mathcal{E}) := \inf\{t \ge 0 | \exists \mathcal{F} \in S \text{ s.t. } \frac{\mathcal{E} + t\mathcal{F}}{\operatorname{tr}[J_{\mathcal{E}}] + t} \in S\}.$$
 (3.47)

Intuitively, the robustness quantifies the distance of \mathcal{E} to S by measuring how much of a "free operation" (where we call operations from S free) can be mixed with \mathcal{E} before it becomes itself "free". For convex decomposition sets consisting of quantum channels, and more generally trace-preserving maps, the robustness is directly related to the quasiprobability extent.

Lemma 3.17: Consider $S \subset HPTP(A, B)$ convex and $\mathcal{E} \in HP(A \to B)$. Then

$$\gamma_S(\mathcal{E}) = \operatorname{tr}[J_{\mathcal{E}}] + 2R_S(\mathcal{E}) \tag{3.48}$$

Note that in this setting, $\gamma_S(\mathcal{E}) = R_S(\mathcal{E}) = \infty$ if \mathcal{E} is not itself proportional to a trace-preserving map.

Proof. We prove this by showing that any valid two-element (see Lemma 3.11) QPD $(a^{\pm}, \mathcal{F}^{\pm})$ induces a valid pair (t, \mathcal{F}) with $tr[J_{\mathcal{E}}] + 2t = a^{+} + a^{-}$ and vice versa.

First, let t, \mathcal{F} be valid. They thus fulfill $(\operatorname{tr}[J_{\mathcal{E}}] + t)\mathcal{G} = \mathcal{E} + t\mathcal{F}$ for some $\mathcal{G} \in S$. Therefore, we get a valid QPD $\mathcal{E} = (\operatorname{tr}[J_{\mathcal{E}}] + t)\mathcal{G} - t\mathcal{F}$.

Conversely, let us assume $\mathcal{E}=a^+\mathcal{F}^+-a^-\mathcal{F}^-$ is a valid QPD. Applying the trace on the Choi representation of both sides of this equation yields $\mathrm{tr}[J_{\mathcal{E}}]=a^+-a^-$ or $a^+=\mathrm{tr}[J_{\mathcal{E}}]+a^-$. Therefore, $\frac{\mathcal{E}+a^-\mathcal{F}^-}{\mathrm{tr}[J_{\mathcal{E}}]+a^-}=\mathcal{F}^+\in S$.

Notice that Lemma 3.17 doesn't apply for (non-trivial) absolutely convex decomposition sets *S*, as they always contain non-trace-preserving maps. Fortunately, the relation to the robustness can still be established in physically relevant settings.

Lemma 3.18: Let $Q \subset QI(A \to B)$ be closed under mixture and $\mathcal{E} \in HP(A \to B)$. Then

$$\gamma_{\operatorname{chan}^*[Q]}(\mathcal{E}) = \operatorname{tr}[J_{\mathcal{E}}] + 2R_{\operatorname{chan}^*[Q]}(\mathcal{E}). \tag{3.49}$$

Proof. First, we note that for any $\mathcal{G} = \sum_j \beta_j \mathcal{G}_j \in \text{chan}^*[Q]$ we have

$$||J_{\mathcal{G}}||_{1} = \left\|\sum_{j} \beta_{j} J_{\mathcal{G}_{j}}\right\|_{1} \le \left\|\sum_{j} |\beta_{j}| J_{\mathcal{G}_{j}}\right\|_{1} = \operatorname{tr}\left[\sum_{j} |\beta_{j}| J_{\mathcal{G}_{j}}\right] \le \operatorname{tr}\left[\sum_{j} J_{\mathcal{G}_{j}}\right] = 1 \quad (3.50)$$

where we used that $\|X - Y\|_1 \le \|X\|_1 + \|Y\|_1 = \|X + Y\|_1$ for two positive operators X, Y. Therefore, for any valid QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ with $\mathcal{F}_i \in \text{chan}^*[Q]$ one has

$$||a||_1 \ge \sum_i |a_i| ||J_{\mathcal{F}_i}||_1$$
 (3.51)

$$= \sum_{i}^{r} \|a_{i} J_{\mathcal{F}_{i}}\|_{1} \tag{3.52}$$

$$\geq \left\| \sum_{i} a_{i} J_{\mathcal{F}_{i}} \right\|_{1} \tag{3.53}$$

$$= \|J_{\mathcal{E}}\|_1 \tag{3.54}$$

$$\geq \operatorname{tr}[J_{\mathcal{E}}] \tag{3.55}$$

and hence $\gamma_{\text{chan}^*[Q]}(\mathcal{E}) \geq \text{tr}[J_{\mathcal{E}}]$. We can thus write $\gamma_{\text{chan}^*[Q]}(\mathcal{E}) = \text{tr}[J_{\mathcal{E}}] + 2r$ for some $r \geq 0$. It remains to show that this r precisely corresponds to $R_{\text{chan}^*[Q]}(\mathcal{E})$.

The proof proceeds analogously to the one of Lemma 3.17. On one side, consider an arbitrary valid QPD $\mathcal{E} = \kappa \mathcal{F}$ with $\mathcal{F} \in \text{chan}^*[Q]$ (recall Lemma 3.14). We can write $\kappa = \text{tr}[J_{\mathcal{E}}] + 2r$ for some $r \geq 0$. Hence, $\frac{\mathcal{E} + r(-\mathcal{F})}{\text{tr}[J_{\mathcal{E}}] + r} = \mathcal{F} \in \text{chan}^*[Q]$ where $(-\mathcal{F})$ is also in chan*[Q] due to the absolute convexity (see Lemma 3.13). This means that $(r, (-\mathcal{F}))$ is a feasible point for the robustness.

Conversely, let t, \mathcal{F} be an arbitrary feasible point for the definition of the robustness. They thus fulfill $(\operatorname{tr}[J_{\mathcal{E}}] + t)\mathcal{G} = \mathcal{E} + t\mathcal{F}$ for some $\mathcal{G} \in \operatorname{chan}^*[Q]$. Therefore, we get a valid QPD $\mathcal{E} = (\operatorname{tr}[J_{\mathcal{E}}] + t)\mathcal{G} - t\mathcal{F}$ with 1-norm $\operatorname{tr}[J_{\mathcal{E}}] + 2t$.

Function properties of the quasiprobability extent We now explore a few properties of $\gamma_{\mathcal{E}}$ as a function.

Lemma 3.19: For any $S \subset HP(A \to B)$, $\mathcal{E}_i \in HP(A \to B)$ for i = 1, ..., m and $a \in \mathbb{R}^m$ one has

$$\gamma_S\left(\sum_{i=1}^m a_i \mathcal{E}_i\right) \le \sum_{i=1}^m |a_i| \, \gamma_S(\mathcal{E}_i) \tag{3.56}$$

In particular, this result shows that γ_S is a convex function on the span of the decomposition set $\operatorname{span}_{\mathbb{R}}(S)$.

Proof. If the left-hand side is infinite, then there must be at least one \mathcal{E}_i for which $\gamma_S(\mathcal{E}_i) = \infty$ (as $\sum_i a_i \mathcal{E}_i$ lies in the span of the \mathcal{E}_i), so the statement trivially holds. We can thus assume without loss of generality that all involved quantities exhibit a finite quasiprobability extent.

The proof follows from the insight that a collection of QPDs for each of the \mathcal{E}_i

$$\mathcal{E}_{i} = \sum_{i=1}^{m_{i}} b_{j}^{(i)} \mathcal{F}_{j}^{(i)}$$
(3.57)

induces a QPD of $\sum_i a_i \mathcal{E}_i$

$$\sum_{i} a_{i} \mathcal{E}_{i} = \sum_{i=1}^{m} \sum_{j=1}^{m_{i}} a_{i} b_{j}^{(i)} \mathcal{F}_{j}^{(i)}$$
(3.58)

which has coefficients with a 1-norm of $\sum_i |a_i| \cdot \left(\sum_j \left|b_j^{(i)}\right|\right)$.

Lemma 3.20: For any $S \subset HP(A \to B)$ the quasiprobability extent γ_S is absolutely homogeneous, i.e.,

$$\forall \lambda \in \mathbb{R}, \forall \mathcal{E} \in HP(A \to B) : \gamma_S(\lambda \mathcal{E}) = |\lambda| \gamma_S(\mathcal{E}). \tag{3.59}$$

Proof. If $\lambda=0$ or $\mathcal{E}\notin \operatorname{span}_{\mathbb{R}}(S)$, then the statement is trivial. For the remaining case, it suffices to consider $\lambda>0$ as clearly $\gamma_S(\mathcal{E})=\gamma_S(-\mathcal{E})$. Showing $\forall \lambda>0, \forall \mathcal{E}: \gamma_S(\lambda\mathcal{E})\leq \lambda\gamma_S(\mathcal{E})$ is sufficient in that regard, as this then implies

$$\lambda \gamma_S(\mathcal{E}) = \lambda \gamma_S(\frac{1}{\lambda}\lambda \mathcal{E}) \le \gamma_S(\lambda \mathcal{E}).$$
 (3.60)

We can easily see $\gamma_S(\lambda \mathcal{E}) \leq \lambda \gamma_S(\mathcal{E})$ by verifying that every QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ induces a QPD of $\lambda \mathcal{E}$ with a 1-norm of the coefficients that is exactly λ multiplied by the 1-norm of the original QPD.

Corollary 3.21: For any bounded decomposition set $S \subset HP(A \to B)$, the quasiprobability extent γ_S is a norm on $\operatorname{span}_{\mathbb{R}}(S)$.

As an example, we will see in Chapter 4 that choosing *S* to be the set of CPTP maps results into the quasiprobability extent being the well-known diamond norm.

Proof. A norm must fulfill three properties: It is *sub-additive* $(\gamma_S(\mathcal{E} + \mathcal{F}) \leq \gamma_S(\mathcal{E}) + \gamma_S(\mathcal{F}))$, which follows from Lemma 3.19), *absolutely homogeneous* (see Lemma 3.20) and *point-separating* $(\gamma_S(\mathcal{E}) = 0 \Leftrightarrow \mathcal{E} = 0)$. It remains to show the last property. Consider some $\mathcal{E} \in \operatorname{HP}(A \to B)$ such that $\gamma_S(\mathcal{E}) = 0$. Therefore, there exists a sequence of QPDs $\mathcal{E} = \sum_i a_i^{(j)} \mathcal{F}_i^{(j)}$ for $j \in \mathbb{N}$ such that $\lim_{j \to \infty} \left\| a^{(j)} \right\|_1 = 0$. Since for all j we have

$$\|\mathcal{E}\| \le \sum_{i} \left| a_{i}^{(j)} \right| \left\| \mathcal{F}_{i}^{(j)} \right\| \le C \left\| a^{(j)} \right\|_{1} \tag{3.61}$$

for some constant $C \ge 0$, we have $||\mathcal{E}|| = 0$ and thus $\mathcal{E} = 0$ (here $||\cdot||$ denotes some arbitrary norm).

The following statements will depend on a choice of basis $\{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_m\} \subset S$ of $\operatorname{span}_{\mathbb{R}}(S)$. The gram matrix G of such a basis is defined w.r.t. the Hilbert-Schmidt inner product of the associated Choi operators $G_{ij} := \langle J_{\mathcal{B}_i}, J_{\mathcal{B}_i} \rangle$.

Lemma 3.22: Consider some $S \subset HP(A \to B)$ and a basis of $\operatorname{span}_{\mathbb{R}}(S)$ contained in S with Gram matrix G. For any $\mathcal{E} \in HP(A \to B)$ one has

$$\gamma_S(\mathcal{E}) \le \sqrt{\frac{\dim(A)\dim(B)}{\sigma_{\min}(G)}} \|J_{\mathcal{E}}\|_2$$
(3.62)

where $\sigma_{\min}(G)$ is the smallest eigenvalue of G.

For superoperators derived from a set of quantum instruments $\mathcal{E} \in \text{chan}^*[Q]$, we always have $||J_{\mathcal{E}}||_2 \leq ||J_{\mathcal{E}}||_1 \leq 1$.

Proof. Let us write the basis as $\{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_m\}$ and denote the expansion of \mathcal{E} into the basis by $\mathcal{E} = \sum_i \lambda_i \mathcal{B}_i$ where $\lambda \in \mathbb{R}^m$. Notice that

$$\frac{\|J_{\mathcal{E}}\|_{2}^{2}}{\dim(AB)} = \frac{\operatorname{tr}[J_{\mathcal{E}}^{\dagger}J_{\mathcal{E}}]}{\dim(AB)} = \langle J_{\mathcal{E}}, J_{\mathcal{E}} \rangle = \sum_{i,j} \lambda_{i} \lambda_{j} G_{ij} \ge \sigma_{\min}(G) \|\lambda\|_{2}^{2}. \tag{3.63}$$

The desired statement follows from

$$\gamma_{S}(\mathcal{E}) \leq \|\lambda\|_{1} \leq \sqrt{m} \|\lambda\|_{2} \leq \frac{\sqrt{m}}{\sqrt{\dim(AB)\sigma_{\min}(G)}} \|J_{\mathcal{E}}\|_{2} \leq \sqrt{\frac{\dim(AB)}{\sigma_{\min}(G)}} \|J_{\mathcal{E}}\|_{2} \quad (3.64)$$

where we used the Cauchy-Schwarz inequality.

As a simple consequence, this allows us to prove the Lipschitz continuity of the quasiprobability extent.

Lemma 3.23: Consider some $S \subset \operatorname{HP}(A \to B)$ and a basis of $\operatorname{span}_{\mathbb{R}}(S)$ contained in S with Gram matrix G. Then γ_S is Lipschitz continuous on $\operatorname{span}_{\mathbb{R}}(S)$

$$\forall \mathcal{F}, \mathcal{G} \in \operatorname{span}_{\mathbb{R}}(S) : |\gamma_{S}(\mathcal{E}) - \gamma_{S}(\mathcal{F})| \leq \sqrt{\frac{\dim(A)\dim(B)}{\sigma_{\min}(G)}} \|J_{\mathcal{E}} - J_{\mathcal{F}}\|_{2} \quad (3.65)$$

where $\sigma_{\min}(G)$ is the smallest eigenvalue of G.

Proof. Define $\Delta := \mathcal{F} - \mathcal{E}$. By Lemma 3.19 we have

$$\gamma_S(\mathcal{F}) = \gamma_S(\mathcal{E} + \Delta) \le \gamma_S(\mathcal{E}) + \gamma_S(\Delta)$$
 and $\gamma_S(\mathcal{E}) \le \gamma_S(\mathcal{F}) + \gamma_S(\Delta)$. (3.66)

It thus suffices to apply Lemma 3.22 and obtain

$$\gamma_S(\Delta) \le \sqrt{\frac{\dim(AB)}{\sigma_{\min}(G)}} \|J_\Delta\|_2.$$
(3.67)

Example 3.24. For single-qubit Pauli operators $Q_1, Q_2 \in P_1$ we define

$$R_{Q_1} := \frac{1}{\sqrt{2}} (\mathbb{1} + iQ_1) \quad R_{Q_1,Q_2} := \frac{1}{\sqrt{2}} (Q_1 + Q_2)$$
 (3.68)

$$\pi_{Q_1} := \frac{1}{2} (\mathbb{1} + Q_1) \quad \pi_{Q_1, Q_2} := \frac{1}{2} (Q_1 + iQ_2).$$
(3.69)

Consider the basis $B = \{\mathcal{B}_1, \dots, \mathcal{B}_{16}\}$ of the 16-dimensional space of single-qubit Hermitian-preserving operators depicted in Table 3.1. Note that R_{Q_1} is a Clifford unitary, since for $Q_3 \in P_1$ we have

$$R_{Q_1}Q_3R_{Q_1}^{\dagger} = \begin{cases} Q_3 & \text{if } [Q_1, Q_3] = 0\\ iQ_1Q_3 & \text{else} \end{cases}$$
 (3.70)

Similarly, R_{Q_1,Q_2} is also a Clifford unitary whenever Q_1 and Q_2 anti-commute

$$R_{Q_{1},Q_{2}}Q_{3}R_{Q_{1},Q_{2}}^{\dagger} = \begin{cases} Q_{3} & \text{if } [Q_{1},Q_{3}] = 0 \text{ and } [Q_{2},Q_{3}] = 0\\ -Q_{3} & \text{if } [Q_{1},Q_{3}] \neq 0 \text{ and } [Q_{2},Q_{3}] \neq 0\\ Q_{1}Q_{2}Q_{3} & \text{if } [Q_{1},Q_{3}] \neq 0 \text{ and } [Q_{2},Q_{3}] = 0\\ -Q_{1}Q_{2}Q_{3} & \text{if } [Q_{1},Q_{3}] = 0 \text{ and } [Q_{2},Q_{3}] \neq 0 \end{cases}$$
(3.71)

Therefore, the \mathcal{B}_i can all be implemented using single-qubit Clifford gates, measurements in the computational basis, classical randomness and classical side information (recall that $\mathcal{M} := \operatorname{Ad}[|0\rangle\langle 0|] - \operatorname{Ad}[|1\rangle\langle 1|]$ can be realized with classical side information as discussed in Section 3.3).

Remarkably, the Choi matrices of these 16 operators are pairwise orthogonal. The Frobenius norm of the Choi matrices is 1 for the first four elements and $1/\sqrt{2}$ for the remaining 12 elements. The associated Gram matrix is therefore diagonal with smallest eigenvalue 1/8.

The basis B can be extended to an orthogonal basis of an arbitrary n-qubit Hilbert space by simply taking the tensor products of the basis elements. The smallest eigenvalue of the Gram matrix is then 2^{-3n} . As such, Lemma 3.23 implies that for any decomposition set S containing $B^{\otimes n}$, the quasiprobability extent γ_S is Lipschitz continuous with Lipschitz constant $2^{5n/2}$.

Compact decomposition sets We will encounter many settings, where the infimum of the quasiprobability extent is always achieved.

Lemma 3.25: If *S* is compact, then the infimum in the quasiprobability extent (see Equations (3.15), (3.36) and (3.41)) is achieved on $\operatorname{span}_{\mathbb{R}}(S)$.

Proof. Without loss of generality, we can assume *S* is absolutely convex, otherwise we can consider its absolute convex hull (see Lemma 3.15, the quasiprobability extent is

```
id
      1
      2
                       Ad[X]
     3
                       Ad[Y]
                       Ad[Z]
     5
                        \frac{1}{2}(\text{Ad}[R_X] - \text{Ad}[R_{-X}])
      6
                               (\operatorname{Ad}[R_Y] - \operatorname{Ad}[R_{-Y}])
                              A(\operatorname{Ad}[R_{Z}] - \operatorname{Ad}[R_{-Z}])
                              \{(\operatorname{Ad}[R_{Y,Z}]-\operatorname{Ad}[R_{Y,-Z}])\}
     9
                            \frac{1}{2}(\text{Ad}[R_{Z,X}] - \text{Ad}[R_{Z,-X}])
                          \frac{1}{2}(\text{Ad}[R_{X,Y}] - \text{Ad}[R_{X,-Y}])
                       \operatorname{Ad}\left[\pi_{X}\right]-\operatorname{Ad}\left[\pi_{-X}\right]
                                                                                                                                                                        = Ad [R_Z^3 R_X^3] \circ \mathcal{M} \circ Ad [R_X R_Z]
11
                     \begin{array}{ll} \operatorname{Ad}\left[R_{X}R_{Z}\right] & \operatorname{Ad}\left[R_{X}R_{Z}\right] \\ \operatorname{Ad}\left[\pi_{Z}\right] - \operatorname{Ad}\left[\pi_{-Z}\right] & = & \operatorname{Ad}\left[R_{X}\right] \circ \mathcal{M} \circ \operatorname{Ad}\left[R_{X}^{3}\right] \\ \operatorname{Ad}\left[\pi_{Z}\right] - \operatorname{Ad}\left[\pi_{-Z}\right] & = & \operatorname{Ad}\left[R_{X}^{3}R_{X}^{3}\right] \circ \mathcal{M} \circ \operatorname{Ad}\left[R_{X}^{3}R_{Z}\right] \\ \operatorname{Ad}\left[\pi_{Z,X}\right] - \operatorname{Ad}\left[\pi_{X,Z}\right] & = & \operatorname{Ad}\left[R_{X}\right] \circ \mathcal{M} \circ \operatorname{Ad}\left[R_{X}^{3}R_{Z}\right] \\ \operatorname{Ad}\left[\pi_{Z,X}\right] - \operatorname{Ad}\left[\pi_{X,Z}\right] & = & \operatorname{Ad}\left[R_{X}\right] \circ \mathcal{M} \circ \operatorname{Ad}\left[R_{X}^{3}R_{Z}\right] \\ \operatorname{Ad}\left[\pi_{Z,X}\right] - \operatorname{Ad}\left[\pi_{X,Z}\right] & = & \operatorname{Ad}\left[R_{X}\right] \circ \mathcal{M} \circ \operatorname{Ad}\left[R_{X}^{3}R_{Z}\right] \end{array}
12
13
14
15
                       \operatorname{Ad}\left[\pi_{X,Y}\right] - \operatorname{Ad}\left[\pi_{Y,X}\right]
                                                                                                                                                                                            \mathcal{M} \circ \operatorname{Ad} \left[ R_{Y}^{2} \right]
16
```

Table 3.1: 16 Clifford operations constituting a basis of the 16-dimensional space of Hermitian-preserving single-qubit superoperators. Here, $\mathcal{M} := \operatorname{Ad}[|0\rangle\langle 0|] - \operatorname{Ad}[|1\rangle\langle 1|]$ as in Section 3.3.

clearly achieved on S if and only if it is achieved on $\operatorname{aconv}(S)$). For any $\mathcal{E} \in \operatorname{span}_{\mathbb{R}}(S)$ there must exist a QPD $\mathcal{E} = \tau \mathcal{F}$ for some $\mathcal{F} \in \operatorname{aconv}(S)$. Hence, we can write

$$\gamma_{S}(\mathcal{E}) = \inf_{\kappa \in [0,\tau], \mathcal{G} \in \text{aconv}(S)} \{ \kappa | \mathcal{E} = \kappa \mathcal{G} \}.$$
(3.72)

Since the absolute convex hull of a compact set is again compact, we are optimizing a continuous function over a compact set, and as such the infimum must be achieved.

The next property shows that the quasiprobability extent can provide a faithful criterion to determine whether an element is part of the decomposition set or not.

Lemma 3.26: Let $S \subset HP(A \to B)$ be compact and $\mathcal{E} \in HP(A \to B)$. Then S is absolutely convex if and only if γ_S fulfills

$$\gamma_S(\mathcal{E}) \le 1 \iff \mathcal{E} \in S.$$
(3.73)

Proof. Consider S to be absolutely convex. Clearly $\mathcal{E} \in S \Rightarrow \gamma_S(\mathcal{E}) \leq 1$ and $\gamma_S(\mathcal{E}) \leq 1$ implies by Lemma 3.14 and Lemma 3.25 that $\mathcal{E} \in S$.

Conversely, consider a S such that γ_S fulfills the faithfulness criterion. Then, for any absolutely convex combination $\sum_i a_i \mathcal{F}_i$ with $\sum_i |a_i| \leq 1$ and $\mathcal{F}_i \in S$ we have $\gamma_S(\sum_i a_i \mathcal{F}_i) \leq \sum_i |a_i| \gamma_S(\mathcal{F}_i) \leq 1$ by Lemma 3.19. Hence, $\sum_i a_i \mathcal{F}_i$ is an element of S.

50

If we can additionally assume \mathcal{E} to be trace-preserving, then a tighter statement can be made.

Lemma 3.27: Let $S \subset HP(A \to B)$ be convex and compact. Furthermore, let either $S \subset TP(A \to B)$ or $S = \operatorname{chan}^*[Q]$ for some $Q \subset QI(A \to B)$ that is closed under mixture. For any $\mathcal{E} \in HPTP(A \to B)$ we have

$$\gamma_S(\mathcal{E}) = 1 \Longleftrightarrow \mathcal{E} \in S. \tag{3.74}$$

Proof. If $\mathcal{E} \in S$ then $\gamma_S(\mathcal{E}) \leq 1$. The other inequality $\gamma_S(\mathcal{E}) \geq 1$ follows directly from Lemma 3.17 or Lemma 3.18 and the non-negativity of the robustness.

Conversely, let $\gamma_S(\mathcal{E}) = 1$. If $S = \text{chan}^*[Q]$, then Lemma 3.26 applies. Otherwise, by Lemmas 3.11 and 3.25 there exists a QPD $\mathcal{E} = a^+\mathcal{F}^+ - a^-\mathcal{F}^-$ with $a^+ + a^- = 1$. Applying the trace on the Choi representation of both sides, we get $a^+ - a^- = 1$ and therefore $a^+ = 1, a^- = 0$. Hence, $\mathcal{E} \in S$.

3.5 The quasiprobability extent in quantum resource theories

QPS lends itself very well to being used in a resource theoretic context. A QRT separates all channels into a set of "free" and "non-free" operations. As such, it is natural to study how expensive it is to simulate a non-free operation using free operations. Most applications of QPS we will consider in this thesis can be framed in the context of some QRT. For instance, the near-Clifford simulation procedure introduced in Section 3.1 is tightly connected to the resource theory of magic. The non-free resources in this case are the magic states and the non-Clifford operations.

More mathematically, for some QRT $\mathfrak O$ we will now study the quasiprobability extent $\gamma_{\mathfrak O^*}$ for

$$\mathfrak{O}^{*}(A \to B) := \operatorname{chan}^{*}[\mathfrak{I}(A \to B)] \tag{3.75}$$

where \Im are the free quantum instruments associated to \Im . This choice of decomposition set allows us to capture QPS with classical side information, as discussed in Section 3.3.

Generally, we will consider convex QRTs, i.e., resource theories where the set of free channels forms a convex set. This property can also be characterized in different ways.

Lemma 3.28: Let \mathfrak{O} be a QRT in which classical processing is free. The following are equivalent.

- (a) $\mathfrak{O}(A \to B)$ is convex for all A, B.
- (b) $\Im(A \to B)$ is closed under mixture for all A, B.
- (c) $\mathfrak{D}^*(A \to B)$ is absolutely convex for all A, B.

Proof. To show (a) \Rightarrow (b), recall that $(\mathcal{E}_i)_i$, $(\mathcal{F}_j)_j \in \mathfrak{I}(A \to B)$ implies that there exist free channels of the form $\sum_i \mathcal{E}_i \otimes |i\rangle \langle i|_{X_1} \in \mathfrak{D}(A \to BX_1)$ and $\sum_j \mathcal{F}_j \otimes |j\rangle \langle j|_{X_2} \in \mathfrak{D}(A \to BX_2)$. Taking a convex mixture of these two measurement channels corresponds exactly to taking a mixture of the two original instruments. To achieve this, appropriate classical processing with padding is required on X_1 and X_2 such that the two measurement channels operate on the same spaces.

The statement (b) \Rightarrow (c) was already proven in Lemma 3.13.

For the final direction (c) \Rightarrow (a), consider two channels $\mathcal{E}, \mathcal{F} \in \mathfrak{O}(A \to B)$. We know that any convex mixture $(1-p)\mathcal{E}+p\mathcal{F}$ lies in $\mathfrak{O}^*(A \to B)$, and we now need to show that it also lies in $\mathfrak{O}(A \to B)$. By the definition of $\mathfrak{O}^*(A \to B)$, we can write

$$(1-p)\mathcal{E} + p\mathcal{F} = \sum_{i} \beta_{i} \mathcal{G}_{i}$$
(3.76)

for some $\beta_i \in [-1,1]$ and $(\mathcal{G}_i)_i \in \mathfrak{I}(A \to B)$. Taking the Choi isomorphism on both sides, as well as the partial trace tr_B thereof, we obtain

$$\frac{1}{\dim(A)}\mathbb{1} = \sum_{i} \beta_{i} \operatorname{tr}_{B}[J_{\mathcal{G}_{i}}]. \tag{3.77}$$

At the same time, we know that $\sum_{i} G_i \in CPTP(A \to B)$, so

$$\frac{1}{\dim(A)}\mathbb{1} = \sum_{i} \beta_{i} \operatorname{tr}_{B}[J_{\mathcal{G}_{i}}] \preccurlyeq \sum_{i} \operatorname{tr}_{B}[J_{\mathcal{G}_{i}}] = \frac{1}{\dim(A)}\mathbb{1}.$$
 (3.78)

This implies that for all i we have $\beta_i = 1$, unless \mathcal{G}_i is itself zero. Therefore, $(1 - p)\mathcal{E} + p\mathcal{F} \in \text{chan}[\mathfrak{I}(A \to B)]$, which is equal to $\mathfrak{O}(A \to B)$ by Lemma 2.14.

In the context of QPS, convexity of a QRT is very natural, as the QPS procedure itself requires the ability to randomly pick and apply a channel during the circuit execution.

The set of free states of a QRT is denoted by $\mathfrak{F}(A)$ for some given system A. It can be understood to be the free channels from the trivial Hilbert space \mathbb{C} to A, i.e., $\mathfrak{F}(A) := \mathfrak{O}(\mathbb{C} \to A) \subset \mathrm{D}(A)$. As such, one might a priori expect that the simulation of non-free states $\omega \in \mathrm{D}(A)$ should involve QPDs w.r.t. $\mathfrak{O}^{\star}(\mathbb{C},A)$ to fully harness the utility of classical side information. However, it turns out that classical side information does not provide any benefit for simulating states.

Lemma 3.29: Let $\mathfrak O$ be a QRT that fulfills the axiom of free instruments and A be some system. Then $\operatorname{aconv}(\mathfrak O^\star(\mathbb C,A)) = \operatorname{aconv}(\mathfrak O(\mathbb C,A))$ and hence

$$\forall \omega \in \mathrm{H}(A) : \gamma_{\mathfrak{D}^{\star}(\mathbb{C}, A)}(\omega) = \gamma_{\mathfrak{D}(\mathbb{C}, A)}(\omega) = \gamma_{\mathfrak{F}(A)}(\omega). \tag{3.79}$$

Proof. The second statement follows from Lemma 3.15. Since $\mathfrak{O}(\mathbb{C}, A) \subset \mathfrak{O}^*(\mathbb{C}, A)$, it suffices to show that $\mathfrak{O}^*(\mathbb{C}, A) \subset \operatorname{aconv}(\mathfrak{O}(\mathbb{C}, A))$. Consider an element $\sum_i \beta_i \mathcal{E}_i \in \mathfrak{O}^*(\mathbb{C}, A)$. The \mathcal{E}_i can be identified with positive operators $\rho_i \in H(A)$ which fulfill $\operatorname{tr}[\sum_i \rho_i] = 1$. By the axiom of free instruments, all $\rho_i / \operatorname{tr}[\rho_i]$ must be

free states $\rho_i/\operatorname{tr}[\rho_i] \in \mathfrak{F}(A)$, respectively $\mathcal{E}_i/\operatorname{tr}[\rho_i] \in \mathfrak{O}(\mathbb{C},A)$, which implies that $\sum_i \beta_i \operatorname{tr}[\rho_i](\mathcal{E}_i/\operatorname{tr}[\rho_i]) \in \operatorname{aconv}(\mathfrak{O}(\mathbb{C},A))$. Note that this proof would not work for channels (i.e. $A \neq \mathbb{C}$) as the \mathcal{E}_i can generally not be normalized by a scalar in order to obtain a valid quantum channel.

For notational convenience, we will usually omit the involved systems in the decomposition set when they are clear from context. For instance, we simply write $\gamma_{\mathfrak{D}^{\star}}(\mathcal{E})$ and $\gamma_{\mathfrak{F}}(\omega)$, as the systems are implicitly specified by $\mathcal{E} \in \mathrm{HP}(A \to B)$ and $\omega \in \mathrm{H}(A)$. The rest of this section is dedicated to studying various properties of the quasiprobability extent $\gamma_{\mathfrak{D}^{\star}}$. While we will focus on proving these properties for $\gamma_{\mathfrak{D}^{\star}}$, many of them do also hold for $\gamma_{\mathfrak{D}}^{7}$.

Chaining property Recall that the free operations of a QRT are by definition closed under composition. We first show that the same holds for $\mathfrak{O}^*(A \to B)$ if the QRT additionally has a tensor product structure.

Lemma 3.30: Let \mathfrak{D} be a QRT with tensor product structure and $\mathcal{E} \in \mathfrak{D}^*(A \to B)$, $\mathcal{F} \in \mathfrak{D}^*(B \to C)$. Then $\mathcal{F} \circ \mathcal{E} \in \mathfrak{D}^*(A \to C)$.

Proof. We can write $\mathcal{E} = \sum_i \alpha_i \mathcal{E}_i$ and $\mathcal{F} = \sum_j \beta_j \mathcal{F}_j$ for $(\mathcal{E}_i)_i \in \mathfrak{I}(A \to B)$, $(\mathcal{F}_j)_j \in \mathfrak{I}(B \to C)$ and $\alpha_i, \beta_j \in [-1, 1]$. By the definition of \mathfrak{I} , there exist free channels $\tilde{\mathcal{E}} \in \mathfrak{D}(A \to BX_1)$ and $\tilde{\mathcal{F}} \in \mathfrak{D}(B \to CX_2)$ given by

$$\tilde{\mathcal{E}}(\rho) = \sum_{i} \mathcal{E}_{i}(\rho) \otimes |i\rangle\langle i| , \quad \tilde{\mathcal{F}}(\rho) = \sum_{j} \mathcal{F}_{j}(\rho) \otimes |j\rangle\langle j| . \tag{3.80}$$

By the tensor product structure, $(\tilde{\mathcal{F}} \otimes \mathrm{id}_{X_1}) \circ \tilde{\mathcal{E}}$ is also a free operation in $\mathfrak{O}(A \to CX_1X_2)$, which implies that the quantum instrument $(\mathcal{F}_j \circ \mathcal{E}_i)_{i,j}$ is free. Therefore, $\mathcal{F} \circ \mathcal{E} = \sum_{i,j} \alpha_i \beta_j \mathcal{F}_j \circ \mathcal{E}_i$ is an element of $\mathfrak{O}^*(A \to C)$.

As a direct consequence, the quasiprobability extent fulfills the *chaining property*.

Lemma 3.31: For any QRT $\mathfrak O$ with tensor product structure and any $\mathcal E\in \mathrm{HP}(A\to B)$, $\mathcal F\in \mathrm{HP}(B\to C)$ one has

$$\gamma_{\mathfrak{D}^{\star}}(\mathcal{F} \circ \mathcal{E}) \le \gamma_{\mathfrak{D}^{\star}}(\mathcal{E})\gamma_{\mathfrak{D}^{\star}}(\mathcal{F}).$$
 (3.81)

To prove this, notice that any two QPDs of \mathcal{E} and \mathcal{F} can be merged into a QPD of $\mathcal{F} \circ \mathcal{E}$. As an important special case, one can choose A to be the trivial Hilbert space, which gives us a relation between the quasiprobability extent of states and of channels.

 $^{^7}$ In fact, these properties for $\gamma_{\mathfrak{D}}$ are rather direct consequences of the definition of a QRT. Showing them for $\gamma_{\mathfrak{D}^*}$ requires a bit more work.

Corollary 3.32: Let $\mathfrak O$ be a QRT with tensor product structure, $\omega \in \mathrm{H}(A)$ and $\mathcal E \in \mathrm{HP}(A \to B)$. Then

$$\gamma_{\mathfrak{D}^{\star}}(\mathcal{E}(\omega)) \le \gamma_{\mathfrak{D}^{\star}}(\mathcal{E})\gamma_{\mathfrak{D}^{\star}}(\omega).$$
 (3.82)

Another consequence is that the quasiprobability extent is invariant under reversible free operations.

Corollary 3.33: Consider a QRT $\mathfrak O$ with tensor product structure. Let $\mathcal U_1 \in \mathfrak O(A)$, $\mathcal U_2 \in \mathfrak O(B)$ be unitary channels s.t. their inverses $\mathcal U_1^{-1}$, $\mathcal U_2^{-1}$ are also free. For any $\mathcal E \in \operatorname{HP}(A \to B)$ one has

$$\gamma_{\mathfrak{D}^{\star}}(\mathcal{U}_{2} \circ \mathcal{E}) = \gamma_{\mathfrak{D}^{\star}}(\mathcal{E}) = \gamma_{\mathfrak{D}^{\star}}(\mathcal{E} \circ \mathcal{U}_{1}). \tag{3.83}$$

Proof. We only show the first equality, the other is analogous. On one hand

$$\gamma_{\mathfrak{D}^{\star}}(\mathcal{U}_{2} \circ \mathcal{E}) \leq \gamma_{\mathfrak{D}^{\star}}(\mathcal{U}_{2})\gamma_{\mathfrak{D}^{\star}}(\mathcal{E}) \leq \gamma_{\mathfrak{D}^{\star}}(\mathcal{E})$$
(3.84)

and conversely

$$\gamma_{\mathfrak{D}^{\star}}(\mathcal{E}) = \gamma_{\mathfrak{D}^{\star}}(\mathcal{U}_{2}^{-1} \circ \mathcal{U}_{2} \circ \mathcal{E}) \leq \gamma_{\mathfrak{D}^{\star}}(\mathcal{U}_{2}^{-1})\gamma_{\mathfrak{D}^{\star}}(\mathcal{U}_{2} \circ \mathcal{E}) \leq \gamma_{\mathfrak{D}^{\star}}(\mathcal{U}_{2} \circ \mathcal{E}). \tag{3.85}$$

Properties under tensor product Here, we investigate properties of the quasiprobability extent that are induced by the tensor product structure of the QRT. First, we note that the decomposition set is closed under tensor product with the identity.

Lemma 3.34: Let \mathfrak{O} be a QRT with tensor product structure, $\mathcal{E} \in \mathfrak{O}^*(A \to B)$ and C some system. Then $(\mathcal{E} \otimes \mathrm{id}_C) \in \mathfrak{O}^*(AC \to BC)$.

Proof. Let $\mathcal{E} = \sum_j \beta_j \mathcal{E}_j$ for $(\mathcal{E}_j)_j \in \mathfrak{I}(A \to B)$ and $\beta_j \in [-1,1]$. Therefore, the channel $\rho \mapsto \sum_j \mathcal{E}_j(\rho) \otimes |j\rangle\langle j|$ is free. Taking the tensor product of this channel with id_C yields another free channel (by the tensor product structure) and therefore $(\mathcal{E}_j \otimes \mathrm{id}_C)_j \in \mathfrak{I}(AC \to BC)$. Hence, $(\mathcal{E} \otimes \mathrm{id}_C) = \sum_j \beta_j (\mathcal{E}_j \otimes \mathrm{id}_C)$ is in $\mathfrak{D}^*(AC \to BC)$.

This result can be directly applied to the quasiprobability extent.

Lemma 3.35: Let \mathfrak{O} be a QRT with tensor product structure and $\mathcal{E} \in HP(A \to B)$. For any system C

$$\gamma_{\mathfrak{D}^{\star}}(\mathcal{E} \otimes \mathrm{id}_{\mathcal{C}}) = \gamma_{\mathfrak{D}^{\star}}(\mathcal{E}).$$
 (3.86)

Proof. For the side \leq , Lemma 3.34 implies that any QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ directly induces a valid QPD $\mathcal{E} \otimes \mathrm{id}_{\mathcal{C}} = \sum_i a_i \mathcal{F}_i \otimes \mathrm{id}_{\mathcal{C}}$. For the converse side \geq , we use Lemma 3.31 and

$$\mathcal{E} = (\mathrm{id}_B \otimes \mathrm{tr}_E) \circ (\mathcal{E} \otimes \mathrm{id}_E) \circ (\mathrm{id}_A \otimes \mathcal{G}) \tag{3.87}$$

54

where $\mathcal{G} \in \operatorname{CPTP}(\mathbb{C} \to E)$ is a state preparation channel that prepares some arbitrary free state. Since tracing out and preparing free states are free operations (see Definition 2.8 and Lemma 2.11), so are $(\operatorname{id}_B \otimes \operatorname{tr}_E)$ and $(\operatorname{id}_A \otimes \mathcal{F})$ by the completely free operations property.

Remark 3.36: In Sections 3.2 and 3.3, we showed that the standard strategy for finding QPDs of large circuits consists of combining QPDs of the constant-sized building blocks that make up that circuit. The results in Lemmas 3.30 and 3.34 justify why this is generally possible. More precisely, they imply that if we each have a QPD of two superoperators $\mathcal{E} \in \mathrm{HP}(A \to BC)$ and $\mathcal{F} \in \mathrm{HP}(CD \to E)$, then they can be combined into a QPD of $(\mathcal{F} \circ \mathcal{E}) \in \mathrm{HP}(AD \to BE)$. By repeatedly applying this procedure on two subsequent building blocks in a circuit, we can combine all individual QPDs into one large QPD of the total circuit.

Together, Lemmas 3.31 and 3.35 imply that the quasiprobability extent is submultiplicative.

Corollary 3.37: Let $\mathfrak O$ be a QRT with tensor product structure and $\mathcal E \in \mathrm{HP}(A \to B)$, $\mathcal F \in \mathrm{HP}(C \to D)$. Then

$$\gamma_{\mathfrak{D}^{\star}}(\mathcal{E} \otimes \mathcal{F}) \le \gamma_{\mathfrak{D}^{\star}}(\mathcal{E})\gamma_{\mathfrak{D}^{\star}}(\mathcal{F}) \tag{3.88}$$

Properties from free classical computation Whenever the classical output of a measurement can be further processed using free operations, the induced decomposition set of the QRT can be written in a simpler form.

Corollary 3.38: Let $\mathfrak O$ be a QRT with tensor product structure where classical computation channels (possibly involving classical randomness) are free. Then, for all systems A and B one has

$$\mathfrak{D}^{\star}(A \to B) = \{ \mathcal{E}^+ - \mathcal{E}^- | (\mathcal{E}^+, \mathcal{E}^-) \in \mathfrak{I}(A \to B) \}. \tag{3.89}$$

This is a direct consequence from Lemma 3.9. Indeed, any resource theory in which classical computation is free, the coarse-graining map (called f in Definition 2.6) can be freely implemented on the classical measurement outcome register. As such, a coarse-graining of a free instrument is again itself a free instrument. Similarly, $\Im(A \to B)$ is also trivially fine-grainable, since one can freely process a measurement result by extending the classical register with a random classical bit.

The quasiprobability extent as a resource measure Even though the definition of the quasiprobability extent is operationally motivated by the simulation overhead of QPS, we have remarkably encountered many properties of it that are desirable for a measure to quantify resources in a QRT. In this paragraph, we briefly summarize these insights and formally frame the quasiprobability extent as a resource measure.

Consider a QRT that fulfills the axiom of free instruments. We start by considering $\gamma_{\mathfrak{F}}(\rho)$ (which is equal to $\gamma_{\mathfrak{D}^*}(\rho)$ by Lemma 3.29) as a measure to quantify the amount of non-free resource in a state ρ . To satisfy the commonly used conventions for normalization and sub-additivity (as opposed to sub-multiplicativity), we will consider the *log-quasiprobability extent* $\log(\gamma_{\mathfrak{F}}(\rho))$.

Proposition 3.39

Let $\mathfrak O$ be a QRT with tensor product structure that fulfills the axiom of free instruments. The log-quasiprobability extent $\log \gamma_{\mathfrak F}$ is a resource measure that is additionally

- quasi-convex
- sub-additive
- strongly monotonous if D is convex and closed
- faithful if $\mathfrak D$ is convex and closed.

Proof. The log-quasiprobability extent is clearly non-negative: For any $\rho \in D(A)$ with QPD $\rho = \sum_i a_i \sigma_i$ one has

$$1 = \|\rho\|_1 \le \sum_{i} |a_i| \|\sigma_i\|_1 = \|a\|_1.$$
 (3.90)

Normalization follows directly from $1 \in \mathfrak{F}(\mathbb{C})$. We have already proven weak monotonicity (Corollary 3.32), quasi-convexity (Lemma 3.19), sub-additivity (Corollary 3.37) and faithfulness (Lemma 3.27). It only remains to show strong monotonicity.

Let $\rho \in D(A)$ and $(\mathcal{E}_i)_i \in \mathfrak{I}(A \to B)$ be a free quantum instrument. Let us write the optimal QPD of ρ as $\rho = (1+t)\rho^+ - t\rho^-$ (recall Lemma 3.14). This allows us to write

$$\frac{\mathcal{E}_{i}(\rho)}{\operatorname{tr}[\mathcal{E}_{i}(\rho)]} = \frac{1}{\operatorname{tr}[\mathcal{E}_{i}(\rho)]} \left((1+t)\mathcal{E}_{i}(\rho^{+}) - t\mathcal{E}_{i}(\rho^{-}) \right) = (1+\tilde{t}_{i}) \frac{\mathcal{E}_{i}(\rho^{+})}{\operatorname{tr}[\mathcal{E}_{i}(\rho^{+})]} - \tilde{t}_{i} \frac{\mathcal{E}_{i}(\rho^{-})}{\operatorname{tr}[\mathcal{E}_{i}(\rho^{-})]}$$
(3.91)

where $\tilde{t}_i := \frac{\operatorname{tr}[\mathcal{E}_i(\rho^-)]}{\operatorname{tr}[\mathcal{E}_i(\rho)]} t$. Hence, we have

$$\sum_{i} \operatorname{tr}[\mathcal{E}_{i}(\rho)] \gamma_{\mathfrak{F}}\left(\frac{\mathcal{E}_{i}(\rho)}{\operatorname{tr}[\mathcal{E}_{i}(\rho)]}\right) \leq \sum_{i} \operatorname{tr}[\mathcal{E}_{i}(\rho)] (1 + 2\tilde{t}_{i}) \leq 1 + 2\sum_{i} \operatorname{tr}[\mathcal{E}_{i}(\rho^{-})] t = \gamma_{\mathfrak{F}}(\rho).$$
(3.92)

While it is remarkable that the quasiprobability extent has strong enough properties to be an interesting resource measure, one should not forget that these properties even hold for general Hermitian operators and not just states.

For a convex resource theory, the quasiprobability extent of $\omega \in H(A)$ is directly related to the robustness $\gamma_{\mathfrak{F}}(\rho)=1+2R_{\mathfrak{F}}(\omega)$ (see Lemma 3.17). In QRT literature, it is more common to instead consider the robustness as a resource measure. Most of the properties of the quasiprobability extent directly translate. The only major

difference is that the robustness lacks the sub-additive behavior, but it is convex instead of quasi-convex.

Corollary 3.40: Let $\mathfrak O$ be a convex QRT with tensor product structure that fulfills the axiom of free instruments. The robustness $R_{\mathfrak F}$ is a resource monotone that is additionally faithful if $\mathfrak O$ is closed.

Remark 3.41: Our definition of the robustness (sometimes called *total robustness*) should not be confounded with the so-called *global robustness* [65], [151] which characterizes the robustness w.r.t. arbitrary (possibly non-free) noise. The global robustness happens to be equivalent to the max-relative entropy [49], [65, Lemma 10.1].

Remark 3.42: Note that in literature, one sometimes encounters a resource measure called *log-robustness* which is defined as $\log(1+R_{\mathfrak{F}})$ as opposed to $\log(\gamma_{\mathfrak{F}}) = \log(1+2R_{\mathfrak{F}})$ [65]. Our quantity $\log(\gamma_{\mathfrak{F}})$ can in some sense be argued to be more natural, as its sub-additivity directly mirrors the sub-multiplicativity of $\gamma_{\mathfrak{F}}$.

We now turn our attention to quantifying the amount of non-free resource in a channel instead of a state. In complete analogy with Proposition 3.39, we summarize as follows.

Proposition 3.43

For any dynamical QRT, the log-quasiprobability extent $\log(\gamma_{\mathfrak{D}^*})$ is a dynamical resource measure that is additionally quasi-convex, sub-additive and faithful if \mathfrak{D}^* is convex and closed.

Proof. This result is an amalgamation of Lemmas 3.19 and 3.27 and Corollary 3.37. The only new property to show is weak monotonicity, which directly follows from proving that for any $\Theta \in \mathfrak{S}(A, B \to C, D)$ one has $\Theta(\mathfrak{D}^*(A \to B)) \subset \mathfrak{D}^*(C \to D)$. For this purpose, let $\mathcal{E} = \sum_j \beta_j \mathcal{E}_j \in \mathfrak{D}^*(A \to B)$. Denote the associated free measurement channel by $\tilde{\mathcal{E}} = \sum_j \mathcal{E}_j \otimes |j\rangle\langle j| \in \mathfrak{D}(A \to BX)$. Since $\Theta \otimes \mathbb{1}_{T(C \to X)}$ is itself also free, Lemma 2.19 implies that

$$(\Theta \otimes \mathbb{1}_{\mathsf{T}(\mathbb{C} \to X)})(\tilde{\mathcal{E}}) = \sum_{j} \Theta(\mathcal{E}_{j}) \otimes |j\rangle\langle j|$$
(3.93)

is a free channel in $\mathfrak{O}(C \to DX)$. Therefore, $(\Theta(\mathcal{E}_j))_j$ is a free instrument in $\mathfrak{I}(C \to D)$ and $\Theta(\mathcal{E}) = \sum_j \beta_j \Theta(\mathcal{E}_j)$ lies in $\mathfrak{O}^*(C \to D)$.

For a convex QRT, the quasiprobability extent of a channel $\mathcal{E} \in \operatorname{CPTP}(A \to B)$ is related to the robustness by $\gamma_{\mathfrak{D}^*}(\mathcal{E}) = 1 + 2R_{\mathfrak{D}^*}(\mathcal{E})$ (see Lemma 3.18). As such, $R_{\mathfrak{D}^*}(\mathcal{E})$ is also a dynamical resource measure that inherits the properties of $\log(\gamma_{\mathfrak{D}^*})$.

Asymptotic simulation cost In Corollary 3.37, we observed that the quasiprobability extent exhibits sub-multiplicative behavior. In fact, we will later see that it is oftentimes even *strictly* sub-multiplicative. This has drastic implications for the sampling overhead of QPS: It can be significantly more efficient to jointly simulate multiple operations in parallel instead of optimally simulating them individually.

When simulating the n-fold copy $\mathcal{E}^{\otimes n}$ of the superoperator \mathcal{E} , we can capture the asymptotic *effective* sampling overhead per instance by regularizing the quasiprobability extent.

Definition 3.44: Let $\mathfrak O$ be a QRT with tensor product structure. We define the regularized quasiprobability extent of $\mathcal E \in \mathrm{HP}(A \to B)$ as

$$\overline{\gamma}_{\mathfrak{D}^{\star}}(\mathcal{E}) := \lim_{n \to \infty} \gamma_{\mathfrak{D}^{\star}}(\mathcal{E}^{\otimes n})^{1/n}. \tag{3.94}$$

Note that this regularization is equivalent to the "additive" regularization of the log-quasiprobability extent

$$\log \overline{\gamma}_{\mathfrak{D}^{\star}}(\mathcal{E}) = \lim_{n \to \infty} \frac{1}{n} \log \gamma_{\mathfrak{D}^{\star}}(\mathcal{E}^{\otimes n}). \tag{3.95}$$

Example 3.45. We again return to the Clifford+T setting from Section 3.1. In Example 3.6, we have already shown how using the magic state injection circuit allows an arbitrary Clifford+T circuit \mathcal{C} can be rewritten as

$$\mathcal{C}(|0\rangle\langle 0|^{\otimes n}) = \mathcal{C}'(|0\rangle\langle 0|^{\otimes n} \otimes |H\rangle\langle H|^{m})$$
(3.96)

for some appropriately chosen Clifford circuit \mathcal{C}' . Notice that this reduction from gate to state allowed us to assume that all magic states are prepared at the very beginning of the circuit, and more importantly, in parallel. The quasiprobability extent of magic states is strictly sub-multiplicative: As previously mentioned $\gamma_{\mathfrak{F}}(|H\rangle\langle H|) = \sqrt{2} = 1.414$ whereas the regularized quasiprobability extent is estimated to be $\overline{\gamma}_{\mathfrak{F}}(|H\rangle\langle H|) \approx 1.283 \pm 0.002$ [79]. As a consequence, the simulation cost for a Clifford+T circuit can be decreased from $\mathcal{O}(\gamma_{\mathfrak{F}}(|H\rangle\langle H|)^{2m})$ to $\mathcal{O}(\overline{\gamma}_{\mathfrak{F}}(|H\rangle\langle H|)^{2m})$.

A more detailed discussion, including a precise definition of the decomposition set, will follow in Chapter 6. $\hfill \triangle$

Interestingly, $\overline{\gamma}$ is arguably not the most meaningful quantity to operationally quantify the asymptotic simulation cost of QPS. In general, one can also consider *approximate* QPS with an asymptotically vanishing error. To capture the minimum overhead for the simulation of some superoperator \mathcal{E} when we allow for some error of magnitude ϵ , we introduce the *smoothed* quasiprobability extent.

Definition 3.46: Let $\mathfrak O$ be a QRT, $\mathcal E \in \operatorname{CPTP}(A \to B)$ and $\epsilon \ge 0$. The smoothed quasiprobability extent of $\mathcal E$ is defined as

$$\gamma_{\mathfrak{D}^{\star}}^{\epsilon}(\mathcal{E}) := \min_{\mathcal{F} \in B_{\epsilon}(\mathcal{E})} \gamma_{\mathfrak{D}^{\star}}(\mathcal{F})$$
(3.97)

where
$$B_{\epsilon}(\mathcal{E}) := \{ \mathcal{F} \in \operatorname{CPTP}(A \to B) | \frac{1}{2} \| \mathcal{E} - \mathcal{F} \|_{\Diamond} \leq \epsilon \}.$$

Remark 3.47: Note that for technical reasons, we chose to smooth only over CPTP maps instead of general HPTP operations. Therefore, we also restricted $\mathcal E$ to be CPTP itself. In later sections, this will allow us to directly relate $\gamma^{\epsilon}_{\mathcal D^{\star}}$ to other smoothed resource measures of QRTs, which are typically also defined by smoothing over physical operations. In principle, one could also define the smoothed quasiprobability extent without the CPTP constraint.

The asymptotic overhead with a vanishing error is thus captured by following quantity.

Definition 3.48: Let $\mathfrak O$ be a QRT with tensor product structure. We define the *asymptotic quasiprobability extent* of some $\mathcal E \in \mathrm{HP}(A \to B)$ as

$$\gamma_{\mathfrak{D}^{\star}}^{\infty}(\mathcal{E}) := \lim_{\epsilon \to 0^{+}} \liminf_{n \to \infty} \gamma_{\mathfrak{D}^{\star}}^{\epsilon}(\mathcal{E}^{\otimes n})^{1/n}. \tag{3.98}$$

Notice that we use a lim inf since it is a priori unclear if the sequence $\gamma_{\mathfrak{D}^*}^{\epsilon}(\mathcal{E}^{\otimes n})^{1/n}$ converges. It is however bounded by

$$\gamma_{\mathfrak{D}^{\star}}^{\infty}(\mathcal{E}) \leq \lim_{\epsilon \to 0^{+}} \liminf_{n \to \infty} \gamma_{\mathfrak{D}^{\star}}(\mathcal{E}^{\otimes n})^{1/n} = \overline{\gamma}_{\mathfrak{D}^{\star}}(\mathcal{E}). \tag{3.99}$$

In the study of QPS for various QRTs, it is an important task to understand the relation between $\gamma_{\mathfrak{D}^*}$, $\overline{\gamma}_{\mathfrak{D}^*}$ and $\gamma_{\mathfrak{D}^*}^{\infty}$. We note that for many QRTs, computing the quasiprobability extent of an operation is a very difficult task and often only possible in restricted cases. In extension, computing the regularized quasiprobability extent is typically harder, and for the asymptotic quasiprobability extent it is even harder.

For the resource theory of entanglement in Chapter 5, we will find an explicit expression for $\overline{\gamma}_{\mathfrak{D}^*}$ and $\gamma_{\mathfrak{D}^*}^{\infty}$ for pure entangled states. Remarkably, we will observe that $\gamma_{\mathfrak{D}^*}^{\infty}$ can be strictly smaller, meaning that erroneous QPS with vanishingly small error is asymptotically more efficient than exact QPS.

3.6 Convex programming for finite decomposition sets

In this section, we will consider decomposition sets that are finite $S = \{\mathcal{F}_1, \dots, \mathcal{F}_M\}$ (or equivalently convex polytopes, by Lemma 3.15). Such situations occur naturally in some settings, for example in the resource theory of magic (see Chapter 6) or for probabilistic error cancellation (see Chapter 7). Even when the decomposition

set is infinite, it can be useful to pick a finite subset to obtain upper bounds to the quasiprobability extent through Lemma 3.10.

For a finite decomposition set $S \subset HP(A \to B)$ with M elements, the quasiprobability extent of $\mathcal{E} \in HP(A \to B)$ is given by

$$\gamma_S(\mathcal{E}) = \begin{cases} \min_{a \in \mathbb{R}^M} & \|a\|_1 \\ \text{s.t.} & \mathcal{E} = \sum_i a_i \mathcal{F}_i \end{cases} . \tag{3.100}$$

Through a standard trick, this optimization problem can be brought into the form of a linear program by introducing a slack variable $b \in \mathbb{R}^M$

$$\gamma_{S}(\mathcal{E}) = \begin{cases} \min_{a,b \in \mathbb{R}^{M}} & \sum_{i} b_{i} \\ \text{s.t.} & \mathcal{E} = \sum_{i} a_{i} \mathcal{F}_{i} \text{ and } \forall i : -a_{i} \leq b_{i}, a_{i} \leq b_{i} \end{cases}$$
(3.101)

as the optimizer (a^*, b^*) must clearly fulfill $b_i^* = |a_i^*|$. Equation (3.101) can be solved numerically through the usage of readily-available linear program solvers.

The smoothed quasiprobability extent of a channel $\mathcal{E} \in \operatorname{CPTP}(A \to B)$ w.r.t. a finite decomposition set

$$\gamma_{S}^{\epsilon}(\mathcal{E}) = \begin{cases}
\min_{a \in \mathbb{R}^{M}} & \|a\|_{1} \\
\text{s.t.} & \frac{1}{2} \|\mathcal{E} - \sum_{i} a_{i} \mathcal{F}_{i}\|_{\Diamond} \leq \epsilon \\
& \sum_{i} a_{i} J_{\mathcal{F}_{i}} \geq 0, \\
\text{tr}_{B} \left[\sum_{i} a_{i} J_{\mathcal{F}_{i}}\right] = \frac{\mathbb{1}_{A}}{\dim(A)}
\end{cases}$$
(3.102)

can also straightforwardly be cast into a convex program, because the diamond norm can be expressed as a SDP [173], [174]. For this purpose, we use following formulation of the channel distinguishability.

Proposition 3.49: [173]

Let $\Delta := \mathcal{G}_1 - \mathcal{G}_2$ be the difference of two quantum channels $\mathcal{G}_1, \mathcal{G}_2 \in \operatorname{CPTP}(A \to B)$. Then the diamond norm of Δ is expressible as the semidefinite program

$$\|\Delta\|_{\Diamond} = \begin{cases} \min_{\lambda \in \mathbb{R}, X \in P(AB)} & 2\lambda \\ \text{s.t.} & \lambda \mathbb{1}_{A} \succcurlyeq \operatorname{tr}_{B}[X], \\ & X \succcurlyeq J_{\Delta} \operatorname{dim}(A) \end{cases}$$
(3.103)

As such, we obtain the semidefinite program

$$\gamma_{S}^{\epsilon}(\mathcal{E}) = \begin{cases}
\min_{a \in \mathbb{R}^{M}, \lambda \in \mathbb{R}, X \in P(AB)} & \|a\|_{1} \\
\text{s.t.} & \lambda \leq \epsilon, \\
\lambda \mathbb{1}_{A} \succcurlyeq \operatorname{tr}_{B}[X], \\
X \succcurlyeq \operatorname{dim}(A)(J_{\mathcal{E}} - \sum_{i} a_{i}J_{\mathcal{F}_{i}}), \\
\sum_{i} a_{i}J_{\mathcal{F}_{i}} \succcurlyeq 0, \\
\operatorname{tr}_{B}\left[\sum_{i} a_{i}J_{\mathcal{F}_{i}}\right] = \frac{\mathbb{1}_{A}}{\operatorname{dim}(A)}
\end{cases} (3.104)$$

Another situation that can occur in practice is that we are given a maximal *budget* for the sampling overhead, i.e., a maximum allowed 1-norm $\kappa \geq 0$ for the QPD. Under this restriction, the approximation error of the best possible QPD is given by

$$\begin{cases}
\min_{a \in \mathbb{R}^{M}} & \|\mathcal{E} - \sum_{i} a_{i} \mathcal{F}_{i}\|_{\diamond} \\
\text{s.t.} & \sum_{i} a_{i} \mathcal{F}_{i} \in \text{CPTP}(A \to B), \\
& \sum_{i} |a_{i}| \leq \kappa.
\end{cases} (3.105)$$

which again can be turned into a semidefinite program using Proposition 3.49 and by introducing a slack variable

$$\min_{\substack{a,b \in \mathbb{R}^{M}, \lambda \in \mathbb{R}, X \in P(AB)}} 2\lambda
\text{s.t.} \quad \sum_{i} b_{i} \leq \kappa,
-a_{i} \leq b_{i}, a_{i} \leq b_{i},
\sum_{i} a_{i} J_{\mathcal{F}_{i}} \geq 0,
\sum_{i} a_{i} \operatorname{tr}_{B}[J_{\mathcal{F}_{i}}] = \frac{\mathbb{1}_{A}}{\dim(A)},
\lambda \mathbb{1}_{A} \geq \operatorname{tr}_{B}[X],
X \geq \dim(A)(J_{\mathcal{E}} - \sum_{i} a_{i} J_{\mathcal{F}_{i}})$$
(3.106)

Note that the way how we set up the optimization problems in Equation (3.102) and Equation (3.105), we explicitly required the approximation $\sum_i a_i \mathcal{F}_i$ of \mathcal{E} to be a physical CPTP map. This constraint could also be dropped to allow for potentially cheaper or better approximations, respectively.

3.7 Further reading

QPS was originally introduced by Pashayan, Wallman and Bartlett in the context of classical simulation of non-Clifford quantum circuits [120]. However, this original work was not based on general QPDs, but rather on the discrete Wigner representation for odd-dimensional qudit systems (and more general quasiprobability representations w.r.t. some frame and dual frame). Later work by Howard and Campbell [86] generalized the algorithm to the qubit setting and applied it to the QPS of Clifford+T circuits by simulating magic states with stabilizer states. General QPS of non-Clifford channels was first treated in Ref. [14].

Initially, the utility of classical side information was not considered in any of the applications of QPS, be that the simulation of non-Clifford channels [14], error mitigation [154] or nonphysical simulation [88], [131]. To our knowledge, the first inclusions of classical side information are due to Endo *et al.* [57] in the context of error mitigation, as well as Regula *et al.* [136] in the context on nonphysical simulation. However, both of these works only considered augmenting the decomposition set with completely positive trace-non-increasing maps. This corresponds to using classical side-information, but with only non-negative weights β_j of the individual quantum instruments. Only later work by Mitarai and Fujii [111], [112] on circuit knitting realized that negative weights were also admissible, though they still didn't

properly formulate and characterize the full associated decomposition set. This was done in later work by Piveteau *et al.* [129] and Lowe *et al.* [106], though both lack the quantum instrument formulation presented here. The full usage of classical side information, including negative weights and a formalism based on quantum instruments, was later also applied to the setting of nonphysical simulation in Ref. [182]. The authors consider the decomposition set of *twisted channels* of the form $\sum_i \beta_i \mathcal{E}_i$ where $\beta_i \in \{+1, -1\}$ and $(\mathcal{E}_i)_i$ is a general quantum instrument.

For an overview of the history of each individual application of QPS, we refer to the "further reading" section of the corresponding chapter.

The (total) robustness was first introduced in the context of entanglement theory by Vidal and Tarrach [166] and later generalized to the context of magic theory [86], nonphysical operations [88], [136] and probabilistic error cancellation [153]. For a comprehensive mathematical treatment of convex resource theories and associated robustness measures, we refer to reader to Ref. [135].

The linear program for the optimal quasiprobability extent w.r.t. a finite decomposition set is due to Ref. [86], [154]. The SDP for the optimal approximation with budgeted 1-norm is due to Piveteau *et al.* [131].

3.8 Contributions

We briefly summarize the author's contributions in this chapter.

- The exposition of QPS in Sections 3.1 and 3.2 summarizes a standard technique in literature and is devoid of new contributions besides a few remarks.
- The framework in Section 3.3, together with results in subsequent sections, is the first comprehensive treatment of classical side information in general QPS and has not been published before. We note that the idea of basing the decomposition set on quantum instruments was proposed in Ref. [182].
- In Section 3.4, the results on absolutely convex decomposition sets induced by quantum instruments (see Lemmas 3.13, 3.14, 3.18 and 3.26) are novel and previously unpublished. Similarly, the Lipschitz continuity result (Lemma 3.23) is new to our knowledge (though it might possibly be related to results in Ref. [139]). We also highlight that the basis in Example 3.24 was found by Thomas Dubach in his master thesis [53] which was supervised by the author.
 - The remaining results in the section are mostly covered explicitly or implicitly by other authors in previous literature.
- Almost all results in Section 3.5 are new and previously unpublished. Though it should be noted that many of them are adaptations of previously know properties of the robustness to the decomposition set \mathfrak{O}^* (as opposed to \mathfrak{O}). We note the results in Proposition 3.39 and Corollary 3.40 are well-established in QRT literature. We also note that the regularized quasiprobability extent has been previously studied [79], [129], but the asymptotic quasiprobability extent has to our knowledge never been considered before.

• The linear program in Section 3.6 is due to Ref. [86], [154]. The semidefinite program for a budgeted QPD was introduced by the author in Ref. [131]. The semidefinite program for the smooth quasiprobability extent has previously not been published.

Chapter 4

Simulating non-physical computation

The set of completely positive trace-preserving (CPTP) superoperators describes all physically admissible ways in which a quantum system can evolve in time. Nonetheless, in various fields of quantum information, the necessity to study non-CPTP maps occasionally arises. For instance, positive non-CP maps play an important role in entanglement detection [124]. Other such examples include the study of non-Markovian dynamics of open quantum systems [121], [143] and quantum error mitigation [88], [154].

This chapter is dedicated to the study of the QPS overhead for simulating non-CPTP maps given one has access to a quantum computer that is capable of running any CPTP operation. From the point of view of resource theories (as discussed in Section 3.5), we will essentially consider the most powerful QRT where the free channels from A to B are precisely all CPTP maps $\mathfrak{D}(A \to B) \coloneqq \text{CPTP}(A \to B)$ and the set of free states on A is the set of all density operators D(A). Clearly, the quasiprobability extent of any physical quantum channel or quantum state is 1, so our focus purely lies on the study of the quasiprobability extent of non-physical (super-)operators. More concretely, we will consider the quasiprobability extent γ_{CPTP^*} w.r.t. the decomposition set $\text{CPTP}^*(A \to B) := \text{chan}^*[QI(A \to B)]$ which captures the utilization of classical side information through intermediate measurements (recall the discussion in Section 3.3 and Equation (3.75)) and how it compares to the quasiprobability extent γ_{CPTP} where intermediate measurements are not considered.

The utility of studying this setting is twofold. On one hand, this "maximal resource theory" can be a useful technical tool to bound the behavior of QPS in other resource theories. For example, for any QRT \mathfrak{D} , $\gamma_{\mathfrak{D}^*}$ is lower bounded by γ_{CPTP^*} due to Lemma 3.10. On the other hand, there are direct practical applications involving the simulation of quantum circuits containing non-physical operations. We will touch upon a few of these applications towards the end of the chapter.

4.1 Utility of classical side information

In this section, we aim to study the difference between $\gamma_{\text{CPTP}^{\star}}$ and γ_{CPTP} in order to evaluate how useful it is to consider classical side information in this setting. We start by mathematically characterizing the decomposition sets. Clearly, by definition the set of all quantum instruments $\text{QI}(A \to B)$ is coarse-grainable, trivially fine-grainable and closed under mixture. So by Lemma 3.9 we can characterize

$$CPTP^{\star}(A \to B) = \{ \mathcal{E}^{+} - \mathcal{E}^{-} | \mathcal{E}^{\pm} \in CP(A \to B), \mathcal{E}^{+} + \mathcal{E}^{-} \in CPTP(A \to B) \}. \quad (4.1)$$

In fact, the trace-preserving condition can be relaxed to trace-non-increasing

$$CPTP^{\star}(A \to B) = \{ \mathcal{E}^{+} - \mathcal{E}^{-} | \mathcal{E}^{\pm} \in CP(A \to B), \mathcal{E}^{+} + \mathcal{E}^{-} \in CPTN(A \to B) \}$$
 (4.2)

since any $\mathcal{E}^+ + \mathcal{E}^- \in \text{CPTN}(A \to B)$ can be completed to a trace-preserving map $(\mathcal{E}^+ + \mathcal{E}^-) + \mathcal{G} \in \text{CPTP}(A \to B)$ and hence $\mathcal{E}^+ - \mathcal{E}^- = (\mathcal{E}^+ + \frac{1}{2}\mathcal{G}) - (\mathcal{E}^- + \frac{1}{2}\mathcal{G})$.

We now turn to the question of which superoperators $\mathcal{E} \in HP(A \to B)$ possess a QPD into the sets $CPTP(A \to B)$ and $CPTP^*(A \to B)$.

Lemma 4.1: For any systems *A* and *B*,

$$\operatorname{span}_{\mathbb{R}}\left(\operatorname{CPTP}(A \to B)\right) = \operatorname{\mathbb{R}HPTP}(A \to B) \coloneqq \left\{r \cdot \mathcal{E} \middle| r \in \mathbb{R}, \mathcal{E} \in \operatorname{HPTP}(A \to B)\right\} \tag{4.3}$$

and

$$\operatorname{span}_{\mathbb{R}}\left(\operatorname{CPTP}^{\star}(A \to B)\right) = \operatorname{HP}(A \to B). \tag{4.4}$$

Proof. We start by proving Equation (4.3). The inclusion \subset follows directly from the fact that any linear combination of trace-preserving maps is itself proportional to a trace-preserving map. Conversely, consider some $r \in \mathbb{R}$ and $\mathcal{E} \in \operatorname{HPTP}(A \to B)$. We will now find a decomposition of $r\mathcal{E}$ in terms of CPTP maps. We can assume without loss of generality that r=1, otherwise the decomposition just needs to be scaled accordingly. Let's denote by $d_A := \dim(A)$, $d_B := \dim(B)$ and by $\mathcal{D}(\rho) := \frac{\operatorname{tr}[\rho]}{d_B} \mathbb{1}_B$ the fully depolarizing channel on B. Furthermore, define $\lambda \geq \sigma_{\max}(J_{\mathcal{E}})$ to be at least the maximum eigenvalue of $J_{\mathcal{E}}$. We can decompose

$$\mathcal{E} = d_A d_B \lambda \cdot \mathcal{D} - (d_A d_B \lambda - 1) \cdot \left(\frac{d_A d_B \lambda}{d_A d_B \lambda - 1} \mathcal{D} - \frac{1}{d_A d_B \lambda - 1} \mathcal{E} \right). \tag{4.5}$$

It remains to show that $\left(\frac{d_A d_B \lambda}{d_A d_B \lambda - 1} \mathcal{D} - \frac{1}{d_A d_B \lambda - 1} \mathcal{E}\right)$ is CPTP. This is straightforward to check as

$$\operatorname{tr}_{B}\left[\frac{d_{A}d_{B}\lambda}{d_{A}d_{B}\lambda-1}J_{\mathcal{D}}-\frac{1}{d_{A}d_{B}\lambda-1}J_{\mathcal{E}}\right]=\left(\frac{d_{A}d_{B}\lambda}{d_{A}d_{B}\lambda-1}-\frac{1}{d_{A}d_{B}\lambda-1}\right)\frac{1}{d_{A}}\mathbb{1}_{A}=\frac{1}{d_{A}}\mathbb{1}_{A}$$
(4.6)

and

$$\left(\frac{d_A d_B \lambda}{d_A d_B \lambda - 1} J_{\mathcal{D}} - \frac{1}{d_A d_B \lambda - 1} J_{\mathcal{E}}\right) \geq 0 \tag{4.7}$$

$$\iff d_A d_B \lambda \frac{1}{d_A d_B} \mathbb{1}_{AB} \succeq J_{\mathcal{E}}$$

$$\tag{4.8}$$

$$\iff \lambda \ge \sigma_{\max}(J_{\mathcal{E}}).$$
 (4.9)

We now turn to proving Equation (4.4). The direction \subset again directly follows from the fact that any linear combination of Hermitian-preserving maps is again Hermitian-preserving. For the converse direction, consider a superoperator $\mathcal{E} \in \mathrm{HP}(A \to B)$. We can split up its Choi representation into the difference of two positive operators $\Lambda^{\pm} \succcurlyeq 0$

$$J_{\mathcal{E}} = \Lambda^{+} - \Lambda^{-} = \lambda^{+} \frac{\Lambda^{+}}{\lambda^{+}} - \lambda^{-} \frac{\Lambda^{-}}{\lambda^{-}}$$
(4.10)

where we introduced the maximal eigenvalues λ^{\pm} of $\dim(A)\operatorname{tr}_B[\Lambda^{\pm}]$. We now observe that $\Lambda^{\pm}/\lambda^{\pm}$ are Choi operators representing completely positive trace-non-increasing maps, as they are positive and $\operatorname{tr}_B[\frac{\Lambda^{\pm}}{\lambda^{\pm}}] \preccurlyeq \frac{1}{\dim(A)} \mathbb{1}_A$. The proof follows from $\operatorname{CPTN}(A \to B) \subset \operatorname{CPTP}^*(A \to B)$ which is a consequence of Equation (4.2). Note that we implicitly assumed $\lambda^{\pm} > 0$, but the proof can be straightforwardly adapted for the case where $J_{\mathcal{E}}$ is positive or negative.

Lemma 4.1 implies that classical side information is crucial in order to simulate Hermitian-preserving maps that do not lie in $\mathbb{R}HPTP(A \to B)$. It is natural to ask, whether using intermediate measurements improves the quasiprobability extent for maps that do lie in $\mathbb{R}HPTP(A \to B)$. The answer to this question is negative.

Proposition 4.2

For all $\mathcal{E} \in \mathbb{R}HPTP(A \to B)$ one has $\gamma_{CPTP}(\mathcal{E}) = \gamma_{CPTP^*}(\mathcal{E})$.

Proof. The proof consists of showing that any QPD of $\mathcal E$ with respect to CPTP* can be reformulated as a QPD with respect to CPTP without increasing the 1-norm of the coefficients. By Lemmas 3.13 and 3.14 we can without loss of generality assume that the QPD is of the form $\mathcal E=\kappa\mathcal F$ for some $\mathcal F\in \operatorname{CPTP}^*(A\to B)$. By Equation (4.1), we can write $\mathcal F=\mathcal F^+-\mathcal F^-$ for $\mathcal F^\pm\in\operatorname{CP}(A\to B)$ such that $\mathcal F^++\mathcal F^-\in\operatorname{CPTP}(A\to B)$. If either $\mathcal F^+=0$ or $\mathcal F^-=0$, then we already have a QPD into $\operatorname{CPTP}(A\to B)$. Otherwise, observe that both $\mathcal F^++\mathcal F^-$ and $\mathcal F^+-\mathcal F^-$ are in $\operatorname{RHPTP}(A\to B)$. Clearly, this is only possible if both $\mathcal F^+$ and $\mathcal F^-$ are themselves in $\operatorname{RHPTP}(A\to B)$. The following is therefore a valid QPD of $\mathcal E$ w.r.t. $\operatorname{CPTP}(A\to B)$

$$\mathcal{E} = \kappa \operatorname{tr}[J_{\mathcal{F}^{+}}] \frac{\mathcal{F}^{+}}{\operatorname{tr}[J_{\mathcal{F}^{+}}]} - \kappa \operatorname{tr}[J_{\mathcal{F}^{-}}] \frac{\mathcal{F}^{-}}{\operatorname{tr}[J_{\mathcal{F}^{-}}]}$$
(4.11)

and the 1-norm of its coefficients is given by $\kappa(\operatorname{tr}[J_{\mathcal{F}^+}] + \operatorname{tr}[J_{\mathcal{F}^-}]) = \kappa$.

4.2 Characterization of quasiprobability extent

In this section, we will derive a simple characterization of $\gamma_{\text{CPTP}^*}(\mathcal{E})$ for Hermitian-preserving operations \mathcal{E} . Recall that since CPTP and CPTP* are convex and absolutely convex (see Lemma 3.13), we can restrict ourselves to QPDs with one and two elements respectively by Lemmas 3.11 and 3.14. Together with the characterization of CPTP* in Equation (4.1), we can write the quasiprobability extent of some $\mathcal{E} \in \mathbb{R}HPTP(A \to B)$ as

$$\gamma_{\text{CPTP}}(\mathcal{E}) = \begin{cases} \min_{a^{\pm} \ge 0, \Lambda^{\pm} \in P(AB)} & a^{+} + a^{-} \\ \text{s.t.} & J_{\mathcal{E}} = a^{+} \Lambda^{+} - a^{-} \Lambda^{-} \\ & \text{tr}_{B} \Lambda^{\pm} = \frac{1}{\dim(A)} \mathbb{1}_{A} \end{cases}$$
(4.12)

and similarly for $\mathcal{E} \in \mathrm{HP}(A \to B)$ we have

$$\gamma_{\text{CPTP}^{\star}}(\mathcal{E}) = \begin{cases} \min_{\kappa \geq 0, \Lambda^{\pm} \in P(AB)} & \kappa \\ \text{s.t.} & J_{\mathcal{E}} = \kappa(\Lambda^{+} - \Lambda^{-}) \\ & \text{tr}_{B}[\Lambda^{+} + \Lambda^{-}] = \frac{1}{\dim(A)} \mathbb{1}_{A} \end{cases}$$
(4.13)

Note that we used a minimum instead of an infimum since both CPTP and CPTP* are closed (recall Lemma 3.25). Using a simple variable substitution $a^{\pm}\Lambda^{\pm} \to \Lambda^{\pm}$, respectively $\kappa\Lambda^{\pm} \to \Lambda^{\pm}$, the constraints in the above two optimization problems can be recast into the form of linear matrix equalities

$$\gamma_{\text{CPTP}}(\mathcal{E}) = \begin{cases} \min_{a^{\pm} \ge 0, \Lambda^{\pm} \in P(AB)} & a^{+} + a^{-} \\ \text{s.t.} & J_{\mathcal{E}} = \Lambda^{+} - \Lambda^{-} \\ & \text{tr}_{B} \Lambda^{\pm} = \frac{a^{\pm}}{\dim(A)} \mathbb{1}_{A} \end{cases}$$
(4.14)

$$\gamma_{\text{CPTP}^{\star}}(\mathcal{E}) = \begin{cases} \min_{\kappa \geq 0, \Lambda^{\pm} \in P(AB)} & \kappa \\ \text{s.t.} & J_{\mathcal{E}} = \Lambda^{+} - \Lambda^{-} \\ & \text{tr}_{B}[\Lambda^{+} + \Lambda^{-}] = \frac{\kappa}{\dim(A)} \mathbb{1}_{A} \end{cases} . \tag{4.15}$$

The resulting optimization problems are now semidefinite programs, which will prove to be very useful. On one hand, it allows for the efficient numerical evaluation of the quasiprobability extent, at least for small Hilbert spaces. On the other hand, it allows us to prove various properties of the optimal solution analytically. We will usually not have the same luxury for other resource theories.

Is will be useful to determine the dual forms to the above SDPs.

Lemma 4.3: The following SDPs are valid dual SDPs to Equations (4.14) and (4.15):

$$\gamma_{\text{CPTP}}(\mathcal{E}) = \begin{cases} \max_{X \in \mathcal{H}(AB), Y^{\pm} \in \mathcal{H}(A)} & \text{tr}[J_{\mathcal{E}}X] \\ \text{s.t.} & -Y^{-} \otimes \mathbb{1}_{B} \preccurlyeq X \preccurlyeq Y^{+} \otimes \mathbb{1}_{B} \\ & \text{tr}[Y^{\pm}] = \dim(A) \end{cases}$$
(4.16)

$$\gamma_{\mathrm{CPTP}^{\star}}(\mathcal{E}) = \begin{cases} \max_{X \in \mathrm{H}(AB), Y \in \mathrm{H}(A)} & \mathrm{tr}[J_{\mathcal{E}}X] \\ \mathrm{s.t.} & -Y \otimes \mathbb{1}_{B} \preccurlyeq X \preccurlyeq Y \otimes \mathbb{1}_{B} \\ & \mathrm{tr}[Y] = \dim(A) \end{cases}$$
 (4.17)

Proof. In the following, we will denote the unnormalized Hilbert-Schmidt inner product by $\langle A, B \rangle := \text{tr}[A^{\dagger}B]$. We start with the SDP in Equation (4.14). The Lagrangian in this case is given by

$$L(a^{\pm}, \Lambda^{\pm}; X, Y^{\pm}) = (a^{+} + a^{-}) - \langle X, J_{\mathcal{E}} - \Lambda^{+} + \Lambda^{-} \rangle$$
$$- \langle Y^{+}, \operatorname{tr}_{B}[\Lambda^{+}] - \frac{a^{+}}{\dim(A)} \mathbb{1}_{A} \rangle - \langle Y^{-}, \operatorname{tr}_{B}[\Lambda^{-}] - \frac{a^{-}}{\dim(A)} \mathbb{1}_{A} \rangle.$$
(4.18)

Using the relation $\langle R_A, \operatorname{tr}_B[S_{AB}] \rangle = \langle R_A \otimes \mathbb{1}_B, S_{AB} \rangle$, we can rewrite this as

$$L(a^{\pm}, \Lambda^{\pm}; X, Y^{\pm}) = -\langle X, J_{\mathcal{E}} \rangle + a^{+} \left(1 + \frac{1}{\dim(A)} \operatorname{tr}[Y^{+}]\right) + a^{-} \left(1 + \frac{1}{\dim(A)} \operatorname{tr}[Y^{-}]\right) + \langle \Lambda^{+}, X - Y^{+} \otimes \mathbb{1}_{B} \rangle + \langle \Lambda^{-}, -X - Y^{-} \otimes \mathbb{1}_{B} \rangle.$$
(4.19)

The dual function $g(X, Y^{\pm}) := \inf_{a^{\pm}, \Lambda^{\pm}} L(a^{\pm}, \Lambda^{\pm}; X, Y^{\pm})$ can clearly only be greater than $-\infty$ if $\pm X \succcurlyeq Y^{\pm} \otimes \mathbb{1}_{B}$ and $\operatorname{tr}[Y^{\pm}] \geq -\dim(A)$. We thus get the dual SDP

$$\gamma_{\text{CPTP}}(\mathcal{E}) = \begin{cases} \max_{X \in \mathcal{H}(AB), Y^{\pm} \in \mathcal{H}(A)} & -\langle X, J_{\mathcal{E}} \rangle \\ \text{s.t.} & Y^{+} \otimes \mathbb{1}_{B} \preccurlyeq X \preccurlyeq -Y^{-} \otimes \mathbb{1}_{B} \\ & \text{tr}[Y^{\pm}] \geq -\dim(A) \end{cases}$$
(4.20)

The desired form of this SDP can be obtained by substituting X, Y^+, Y^- by $-X, -Y^+, -Y^-$ and by realizing that the inequality in the trace can without loss of generality be assumed to be achieved.

We now turn to the second SDP in Equation (4.14). The Lagrangian here is given by

$$L(\kappa, \Lambda^{\pm}; X, Y) = \kappa - \langle X, J_{\mathcal{E}} - \Lambda^{+} + \Lambda^{-} \rangle$$

$$- \langle Y, \operatorname{tr}_{B}[\Lambda^{+} + \Lambda^{-}] - \kappa \frac{1}{\dim(A)} \mathbb{1}_{A} \rangle \qquad (4.21)$$

$$= -\langle X, J_{\mathcal{E}} \rangle + \kappa (1 + \frac{1}{\dim(A)} \operatorname{tr}[Y])$$

$$+ \langle \Lambda^{+}, X - Y \otimes \mathbb{1}_{B} \rangle + \langle \Lambda^{-}, -X - Y \otimes \mathbb{1}_{B} \rangle. \qquad (4.22)$$

By the same argument as before, we get a dual SDP

$$\gamma_{\text{CPTP}^{\star}}(\mathcal{E}) = \begin{cases} \max_{X \in \text{H}(AB), Y \in \text{H}(A)} & -\langle X, J_{\mathcal{E}} \rangle \\ \text{s.t.} & -Y \otimes \mathbb{1}_{B} \preccurlyeq X \preccurlyeq Y \otimes \mathbb{1}_{B} \\ & \text{tr}[Y] \geq -\dim(A) \end{cases}$$
(4.23)

Again, with a suitable substitution and realizing the inequality $tr[Y] \le dim(A)$ is achieved, we get the desired result.

In fact, it turns out that the above semidefinite program for γ_{CPTP^*} is equivalent to the well-known diamond norm.

Theorem 4.4: [182, Theorem 3]

For any
$$\mathcal{E} \in HP(A \to B)$$
 one has $\gamma_{CPTP^*}(\mathcal{E}) = ||\mathcal{E}||_{\Diamond}$.

In turn, this also means that $\gamma_{\text{CPTP}}(\mathcal{E}) = \|\mathcal{E}\|_{\Diamond}$ for any $\mathcal{E} \in \mathbb{R}\text{HPTP}(A \to B)$ by Proposition 4.2. The standard operational interpretation of the diamond norm is typically stated in terms of the single-shot distinguishability of two quantum channels. Our result can be considered an alternative and new operational interpretation: The diamond norm measures the non-physicalness of a certain operation by quantifying the cost of simulating it using physical operations.

When specialized to states, Theorem 4.4 implies

$$\forall \omega \in \mathcal{H}(A) : \gamma_{\mathcal{D}^{\star}(A)}(\omega) = \|\omega\|_{1}. \tag{4.24}$$

This is perhaps not too surprising: By Lemma 3.29 we know that $\gamma_{D^*} = \gamma_D$ and clearly the optimal QPD w.r.t. D(A) of ω is given by separating its positive and negative eigenspace projectors.

Remarkably, Theorem 4.4 thus implies that the quasiprobability extent of a super-operator $\gamma_{\text{CPTP}^{\star}}(\mathcal{E})$ is equal to the largest extent $\gamma_{\text{D}}(\mathcal{E}(\rho))$ optimized over all Hermitian operators ρ (on a possibly extended Hilbert space) with quasiprobability extent $\gamma_{\text{D}}(\rho) \leq 1$.

Proof. This proof essentially follows [136, Lemma 1] which could also be derived from [87, Lemma 4, Theorem 2]. We start with the dual formulation of $\gamma_{\text{CPTP}^{\star}}$ from Lemma 4.3. In the following, we will denote the unnormalized Hilbert-Schmidt inner product by $\langle A, B \rangle := \text{tr}[A^{\dagger}B]$.

$$\gamma_{\operatorname{CPTP}^{\star}}(\mathcal{E}) = \max\{\langle J_{\mathcal{E}}, X \rangle | -Y \otimes \mathbb{1}_{\mathcal{B}} \preccurlyeq X \preccurlyeq Y \otimes \mathbb{1}_{\mathcal{B}}, \operatorname{tr}[Y] = \dim(A)\}$$
 (4.25)

$$= \max\{\langle J_{\mathcal{E}}, X \rangle | -Y \otimes \mathbb{1}_{B} \preccurlyeq \frac{X}{\dim(A)} \preccurlyeq Y \otimes \mathbb{1}_{B}, Y \in D(A)\}$$
 (4.26)

$$= \sup\{\langle J_{\mathcal{E}}, X \rangle | -Y \otimes \mathbb{1}_{B} \preccurlyeq \frac{X}{\dim(A)} \preccurlyeq Y \otimes \mathbb{1}_{B}, Y \in D_{>0}(A)\}$$
(4.27)

$$= \sup\{\langle \sqrt{Y \otimes \mathbb{1}_B} J_{\mathcal{E}} \sqrt{Y \otimes \mathbb{1}_B}, X \rangle | \mathbb{1}_{AB} \preccurlyeq \frac{X}{\dim(A)} \preccurlyeq \mathbb{1}_{AB}, Y \in D_{>0}(A)\}$$

$$(4.28)$$

where $D_{>0}(A)$ denotes the set of full-rank density operators and the last equality used the variable substitution $X \mapsto \sqrt{Y^{-1} \otimes \mathbb{1}_B} X \sqrt{Y^{-1} \otimes \mathbb{1}_B}$. The trace norm of a Hermitian matrix M can be expressed as

$$||M||_1 = \max_{||N||_{\infty} \le 1} \operatorname{tr}[MN].$$
 (4.29)

Thus, we get

$$\gamma_{\text{CPTP}^{\star}}(\mathcal{E}) = \sup_{Y \in D_{>0}(A)} \dim(A) \left\| \sqrt{Y \otimes \mathbb{1}_B} J_{\mathcal{E}} \sqrt{Y \otimes \mathbb{1}_B} \right\|_1$$
(4.30)

$$= \max_{Y \in D(A)} \dim(A) \left\| \sqrt{Y \otimes \mathbb{1}_B} J_{\mathcal{E}} \sqrt{Y \otimes \mathbb{1}_B} \right\|_1$$
 (4.31)

$$= \max_{Y \in D(A)} \left\| (id_A \otimes \mathcal{E}) \left[\sqrt{Y} \otimes \mathbb{1}_{A'} | \Psi \rangle \langle \Psi | \sqrt{Y} \otimes \mathbb{1}_{A'} \right] \right\|_1$$
(4.32)

$$= \max_{Y \in D(A)} \left\| (id_A \otimes \mathcal{E}) \left[\mathbb{1}_A \otimes \sqrt{Y} |\Psi\rangle\langle\Psi| \, \mathbb{1}_A \otimes \sqrt{Y} \right] \right\|_1 \tag{4.33}$$

$$= \max_{\phi} \left\| (\mathrm{id}_A \otimes \mathcal{E})(\phi) \right\|_1 \tag{4.34}$$

$$= \max_{\rho \in \mathcal{D}(AA')} \left\| (\mathrm{id}_A \otimes \mathcal{E})(\rho) \right\|_1 \tag{4.35}$$

$$= \|\mathcal{E}\|_{\Diamond} \tag{4.36}$$

where we used $(\mathbb{1}_A \otimes Z_{A'}) | \Psi \rangle = (Z_A^T \otimes \mathbb{1}_{A'}) | \Psi \rangle$ for the unnormalized maximally entangled state $| \Psi \rangle_{AA'} = \sum_i |i\rangle_A \otimes |i\rangle_{A'}$ and the fact that any pure state can be seen as the canonical purification $(\mathbb{1} \otimes \sqrt{\rho}) | \Psi \rangle$ of some mixed state ρ .

We summarize the results from this section in Figure 4.1. Note that the diamond norm is multiplicative under the tensor product $\|\mathcal{E}\otimes\mathcal{F}\|_{\Diamond}=\|\mathcal{E}\|_{\Diamond}\|\mathcal{F}\|_{\Diamond}$, see for example Ref. [175, Theorem 3.49]. As such, the regularized quasiprobability extent remains unchanged $\overline{\gamma}_{\text{CPTP}^*}=\gamma_{\text{CPTP}^*}$.

4.3 Applications

In this section, we investigate a few practical applications of QPS of non-physical operations. Before we get to the applications themselves, we briefly introduce a technical result on the quasiprobability extent of superoperators with diagonal Pauli transfer matrices. In the following, we will denote by $\{Q_1, \ldots, Q_{4^n}\}$ the Pauli strings $\{1, X, Y, Z\}^{\otimes n}$ endowed with some arbitrary order. Furthermore, $[\![Q_\alpha, Q_\beta]\!]$ denotes the symplectic inner product which is = 0 if Q_α and Q_β commute and = 1 if they anti-commute.

Definition 4.5: The *n*-qubit symplectic Walsh-Hadamard transformation is the linear map $\mathbf{W}_n : \mathbb{C}^{4^n} \to \mathbb{C}^{4^n}$ defined by

$$(\mathbf{W}_n(x))_i := \sum_{j=1}^{4^n} (-1)^{[Q_i,Q_j]} x_j.$$
 (4.37)

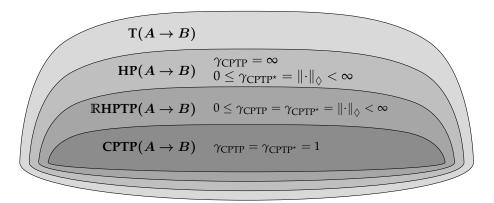


Figure 4.1: Set diagram representing the relations between the sets of linear superoperators, Hermitian-preserving maps, maps proportional to a trace-preserving map and quantum channels. For each shaded region, the possible ranges of values taken by the quasiprobability extents γ_{CPTP} and γ_{CPTP^*} are depicted.

By definition, the Walsh-Hadamard transform is a linear and symmetric operator. Furthermore, it is invertible, and its inverse is proportional to itself $\mathbf{W}_n^2 = 4^n \mathbb{1}$ which can be readily verified from its definition

$$\left(\mathbf{W}_{n}^{2}(x)\right)_{i} = \sum_{j,k} (-1)^{[[Q_{i},Q_{j}]]} (-1)^{[[Q_{j},Q_{k}]]} x_{k}$$
(4.38)

$$= \sum_{j,k} (-1)^{[[Q_i Q_k, Q_j]]} x_k$$

$$= \sum_{k} 4^n \delta_{i,k} x_k$$
(4.39)

$$=\sum_{k}4^{n}\delta_{i,k}x_{k}\tag{4.40}$$

$$=4^{n}x_{i} \tag{4.41}$$

where we used that any Pauli string (except the identity) precisely commutes with half of all Pauli strings. We refer readers interested in an exposition with more mathematical details to Ref. [59].

Proposition 4.6

Let $\mathcal{E} \in \mathrm{HP}(A)$ be an n-qubit superoperator with diagonal Pauli transfer matrix representation $M = \text{diag}(m_1, \dots, m_{4^n})$ for some $m \in \mathbb{R}^{4^n}$. Then

$$\gamma_{\text{CPTP}^{\star}}(\mathcal{E}) = \gamma_{\mathcal{P}_n}(\mathcal{E}) = \frac{1}{4^n} \|\mathbf{W}_n(m)\|_1$$
 (4.42)

where \mathcal{P}_n denotes the *n*-qubit Pauli channels.

Proof. Consider an *n*-qubit superoperator of the form

$$\mathcal{F}(\rho) := \sum_{i} r_i Q_i \rho Q_i \tag{4.43}$$

for some $r \in \mathbb{R}^{4^n}$. One can show that $\|\mathcal{F}\|_{\Diamond} = \|r\|_1$ by using the triangle inequality

$$\|\mathcal{F}\|_{\Diamond} \le \sum_{i} |r|_{i} \|\text{Ad}[Q_{i}]\|_{\Diamond} = \|r\|_{1}$$
 (4.44)

and conversely by using the definition of the diamond norm

$$\|\mathcal{F}\|_{\Diamond} \ge \|(\mathcal{F} \otimes \mathrm{id})(|\Psi\rangle\langle\Psi|)\|_{1}$$
 (4.45)

$$= \left\| \sum_{i} r_{i}(Q_{i} \otimes \mathbb{1}) |\Psi\rangle\langle\Psi| (Q_{i} \otimes \mathbb{1}) \right\|_{1} \tag{4.46}$$

$$= \sum_{i} |r_{i}| \|(Q_{i} \otimes \mathbb{1}) |\Psi\rangle\langle\Psi| (Q_{i} \otimes \mathbb{1}) \|_{1}$$

$$(4.47)$$

$$= \|r\|_1$$
 (4.48)

where $|\Psi\rangle$ is the maximally-entangled 2n-qubit state and we used the fact that the states $\{(Q_i \otimes 1) | \Psi\rangle\}_i$ form an orthonormal basis. So our desired statement follows from Theorem 4.4 if we can show that \mathcal{E} must be of the form of Equation (4.43) with $r = \mathbf{W}_n^{-1}(m) = \mathbf{W}_n(m)/4^n$.

To this end, we note that there is a one-to-one correspondence between superoperators of the form in Equation (4.43) and superoperators with diagonal Pauli transfer matrix. On one hand, consider $\mathcal F$ as in Equation (4.43). The associated Pauli transfer matrix reads

$$M_{\alpha,\beta}^{(\mathcal{F})} = \frac{1}{2^n} \operatorname{tr} \left[Q_{\alpha} \mathcal{F}(Q_{\beta}) \right] \tag{4.49}$$

$$= \frac{1}{2^n} \sum_{i} r_i \operatorname{tr} \left[Q_{\alpha} Q_i Q_{\beta} Q_i \right] \tag{4.50}$$

$$= \frac{1}{2^n} \sum_{i} r_i (-1)^{\llbracket Q_i, Q_\beta \rrbracket} \operatorname{tr} \left[Q_\alpha Q_\beta \right]$$
 (4.51)

$$= \delta_{\alpha,\beta} \sum_{i} r_i (-1)^{\llbracket Q_i, Q_\beta \rrbracket} \tag{4.52}$$

and the diagonal elements are precisely given by $\mathbf{W}_n(r)$. Conversely, thanks to the invertibility of the Walsh-Hadamard transformation, for every superoperator with diagonal Pauli transfer matrix, one can find a vector r such that it can be written in the form of Equation (4.43).

We now present a few applications of non-physical QPS.

Noise inversion Consider following situation: In some part of a quantum computation, a subset of the qubits, denoted by system A, undergoes some noise described by a quantum channel $\mathcal{N} \in \operatorname{CPTP}(A)$. For example, at this point of the circuit the qubits could be sent from one computational device to another through a faulty transmission channel. To undo the effect of the noise, one would ideally perform the operation

 \mathcal{N}^{-1} to reverse the effect $\mathcal{N}^{-1} \circ \mathcal{N}(\rho) = \rho$ for any arbitrary input state ρ^1 . Unfortunately, \mathcal{N}^{-1} is not in CPTP(A) unless \mathcal{N} is unitary, making it impossible to fully recover the quantum information in physically realistic scenarios. Using QPS, we can however simulate the channel \mathcal{N}^{-1} with a sampling overhead of $\gamma_{\text{CPTP}^*}(\mathcal{N}^{-1})$.

First, we note that the inverse of a CPTP map is itself trace-preserving.

Lemma 4.7: Let $\mathcal{N} \in \operatorname{CPTP}(A)$ be an invertible quantum channel. Then \mathcal{N}^{-1} is Hermitian-preserving and trace-preserving.

As such, the quasiprobability extent of \mathcal{N}^{-1} is always well-defined and

$$\gamma_{\text{CPTP}^{\star}}(\mathcal{N}^{-1}) = \gamma_{\text{CPTP}}(\mathcal{N}^{-1}) = \left\| \mathcal{N}^{-1} \right\|_{\Diamond}. \tag{4.53}$$

Proof. The only nontrivial step is to show that a Hermitian-preserving superoperator cannot map a non-Hermitian map onto a Hermitian map. An arbitrary operator $O \in L(A)$ can always be uniquely decomposed into a Hermitian and skew-Hermitian part O = X + iY where $X, Y \in H(A)$. By linearity, a Hermitian-preserving superoperator $\mathcal{E} \in HP(A)$ will preserve this decomposition

$$O \mapsto \mathcal{E}(X) + i\mathcal{E}(Y)$$
 (4.54)

as $\mathcal{E}(X)$ and $i\mathcal{E}(Y)$ are still Hermitian and skew-Hermitian respectively. Due to the invertibility of \mathcal{E} we have $\mathcal{E}(Y) = 0 \Rightarrow Y = 0$.

Example 4.8. The one-qubit depolarizing channel

$$\mathcal{N}(\rho) = (1 - p)\rho + p \operatorname{tr}[\rho] \frac{1}{2}$$
(4.55)

has the Pauli transfer matrix representation² diag(1, 1 - p, 1 - p, 1 - p). By using the inverse Walsh-Hadamard transformation we obtain

$$\mathcal{N}^{-1} = \frac{4-p}{4(1-p)} id - \frac{p}{4(1-p)} Ad[X] - \frac{p}{4(1-p)} Ad[Y] - \frac{p}{4(1-p)} Ad[Z]. \quad (4.56)$$

By Proposition 4.6, the quasiprobability extent for the noise inversion is given by

$$\gamma_{\text{CPTP}^*}(\mathcal{N}^{-1}) = \frac{4-p}{4(1-p)} + 3\frac{p}{4(1-p)} = \frac{1+\frac{p}{2}}{1-p}.$$
 (4.57)

Note that the quasiprobability extent converges to 1 as p approaches zero. Conversely, towards $p \to 1$ the extent diverges, as \mathcal{N} is not invertible at p = 1.

 $^{^1}$ Note that here we make an implicit assumption that a (possibly non-CPTP) inverse of the channel $\mathcal N$ exists in the first place. This is not necessarily the case, as exemplified by choosing $\mathcal N$ to be the fully depolarizing channel or by choosing differing input and output spaces. For non-invertible channels one could utilize a pseudo-inverse like the Moore-Penrose inverse or the Drazin inverse. See [34] for more discussion about pseudoinverses in noise inversion.

²The Pauli order is chosen to be $\mathbb{1}$, X, Y, Z.

Example 4.9. Consider the one-qubit amplitude damping noise channel

$$\mathcal{N}(\rho) := K_0 \rho K_0^{\dagger} + K_1 \rho K_1^{\dagger} \quad \text{where} \quad K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, K_1 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}$$
(4.58)

and $p \in [0,1]$. The quasiprobability extent of its inverse can be shown to be

$$\gamma_{\text{CPTP}^*}(\mathcal{N}^{-1}) = \|\mathcal{N}^{-1}\|_{\Diamond} = \frac{1+p}{1-p}.$$
(4.59)

For a proof, we refer the reader to [136].

Entanglement detection Positive but not completely positive maps $\mathcal{E} \in HP(A)$ provide a simple criterion to detect entanglement in a bipartite state ρ_{AB} . It is easy to show that $(\mathcal{E} \otimes id_B)(\rho_{AB})$ can only contain negative eigenvalues if ρ_{AB} is entangled. The canonical example of such an entanglement criterion is provided by choosing \mathcal{E} to be the transpose map

$$T_A\left(\sum_{i,j}\alpha_{i,j}|b_i\rangle\langle b_j|\right) := \sum_{i,j}\alpha_{j,i}|b_i\rangle\langle b_j|. \tag{4.60}$$

which is defined w.r.t. some fixed orthonormal basis $\{|b_i\rangle\}_i$ of A.

While useful in theory, it is difficult to translate such an entanglement criterion into an experimentally implementable protocol to detect entanglement. Using QPS, we can address this problem by simulating the transpose map T_A and then estimate the smallest eigenvalue of the state. The associated sampling overhead is as follows.

Proposition 4.10

For any system *A* one has $\gamma_{\text{CPTP}^*}(T_A) = \dim(A)$.

Note that for qubit systems, T_A has a diagonal Pauli transfer matrix and hence the decomposition set can be chosen to only contain efficiently implementable Pauli channels (recall Proposition 4.6).

Proof. Let's denote $d_A := \dim(A)$. By Theorem 4.4, we need to evaluate $||T_A||_{\diamondsuit}$. It is insightful to realize that the Choi representation of T_A is simply the qudit swap operator

$$J_{T_A} = \frac{1}{d_A} (id \otimes T) \left(\sum_{i,j} |ii\rangle\langle jj| \right) = \frac{1}{d_A} \sum_{i,j} |ij\rangle \langle ji| = \frac{1}{d_A} SWAP_{AA'}. \tag{4.61}$$

Since SWAP_{AA'} is both Hermitian and unitary, its eigenvalues are precisely ± 1 and

$$||T_A||_{\diamondsuit} \ge ||J_{T_A}||_1 = \left\|\frac{1}{d_A} SWAP_{AA'}\right\|_1 = d_A.$$
 (4.62)

 \triangle

For the converse, we briefly discuss the spectral decomposition of SWAP_{AA'}. The +1 eigenstates are $|ii\rangle$ and $\frac{|ij\rangle+|ji\rangle}{\sqrt{2}}$ for i < j and the -1 eigenstates are $\frac{|ij\rangle-|ji\rangle}{\sqrt{2}}$ for i < j. The decomposition of J_{T_A} into the difference of the +1 and -1 eigenspace projectors

$$J_{T_A} = \Lambda^+ - \Lambda^- = \frac{d_A + 1}{2} \frac{\Lambda^+}{\text{tr}[\Lambda^+]} - \frac{d_A - 1}{2} \frac{\Lambda^-}{\text{tr}[\Lambda^-]}$$
(4.63)

induces a valid QPD of T_E w.r.t. CPTP because

$$\operatorname{tr}_{A'}\left[\frac{\Lambda^{+}}{\operatorname{tr}[\Lambda^{+}]}\right] = \frac{2}{d_{A}(d_{A}+1)}\operatorname{tr}_{A'}\left[\sum_{i}|ii\rangle\langle ii|\right] \tag{4.64}$$

$$+\frac{2}{d_A(d_A+1)}\operatorname{tr}_{A'}\left[\sum_{i\leq j}\frac{(|ij\rangle+|ji\rangle)(\langle ij|+\langle ji|)}{2}\right]$$
(4.65)

$$=\frac{2}{d_A(d_A+1)}\sum_i|i\rangle\langle i|+\frac{2}{d_A(d_A+1)}\sum_{i< j}\frac{|i\rangle\langle i|+|j\rangle\langle j|}{2} \tag{4.66}$$

$$= \frac{2}{d_A(d_A+1)} \mathbb{1}_A + \frac{d_A-1}{d_A(d_A+1)} \mathbb{1}_A \tag{4.67}$$

$$=\frac{\mathbb{1}_A}{d_A}\tag{4.68}$$

and analogously $\operatorname{tr}_{A'}\left[\frac{\Lambda^-}{\operatorname{tr}[\Lambda^-]}\right]=\frac{\mathbb{1}_A}{d_A}.$ Hence, we have

$$\gamma_{\text{CPTP}^*} \le \frac{d_A + 1}{2} + \frac{d_A - 1}{2} = d_A.$$
(4.69)

Quantum universal-NOT gate The universal-NOT (UNOT) gate is an anti-unitary single-qubit operation that maps a vector on the Bloch sphere to its opposite point

$$\alpha |0\rangle + \beta |1\rangle \mapsto \beta^* |0\rangle - \alpha^* |1\rangle$$
 (4.70)

As the basis-independent quantum analog of the classical NOT gate, the UNOT gate and its best CPTP approximations have been extensively studied [30], [58]. In fact, it has been shown to be closely related to the transpose superoperator and quantum cloning. Its action as a linear superoperator

$$q\mathbb{1} + \vec{r} \cdot \vec{\sigma} \mapsto q\mathbb{1} - \vec{r} \cdot \vec{\sigma} \tag{4.71}$$

for $q \in \mathbb{R}$, $\vec{r} \in \mathbb{R}^3$ is trace-preserving but not completely positive. Here, $\vec{r} \cdot \vec{\sigma}$ is shorthand notation for $r_x X + r_y Y + r_z Z$. Still, it is possible to simulate a UNOT gate using QPS. In fact, Equation (4.71) precisely provides us with its Pauli transfer representation diag(1, -1, -1, -1). By Proposition 4.6, this implies that $\gamma_{\text{CPTP}^*}(\text{UNOT}) = 2$ with the achieving QPD being

UNOT =
$$-\frac{1}{2}id + \frac{1}{2}Ad[X] + \frac{1}{2}Ad[Y] + \frac{1}{2}Ad[Z]$$
. (4.72)

Quantum cloning The famous quantum no-cloning theorem states that there exists no physical quantum process that maps every state $|\psi\rangle$ onto two copies of itself $|\psi\rangle \mapsto |\psi\rangle \otimes |\psi\rangle$. As such, much research has been devoted to finding the CPTP maps which best approximate the cloning operation. We refer the reader to [138] for a comprehensive overview of results on approximate cloning.

Since QPS allows for the simulation of nonphysical quantum operations, it stands that it could be used to effectively realize improved or even perfect cloning machines. Here, we investigate the arguably simplest setting of state-dependent cloning of two pure states. More precisely, we consider two pure states $|\psi_0\rangle$ and $|\psi_1\rangle$ and investigate cloning machines $\mathcal E$ which aim to maximize the average fidelity

$$\frac{1}{2}\left(\left\langle \psi_{0}\right|\otimes\left\langle \psi_{0}\right|\right)\mathcal{E}\left(\left|\psi_{0}\right\rangle\!\left\langle \psi_{0}\right|\right)\left(\left|\psi_{0}\right\rangle\otimes\left|\psi_{0}\right\rangle\right)+\frac{1}{2}\left(\left\langle \psi_{1}\right|\otimes\left\langle \psi_{1}\right|\right)\mathcal{E}\left(\left|\psi_{1}\right\rangle\!\left\langle \psi_{1}\right|\right)\left(\left|\psi_{1}\right\rangle\otimes\left|\psi_{1}\right\rangle\right).$$

Without loss of generality, we can assume that the two states are of the form

$$|\psi_0\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle$$
, $|\psi_1\rangle = \cos\frac{\theta}{2}|0\rangle - \sin\frac{\theta}{2}|1\rangle$ (4.73)

for $\theta \in [0, \pi/2]$ such that $\cos(\theta)$ characterizes the overlap $\langle \psi_0 | \psi_1 \rangle$.

Let us denote by A and B the input 1-qubit system and the output 2-qubit system. The maximum achievable fidelity F_{max} with a physical CPTP cloner can be expressed in terms of a simple SDP

$$F_{\max}(\theta) = \begin{cases} \max_{J_{\mathcal{E}} \in P(AB)} & \frac{1}{2} \left(\langle \psi_{0} | \otimes \langle \psi_{0} | \right) \mathcal{E} \left(| \psi_{0} \rangle \langle \psi_{0} | \right) \left(| \psi_{0} \rangle \otimes | \psi_{0} \rangle \right) \\ & + \frac{1}{2} \left(\langle \psi_{1} | \otimes \langle \psi_{1} | \right) \mathcal{E} \left(| \psi_{1} \rangle \langle \psi_{1} | \right) \left(| \psi_{1} \rangle \otimes | \psi_{1} \rangle \right) \\ \text{s.t.} & \operatorname{tr}_{B}[J_{\mathcal{E}}] = \frac{1}{2} \mathbb{1}_{A} \end{cases}$$

$$(4.74)$$

where $\mathcal{E}(\sigma)$ should be understood as the evolution in terms of the Choi matrix, i.e., $\mathcal{E}(\sigma) = \dim(A) \operatorname{tr}_A \left[J_{\mathcal{E}}(\sigma^T \otimes \mathbb{1}_B) \right]$.

While no perfect CPTP cloner exists (unless $\theta=0$ or $\theta=\pi/2$), it is easy to see that there always exists a Hermitian-preserving superoperator which achieves optimal cloning. At this point, it is natural to ask what is the smallest quasiprobability extent $\gamma_{\text{CPTP}^*}(\mathcal{E})$ over all the maps $\mathcal{E}\in \text{HP}(A\to B)$ which achieve perfect cloning. Again, this can be expressed as a SDP

$$\gamma_{\min}(\theta) = \begin{cases}
\max_{\substack{J_{\mathcal{E}} \in \mathcal{H}(AB) \\ Y_{0}, Y_{1} \in \mathcal{P}(AB)}} \frac{1}{2} \| \operatorname{tr}_{B}[Y_{0}] \|_{\infty} + \frac{1}{2} \| \operatorname{tr}_{B}[Y_{1}] \|_{\infty} \\
\text{s.t.} \quad \mathcal{E}(|\psi_{0}\rangle\langle\psi_{0}|) = |\psi_{0}\rangle\langle\psi_{0}| \otimes |\psi_{0}\rangle\langle\psi_{0}|, \\
\mathcal{E}(|\psi_{1}\rangle\langle\psi_{1}|) = |\psi_{1}\rangle\langle\psi_{1}| \otimes |\psi_{1}\rangle\langle\psi_{1}|, \\
\begin{pmatrix} Y_{0} & -J_{\mathcal{E}} \\ -J_{\mathcal{E}} & Y_{1} \end{pmatrix} \approx 0
\end{cases} (4.75)$$

where we used Theorem 4.4 as well as the SDP characterization of the diamond norm from Ref. [174].

The numerical evaluation of the SDPs in Equations (4.74) and (4.75) is depicted in Figure 4.2. Interestingly, the minimal quasiprobability extent does not directly

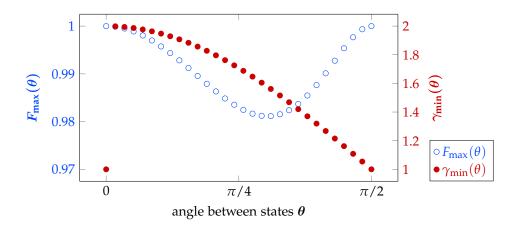


Figure 4.2: Numerics on state-dependent cloning machines for two pure input states with overlap $\cos\theta$. F_{max} denotes the maximum cloning fidelity for a physical (CPTP) cloning machine and γ_{min} denotes the minimal quasiprobability extent of a non-physical optimal cloning machine. The limits $\theta=0$ and $\theta=\pi/2$ correspond to two identical states, respectively two orthogonal states.

correlate with the maximum fidelity of the optimal CPTP cloner. Indeed, as the overlap between the states approaches 1, the approximate CPTP cloner becomes more accurate whereas the minimal quasiprobability extent seems to converge towards a value of 2. Notably, $\gamma_{\min}(\theta)$ is discontinuous at $\theta=0$, as cloning is trivial for identical states. Our results show that *exact* cloning is most expensive to simulate for θ close to zero, even though *approximate* cloning is very easy in this regime (by essentially doing almost nothing).

Imaginary time evolution Disclaimer: The idea described in this paragraph was orally conveyed to the author by Ray et al. [134]. The original authors plan to publish a detailed analysis in the near future. The current paragraph was written with their explicit permission and all credit for the idea is due to them.

Imaginary time evolution is a standard technique in computational quantum mechanics to find the ground state of a given Hamiltonian H. Consider evolving a random initial state $|\psi\rangle$ with the operator $e^{-\tau H}$ for some $\tau\geq 0$. This evolution can be interpreted as "imaginary time evolution", since it corresponds to substituting imaginary time $\tau=it$ into the time evolution operator e^{-iHt} derived from the Schrödinger equation (note that we are fixing $\hbar=1$). By choosing τ large enough, the ground state component becomes dominant.

More precisely, let $H|b_i\rangle = E_i|b_i\rangle$ be the eigenvectors and eigenvalues of H. The imaginary time evolution expressed in the eigenbasis reads

$$|\psi\rangle = \sum_{i} c_{i} |b_{i}\rangle \mapsto e^{-\tau H} |\psi\rangle = \sum_{i} c_{i} e^{-\tau E_{i}} |b_{i}\rangle$$
 (4.76)

so the ground state component decays exponentially slower than the other components. The convergence time τ_{conv} is governed by the inverse gap $\tau_{conv} = \mathcal{O}(\frac{1}{\Delta})$ where

 $\Delta := E_1 - E_0$. Using QPS, we can simulate the imaginary time evolution operator and hence estimate the ground state energy

$$E_0 \approx \frac{\langle \psi' | H | \psi' \rangle}{\langle \psi' | \psi' \rangle} \text{ where } | \psi' \rangle \coloneqq e^{-\tau H} | \psi \rangle$$
 (4.77)

by estimating the two expectation values $\langle \psi'|H|\psi'\rangle$ and $\langle \psi'|\psi'\rangle$. Recalling Proposition 3.1 and Theorem 4.4, we can estimate both to an additive error ϵ with a sampling overhead of $\mathcal{O}\left(\|\operatorname{Ad}\left[e^{-\tau H}\right]\|_{\Diamond}^{2}\epsilon^{-2}\right)$. As τ increases, the quasiprobability extent $\|\operatorname{Ad}\left[e^{-\tau H}\right]\|_{\Diamond}$ decreases, but so does also the required additive error ϵ . We refrain from a full analysis of the scheme and instead refer to Ref. [134].

4.4 Further reading

Two separate works by Piveteau et al. [126], [131] and Jiang et al. [88] independently introduced the idea of decomposing RHPTP linear maps into CPTP operations. The former calls the optimal QPD the channel difference decomposition and uses it as an intermediate technical results for an error mitigation protocol. The latter calls the log-quasiprobability extent the physical implementability, and shows some basic properties (faithfulness, additivity, unitary invariance, relation to robustness) as well as weak monotonicity under superchannels. Both express the quasiprobability extent in terms of a semidefinite program, but it was only later work by Regula et al. [136] that realized that this SDP is equivalent to the diamond norm (at least for RHPTP maps). This realization essentially implies that many proofs in the previous papers just re-stated known properties of the diamond norm. The authors also relate the quasiprobability extent to some advantage that non-physical maps can provide in discrimination-based quantum games. In addition, this third work also realized that all Hermitian-preserving maps can be simulated if the decomposition set is enlarged from CPTP to the set of all completely positive trace-non-increasing maps CPTN. Note that this does not capture the full extent of using classical side information, as it effectively only allows for non-negative weighting of the quantum instruments. Indeed, there exist explicit non-TP Hermitian-preserving maps $\mathcal E$ for which $\gamma_{\mathrm{CPTN}}(\mathcal E)$ is strictly greater than $\gamma_{CPTP^*}(\mathcal{E})$, but still finite. Only further work by Xuanqiang *et* al. [182] incorporated the full power of classical side information and realized that the diamond norm characterization also extends to γ_{CPTP^*} .

The study of QPDs of non-physical maps is related to the technique of *structural physical approximation* (SPA) [83]. The underlying motivation is that positive, but not completely positive, maps $\mathcal{E} \in \mathrm{HP}(A)$ can serve as entanglement witnesses: $\mathcal{E} \otimes \mathbb{1}_B(\rho_{AB})$ can only have negative eigenvalues if ρ_{AB} is itself entangled. The most famous instance is the Peres-Horodecki criterion, where \mathcal{E} is chosen to be the transpose map [124]. Unfortunately, such positive non-CP maps are not physically realizable, making it difficult to directly translate these mathematical criteria into experimentally realizable protocols to detect entanglement. It can be shown that taking a mixture of a non-CP map with the fully-depolarizing channel $\mathcal{D}(\rho) \coloneqq \mathbb{1}_A / \dim(A)$

will produce a CPTP map

$$\tilde{\mathcal{E}} := (1 - p)\mathcal{E} + p\mathcal{D} \in CPTP(A) \tag{4.78}$$

provided the mixing probability p is chosen large enough. This map $\tilde{\mathcal{E}}$ for the smallest admissible p is called a structural physical approximation of \mathcal{E} and it can be used to physically realize entanglement detection protocols. We refer the reader to [6], [84], [98] for more details. Clearly, any SPA can be seen as a QPD

$$\mathcal{E} = \frac{1}{1-p}\tilde{\mathcal{E}} - \frac{p}{1-p}\mathcal{D}. \tag{4.79}$$

Therefore, QPDs can be considered a generalization of SPAs which allows for mixing with operations beyond the depolarizing channel to achieve complete positivity. The minimal admissible probability p which gives a valid SPA is thus an upper bound for the quasiprobability extent and can easily be shown to be related to a variant of the robustness measure from Definition 3.16, called the *random robustness* [43], [166].

4.5 Contributions

The idea of decomposing and simulating general $\mathbb{R}HPTP$ maps with CPTP maps was proposed by the author in Ref. [126], [131]. Later work in Ref. [88] indpendently proposed the same idea. The SDP formulation of γ_{CPTP} is due to the author. However, the connection to the diamond norm as well as the generalization to γ_{CPTP^*} were first introduced by other researchers.

The applications in Section 4.3, including Proposition 4.6, are novel work and have not been published before. The only exception is the paragraph *imaginary time evolution*, which is due to Ref. [134].

Chapter 5

Simulating nonlocal computation

In this chapter, we will study QPS in the context of the QRT of entanglement, where the free operations across two parties are given by local operations with classical communication (LOCC). In the broadest sense, this means that we will study the quasiprobability extent $\gamma_{\text{LOCC}^*}(\mathcal{E})$ of bipartite nonlocal channels $\mathcal{E} \in \text{CPTP}(AB)$ across two parties A and B. As before, the set LOCC* (as opposed to LOCC) captures the possibility of utilizing classical side information from intermediate measurements in the sampling procedure (recall the discussion in Section 3.3 and Equation (3.75)). In contrast to the QRT of physical operations we encountered in Chapter 4, the characterization of $\gamma_{\text{LOCC}^*}(\mathcal{E})$ will turn out to be extremely difficult, and we will only find closed-form expressions for a few simple cases. This can be attributed to the very rich structure of the QRT, and we will establish many connections to the study of entanglement.

The main practical application of QPS of nonlocal operations is for the purpose of realizing *circuit knitting* (also called *circuit cutting*) to simulate a *large* quantum computer using *small* quantum computers. Suppose one wants to run a quantum circuit on a given quantum device, but its number of available qubits is too small to fit the circuit. The premise of circuit knitting is to subdivide the circuit into smaller partitions which are each small enough to fit on the available quantum computer. The goal is to reconstruct the outcome of the original circuit by executing the smaller partitions on the quantum computer and performing appropriate classical post-processing.

It is very natural to realize circuit knitting through QPS. Consider for instance the setup depicted in Figure 5.1. The qubits in a circuit are grouped into two systems *A* and *B*. Every gate that acts nonlocally across *A* and *B* is simulated using LOCC operations via QPS. This effectively means that we express the original large circuit as a quasiprobabilistic mixture of two sub-circuits which interact with each other only through classical communication. As such, we can simulate the original quantum circuit with two smaller quantum computers, one executing *A* and the other executing *B*. Note that this approach also works when the circuit is divided into more than 2 partitions.

The partitioning of a circuit in circuit knitting can generally be achieved using two kinds of "cuts". The previous example illustrated *space-like cuts* (sometimes also called *gate cuts*) which horizontally split nonlocal gates into local operations. There

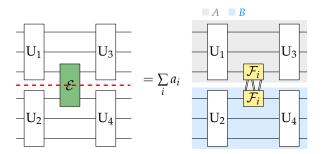


Figure 5.1: Example of a space-like cut (red dashed line) which partitions a large 6-qubit circuit into two 3-qubit sub-circuits. Using a QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ of \mathcal{E} into LOCC operations \mathcal{F}_i , the total circuit can be written as a probabilistic mixture of two separate 3-qubit circuits A and B which only interact through classical communication.

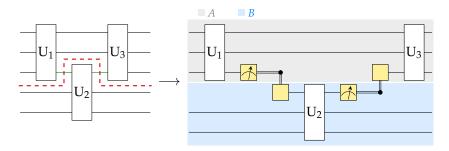


Figure 5.2: Example of a wire cut (red dashed line) which partitions a large 5-qubit circuit into two 3-qubit sub-circuits. Using a QPD of the identity channel into entanglement breaking channels, the total circuit can be written as a probabilistic mixture of two separate 3-qubit circuits *A* and *B* which only interact through classical communication.

is also a second kind of cut, called *time-like cuts* (or *wire cuts*), which vertically split a qubit wire as depicted on the left side of Figure 5.2. A wire cut can be achieved through QPS by decomposing the qubit identity channel into a quasiprobabilistic mixture of *entanglement breaking channels*¹ which consist of a measurement followed by a state preparation which may depend on the measurement outcome. As depicted in Figure 5.2, this again allows us to write the total circuit as a quasiprobabilistic mixture of two sub-circuits which only communicate via classical communication, and hence can be executed on two smaller quantum devices.

In this chapter, we will study both space-like cuts and time-like cuts and their associated quasiprobability extents γ_{LOCC^*} and γ_{EBC^*} where EBC stands for entanglement breaking channels. Correspondingly, the chapter is split up into Sections 5.1 and 5.2 treating the two cases. Below, we briefly outline some specific questions that will be considered in detail.

¹The reason for the name "entanglement breaking channel" will become more apparent in Section 5.2.

Utility of classical communication In the above descriptions of space-like and time-like cutting, we assumed that the individual sub-circuits can exchange classical communication with each other. Interestingly, circuit knitting remains possible even without this capability. We will investigate whether classical communication between the sub-circuits can allow for a lower simulation overhead. Mathematically speaking, for space-like cuts this amounts to comparing γ_{LOCC^*} with γ_{LO^*} where LO stands for local operations *without* classical communication. Similarly, for time-like cuts, we will also introduce a decomposition set which is based on channels consisting of measurements and state preparations, but in constrast to EBC, the state may not depend on the measurement outcome.

While the importance of classical communication is academically very interesting, it also has some important practical ramifications for circuit knitting. Allowing for classical communication could a priori lead to a lower sampling overhead, since we are extending the decomposition set to a larger set. However, this comes at an additional price in terms of practical implementation costs: Without classical communication, the individual sub-circuits could be run sequentially and in principle on the same quantum device. However, once classical communication is involved, one will generally require multiple small quantum processors executing the sub-circuits simultaneously.

Somewhat surprisingly, we will observe quite different results for the settings of space-like and time-like cutting. We will show that for a large range of nonlocal gates, their quasiprobability extents $\gamma_{\text{LOCC}^{\star}}$ and $\gamma_{\text{LO}^{\star}}$ are identical. In comparison, time-like cuts are significantly less expensive with classical communication.

Note that when we are referring to the utilization of classical communication here, we explicitly mean classical communication *during* the circuit execution. Circuit knitting inherently requires some form of communication to combine the final measurement results of the individual sub-circuits. However, this final communication can be considered part of the post-processing and doesn't entail the same technological requirements as mid-circuit classical communication.

Utility of classical side information Following the discussion in Section 3.3, we will consider the decomposition sets that capture the utilization of classical side information, e.g., LOCC* instead of LOCC and EBC* instead of EBC (recall Equation (3.75)). It is a natural question to ask whether this is really useful and if we could achieve the same overhead without classical side information.

Intriguingly, the answer seems to be very different in the settings with and without classical communication. For instance, $\gamma_{\text{LOCC}^*}(\mathcal{U}) = \gamma_{\text{LOCC}}(\mathcal{U})$ when \mathcal{U} is a Clifford gate, so classical side information is generally not always required to reach the lowest extent. In contrast, we will show in Section 5.1.2 that intermediate measurements are absolutely crucial in the absence of classical communication. Mathematically, this manifests as $\gamma_{\text{LO}}(\mathcal{E}) = \infty$ for any signalling operation \mathcal{E} whereas $\gamma_{\text{LO}^*}(\mathcal{E})$ is finite for all Hermitian-preserving superoperators \mathcal{E} . An analogous result also holds for time-like cuts, as discussed in Section 5.2.2.

Space-like cuts of states Besides just studying space-like cuts of quantum channels, we will also spend significant effort in Section 5.1.3 considering space-like cuts of quantum states, i.e., QPDs of bipartite nonlocal states into separable states (or equivalently product states). This can not only be useful as a technical tool for bounding the quasiprobability extent of channels, but also on its own for cutting circuits that contain nonlocal states.

To illustrate the importance, we highlight that there exist quantum gates that can be realized from a pre-shared entangled state using only LOCC. A great example for this is the CNOT gate: The gate-teleportation protocol [64] depicted in Figure 5.3 allows for the realization of a CNOT gate using only a pre-shared Bell pair $|\Psi\rangle$ and LOCC operations. As such, any QPD of a Bell pair into product states $|\Psi\rangle\langle\Psi|=\sum_i a_i\rho_i\otimes\sigma_i$ naturally induces a QPD of the CNOT gate into LOCC channels. We note that this is completely analogous to the magic state injection procedure discussed in Example 3.6.

Scaling of quasiprobability extent under tensor product. Another important aspect that we will study in this chapter is the scaling of the quasiprobability extent under the tensor product. In many instances, the quasiprobability extent is strictly sub-multiplicative, which has direct ramifications for the asymptotic overhead of circuit knitting. Consider for instance the extent of a non-local Bell-pair. We will later see that it is given by $\gamma(|\Psi\rangle\langle\Psi|) = 3$, while $\gamma(|\Psi\rangle\langle\Psi|^{\otimes n}) = 2^{n+1} - 1 < 3^n$ and $\overline{\gamma}(|\Psi\rangle\langle\Psi|) = \gamma^{\infty}(|\Psi\rangle\langle\Psi|) = 2$. It is therefore cheaper to quasiprobabilistically simulate the joint preparation of n nonlocal Bell pairs instead of optimally simulating them separately. When simulating c nonlocal CNOT gates through the gate teleportation approach (as discussed in Figure 5.3), this joint simulation strategy constitutes a large reduction in sampling overhead from $\mathcal{O}(9^c)$ to $\mathcal{O}(4^c)$.

While the regularized and asymptotic quasiprobability extent for the Bell state $|\Psi\rangle$ are identical, we will show that this is surprisingly not the case for almost all other pure states. More precisely, we will show that for a pure bipartite state $|\phi\rangle_{AB}$ one has $\log \overline{\gamma}(|\phi\rangle\langle\phi|) = H_{1/2}(\mathrm{tr}_B[|\phi\rangle\langle\phi|])$ and $\log \gamma^\infty(|\phi\rangle\langle\phi|) = H(\mathrm{tr}_B[|\phi\rangle\langle\phi|])$ where H and $H_{1/2}$ denote the von Neumann entropy and the quantum Rényi-1/2 entropy.

The quasiprobability extent of an n-fold copy of some gate $\gamma(\mathcal{U}^{\otimes n})$ captures the optimal overhead for simulating n instances of the nonlocal gate \mathcal{U} in parallel, i.e., in the same time slice of the circuit. This is a strong assumption, as typically the nonlocal gates in a circuit can occur far apart from each other. We will see that in some instances, it remains possible to harness the submultiplicative behavior of the quasiprobability extent for distant gates and reduce the sampling overhead. We refer to this as the $black\ box\ setting$, and it will be discussed in Section 5.1.5.

²We are being intentionally imprecise here by omitting the decomposition set. We defer the formal mathematical discussion to later in the chapter.

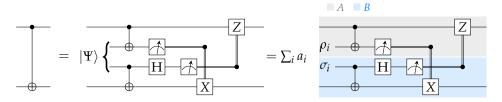


Figure 5.3: The first equality shows how a nonlocal CNOT gate between two parties can be realized using a LOCC protocol that consumes a pre-shared Bell state $|\Psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. By writing the Bell pair in terms of a quasiprobabilistic mixture of product state $|\Psi\rangle\langle\Psi| = \sum_i a_i \rho_i \otimes \sigma_i$, we directly obtain a QPD of the CNOT gate into LOCC protocols. The equalities should be understood in terms of superoperators.

Generally speaking, there is no efficient methodology for evaluating the quasiprobability extent for space-like and time-like cuts of arbitrary channels. Instead, we will have to content ourselves with various techniques to find lower and upper bounds. Fortunately, in some instances, these bounds match. The most important such examples that we will encounter are

- space-like cuts of pure states (see Theorem 5.13 for the single-shot result and Corollaries 5.16 and 5.32 for the asymptotic results),
- space-like cuts of Clifford gates (see Corollary 5.41),
- space-like cuts of parallel two-qubit gates (see Theorem 5.44) and
- time-like cuts of the identity channel (see Theorems 5.75 and 5.77).

Generally, we will find upper bounds by explicitly constructing a QPD with the corresponding 1-norm of its coefficients. To find lower bounds, we will often (but not always) proceed by relaxing the decomposition set to a larger set, for which the quasiprobability extent is easier to evaluate. For instance, LOCC has a notoriously complicated structure. To avoid dealing with its difficulties, we will often consider the relaxations to the sets of separable channels (SEPC) and positive partial transpose channels (PPTC). Especially the latter exhibits a much simpler characterization of the quasiprobability extent in terms of a semidefinite program.

In the context of space-like cuts, we briefly note that the set LOCC contains protocols that incur an unbounded number of communication rounds. As such, the time complexity for realizing arbitrary LOCC protocols can be exceedingly high, possibly overshadowing the sampling overhead of QPS in the first place. A practically more meaningful decomposition set would be based on the set LOCC $_r$ with at most r rounds, for some appropriately chosen r. However, in this thesis, we will only consider the unbounded round LOCC setting for two reasons. First, it is already exceedingly difficult to make statements about γ_{LOCC^*} without having to worry about the number of communication rounds. Furthermore, we are fortunate that all protocols that we will encounter (like the ones mentioned in the list above) anyway only involve a small number of rounds.

5.1 Space-like cuts

5.1.1 Decomposition sets

In this section, we start by formally introducing all relevant decomposition sets that will play a major role for our study of space-like cuts. We start by characterizing the sets of states that are important for the resource theory of entanglement.

Definition 5.1: For two quantum systems A, B, we define the set of separable states SEP(A; B) to be all bipartite states $\rho_{AB} \in D(AB)$ of the form

$$\rho_{AB} = \sum_{i=1}^{m} p_i \rho_A^{(i)} \otimes \sigma_B^{(i)}$$
 (5.1)

where $m \in \mathbb{N}$, $p \in \mathbb{R}^m$ is a probability vector and $\rho_A^{(i)} \in D(A)$, $\sigma_B^{(i)} \in D(B)$.

The set SEP(A; B) precisely contains all states that are reachable from an initial product state using local operations and classical communication. By definition, they are the states that do not possess any entanglement, as their correlations can be thought to be of purely classical origin.

Since separability is not expressible as a semidefinite constraint, one often encounters the challenge that optimizing over SEP(A;B) is difficult. To address this, a common approach is to relax such problems by instead optimizing over the larger set of *positive partial transpose states*. For this purpose, we define the transpose map T_A for some system A w.r.t. an orthonormal basis $\{|b_1\rangle, \ldots, |b_n\rangle\}$ by

$$T_A\left(\sum_{i,j}\alpha_{i,j}|b_i\rangle\langle b_j|\right) := \sum_{i,j}\alpha_{j,i}|b_i\rangle\langle b_j|. \tag{5.2}$$

Definition 5.2: For two quantum systems A, B, we define the set of positive partial transpose states PPT(A; B) to be all bipartite states $\rho_{AB} \in D(AB)$ which fulfill $(id_A \otimes T_B)(\rho_{AB}) \geq 0$.

We leave it as an exercise to the reader to verify that this definition does not depend on the choice of orthonormal basis or the system on which the transpose is taken. It is also evident to see that any separable operator has a positive partial transpose, i.e., $SEP(A; B) \subset PPT(A; B)$. This gives rise to the well-known Peres-Horodecki criterion: A quantum state that does not exhibit a positive partial transpose must be entangled [124]. For small systems, this criterion is faithful.

Lemma 5.3 ([81]): If
$$\dim(A) \cdot \dim(B) \le 6$$
 then $SEP(A; B) = PPT(A; B)$.

However, for larger systems, the inclusion $SEP(A; B) \subset PPT(A; B)$ becomes strict. The following state, due to Ref. [13], provides an example of an entangled PPT state for dim(A) = dim(B) = 3.

Example 5.4. Consider the two systems A, B with $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^3$ and the five states

$$|\psi_1
angle = rac{1}{\sqrt{2}}|0
angle \otimes (|0
angle - |1
angle), \qquad |\psi_2
angle = rac{1}{\sqrt{2}}(|0
angle - |1
angle) \otimes |2
angle,
onumber \ |\psi_3
angle = rac{1}{\sqrt{2}}|2
angle \otimes (|1
angle - |2
angle), \qquad |\psi_4
angle = rac{1}{\sqrt{2}}(|1
angle - |2
angle) \otimes |0
angle,
onumber \ |\psi_5
angle = rac{1}{3}(|0
angle + |1
angle + |2
angle) \otimes (|0
angle + |1
angle + |2
angle) \qquad .$$

Any product state has non-zero overlap with at least one of these states. Therefore, any state in the orthogonal complement of these five states \mathcal{H}^{\perp} must be entangled. Denote $\rho \coloneqq \frac{1}{4}\Pi$ where Π is the projection onto \mathcal{H}^{\perp} . On one hand, ρ must be entangled since $\Pi\rho\Pi = \rho$. On the other hand, ρ is invariant under partial transposition, because $\rho \propto \mathbb{1} - \sum_{i=1}^{5} |\psi_i\rangle\langle\psi_i|$ and each $|\psi_i\rangle\langle\psi_i|$ is clearly invariant under partial transposition. Therefore, $\rho \in \operatorname{PPT}(A;B)$ but $\rho \notin \operatorname{SEP}(A;B)$.

Next, we introduce four resource theories which are called LO, LOCC, SEPC and PPTC. The QRT LOCC constitutes the cornerstone of the study of entanglement in quantum information theory. It is operationally defined as the set of channels that can be realized by two³ parties which can perform arbitrary quantum operations on their respective local system, but are limited to classical communication between each other. As such, it has a very rich structure and far-reaching applications for many entanglement manipulation tasks, such as the preparation, discrimination and transformation of entangled states. In contrast, LO corresponds to the more restricted setting where the two parties cannot exchange classical communication with each other. It doesn't have the same operational applications and is mathematically easy to handle, at least compared to LOCC. For the purposes of QPS, LO captures the setting where the two sub-circuits do not have the capabilities to exchange classical information.

The set LOCC is notoriously difficult to work with, and as such, it is often necessary to resort to relaxations of it. SEPC and PPTC are two supersets of LOCC that are mathematically easier to handle. Compared to LO and LOCC, they have no operational meaning for the purposes of QPS and circuit knitting, and instead will serve purely as mathematical tools to bound the quasiprobability extent w.r.t. LOCC.

We now define the four resource theories in full detail.

Local operations (LO): The set of free transformations for this QRT is given by all product channels

$$LO(A; B \to A'; B') := \{ \mathcal{E} \otimes \mathcal{F} | \mathcal{E} \in CPTP(A \to A'), \mathcal{F} \in CPTP(B \to B') \}$$
 (5.3)

and the associated free instruments are

$$\mathfrak{I}_{LO}(A; B \to A'; B') := \left\{ (\mathcal{E}_i \otimes \mathcal{F}_j)_{i,j} \middle| \begin{array}{l} (\mathcal{E}_i)_i \in \mathrm{QI}(A \to A') \\ (\mathcal{F}_j)_j \in \mathrm{QI}(B \to B') \end{array} \right\}. \tag{5.4}$$

³The four QRTs can straightforwardly be defined for arbitrarily many parties, but we restrict ourselves to the two-party case here.

For the purposes of QPS, we are interested in the induced decomposition set that accounts for classical side information (recall the discussion in Section 3.3)

$$LO^{\star}(A; B \to A'; B') := \operatorname{chan}^{*}[\Im_{LO}(A; B \to A'; B')]. \tag{5.5}$$

The free states in the QRT of LO are precisely the product states of the form $\rho_A \otimes \sigma_B$ where $\rho_A \in D(A)$ and $\sigma_B \in D(B)$.

Despite its very simple definition, LO unfortunately lacks certain structure compared to the other three QRTs (LOCC, SEPC, PPTC) that makes it a bit cumbersome to work with in the context of QPS. For instance, it is not a convex resource theory, so Lemmas 3.11, 3.14 and 3.28 don't apply, and we generally need to consider QPDs with an unbounded number of terms. Furthermore, \mathfrak{I}_{LO} is neither coarse-grainable nor trivially fine-grainable⁴, and hence no simple characterization of LO* in the sense of Lemma 3.9 is possible. In the context of Corollary 3.38, this can be attributed to the fact that general classical computation is not free under LO (only *local* classical computation is).

Some previous work instead considered the decomposition set [129]

$$\{\mathcal{E} \otimes \mathcal{F} | \mathcal{E} \in CPTP^{*}(A), \mathcal{F} \in CPTP^{*}(B)\}$$
 (5.6)

(recall Equation (4.1) for the characterization of CPTP*). While this set spans the same space as LO* (as we will see later), it is a strict subset of LO*(A; $B \rightarrow A'$; B') since the weighting based on the measurement outcomes on A and B is done independently.

Local operations with classical communication (LOCC): The formal mathematical construction of the set LOCC is quite involved and typically phrased in terms of the associated free instruments \mathfrak{I}_{LOCC} . Since this formal definition will not be particularly insightful for our considerations, we refrain from reproducing it here and instead refer the interested reader to the excellent exposition in Ref. [44]. It suffices to note that the set of free instruments $\mathfrak{I}_{LOCC}(A; B \to A'B') \subset QI(AB \to A'B')$ is

- coarse-grainable by construction.
- trivially fine-grainable, as any protocol can be extended by having one of the two parties sample a random bit with probability depending on all previous measurement outcomes.
- closed under mixture, as the two parties can randomly choose between one of two protocols using one round of classical communication at the beginning.

By Lemmas 3.9 and 3.28, the induced decomposition set LOCC* $(A; B \rightarrow A'; B') := \text{chan}^*[\mathfrak{I}_{LOCC}(A; B \rightarrow A'B')]$ is absolutely convex and can be characterized as

$$LOCC^{\star}(A; B \to A'; B') = \{\mathcal{E} - \mathcal{F} | (\mathcal{E}, \mathcal{F}) \in \mathfrak{I}_{LOCC}(A; B \to A'; B')\}. \tag{5.7}$$

⁴Note that the two instruments on each bipartition can be individually coarse-grained or finegrained. However, the combined instrument can generally not, as that would require additional classical communication.

Under the resource theory of LOCC, the free states on a bipartite system AB are precisely given by SEP(A; B).

Separable channels (SEPC): We start by defining the set of separable quantum instruments $\mathfrak{I}_{SEPC}(A; B \to A'; B')$ to contain all instruments $(\mathcal{E}_i)_i \in QI(AB \to A'B')$ which have a separable Kraus representation, i.e., for all i we can write

$$\mathcal{E}_i(\rho) = \sum_j (M_j \otimes N_j) \rho(M_j^{\dagger} \otimes N_j^{\dagger})$$
 (5.8)

for some $M_j \in L(A,A')$, $N_j \in L(B,B')$. As an alternative characterization, the instruments $\mathcal{I} = (\mathcal{E}_i)_i \in \mathfrak{I}_{SEPC}$ are precisely those for which the associated Choi representations are separable $(J_{\mathcal{E}_i}/\operatorname{tr}[J_{\mathcal{E}_i}]) \in \operatorname{SEP}(AA';BB')$ [47]. Correspondingly, the set of separable channels $\operatorname{SEPC}(A;B \to A';B')$ is defined to be the subset of all quantum channels $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ that exhibit a separable Choi operator $J_{\mathcal{E}} \in \operatorname{SEP}(AA';BB')$. It can also be shown that SEPC precisely constitutes the set of quantum channels which are *completely resource non-generating*, i.e, which cannot generate entanglement even when applied to a subsystem [45, Proposition 17].

By construction, it can easily be seen that any LOCC instrument is separable, and hence $\mathfrak{I}_{LOCC}(A; B \to A'; B') \subset \mathfrak{I}_{SEPC}(A; B \to A'; B')$. In fact, it has been shown that SEPC($A; B \to A'; B'$) is in general strictly larger than LOCC($A; B \to A'; B'$) [12].

It is easy to verify that \mathfrak{I}_{SEPC} is coarse-grainable, trivially fine-grainable and closed under mixture. By Lemmas 3.9 and 3.28, the induced decomposition set $SEPC^*(A; B \to A'; B') := chan^*[\mathfrak{I}_{SEPC}(A; B \to A'B')]$ is absolutely convex and can be characterized as

$$SEPC^{\star}(A; B \to A'; B') = \left\{ \mathcal{E} - \mathcal{F} \middle| \begin{array}{l} (\mathcal{E}, \mathcal{F}) \in QI(AB \to A'B'), \\ \frac{J_{\mathcal{E}}}{\operatorname{tr}[J_{\mathcal{E}}]}, \frac{J_{\mathcal{F}}}{\operatorname{tr}[J_{\mathcal{F}}]} \in SEP(AA'; BB') \end{array} \right\}.$$
 (5.9)

By the Choi characterization, the free states of this QRT are precisely the separable states SEP(A; B).

Positive partial transpose channels (PPTC): SEPC provides a relaxation of the complicated LOCC constraint with a much simpler separability constraint on the Choi matrix. However, even separability is not a semidefinite constraint, and as such, it is sometimes useful to further relax the problem using the PPT criterion. We define the PPTC instrument as

$$\mathfrak{I}_{\mathrm{PPTC}}(A; B \to A'; B') := \{ (\mathcal{E}_i)_i \in \mathrm{QI}(AB \to A'B') | \frac{J_{\mathcal{E}_i}}{\mathrm{tr}[J_{\mathcal{E}_i}]} \in \mathrm{PPT}(AA'; BB') \}$$

$$(5.10)$$

and correspondingly, the set of PPT channels $PPTC(A; B \to A'; B')$ is given by all maps $\mathcal{E} \in CPTP(AB \to A'B')$ with PPT Choi representation $J_{\mathcal{E}} \in PPT(AA'; BB')$. Again, it can be shown that PPTC precisely constitutes the set of quantum channels which are completely PPT-preserving, i.e., which cannot

create non-PPT states from PPT states, even when only applied to a subsystem [45, Proposition 17]. The QRT PPTC is sometimes also called the *resource theory of NPT-entanglement*.

It is easy to verify that \mathfrak{I}_{PPTC} is coarse-grainable, trivially fine-grainable and closed under mixture. By Lemmas 3.9 and 3.28, the induced decomposition set $PPTC^*(A; B \to A'; B') \coloneqq chan^*[\mathfrak{I}_{PPTC}(A; B \to A'B')]$ is absolutely convex and can be characterized as

$$PPTC^{\star}(A; B \to A'; B') = \left\{ \mathcal{E} - \mathcal{F} \middle| \begin{array}{l} (\mathcal{E}, \mathcal{F}) \in QI(AB \to A'B'), \\ (\mathrm{id}_{AA'} \otimes T_{BB'})(J_{\mathcal{E}}) \geq 0, \\ (\mathrm{id}_{AA'} \otimes T_{BB'})(J_{\mathcal{F}}) \geq 0 \end{array} \right\}.$$
 (5.11)

By the Choi characterization, the free states of this QRT are precisely the PPT states PPT(A; B).

To simplify the notation, we will omit the output system whenever it is equal to the input system, e.g., $PPTC^*(A;B) := PPTC^*(A;B \to A;B)$ or $\mathfrak{I}_{LOCC}(A;B) := \mathfrak{I}_{LOCC}(A;B \to A;B)$. It is easy to verify that LO, LOCC, SEPC and PPTC are all valid QRTs with tensor product structure. Thus many properties in Sections 3.4 and 3.5 directly apply. For instance, the associated quasiprobability extents

- fulfill the chaining property (Lemma 3.31),
- are sub-multiplicative (Corollary 3.37),
- are invariant under local unitaries (Corollary 3.33),
- are Lipschitz continuous with $L = \dim(AB)^{5/2}$ on qubit systems A = A', B = B' (Example 3.24).

The QRTs LOCC, SEP and PPTC are even convex, so the associated quasiprobability extents are also

- directly related to a robustness measure (Lemmas 3.17 and 3.18),
- expressible in terms of single-element QPDs (Lemma 3.14).

Our four QRTs constitute a hierarchy in the sense that

$$LO(A; B \to A'; B') \subset LOCC(A; B \to A'; B') \subset SEPC(A; B \to A'; B') \subset PPTC(A; B \to A'; B')$$
 and correspondingly

$$LO^{\star}(A; B \to A'; B') \subset LOCC^{\star}(A; B \to A'; B') \subset SEPC^{\star}(A; B \to A'; B') \subset PPTC^{\star}(A; B \to A'; B').$$
(5.13)

In terms of the quasiprobability extent, Lemma 3.10 thus directly implies

$$\gamma_{\text{LO}^*} \geq \gamma_{\text{LOCC}^*} \geq \gamma_{\text{SEP}^*} \geq \gamma_{\text{PPT}^*}$$
 $\wedge | \qquad \wedge | \qquad \wedge | \qquad \wedge |$
 $\gamma_{\text{LO}} \geq \gamma_{\text{LOCC}} \geq \gamma_{\text{SEP}} \geq \gamma_{\text{PPT}}$ (5.14)

As we proceed further in this chapter, we will obtain a more advanced understanding of these inequalities and whether they are tight or not. The culminated results will later be summarized in Figure 5.11. Let us highlight once again that only the two quantities γ_{LO^*} and γ_{LOCC^*} have practical relevance in the context of circuit knitting. They precisely characterize the optimal sampling overhead for the two technologically relevant cases when the smaller sub-partitions can exchange classical communication or not. The main effort is therefore geared towards characterizing these two quantities, and the other ones purely serve as mathematical tools to find bounds.

When specializing from channels to states (i.e. channels with trivial input space), the situation in Equation (5.14) simplifies significantly. First of all, since LOCC is a resource theory that fulfills the axiom of free instruments, Lemma 3.29 implies that the quasiprobability extent is not impacted by the inclusion of intermediate measurements

$$\forall \omega \in H(AB) : \gamma_{LOCC^{*}(C;C \to A;B)}(\omega) = \gamma_{LOCC(C;C \to A;B)}(\omega) = \gamma_{SEP(A;B)}(\omega)$$
 (5.15)

and we can focus our study purely on the quasiprobability extent w.r.t. the set of separable channels. In fact, the same also holds for LO and SEPC⁵, and an analogous statement holds for PPTC.

Lemma 5.5: Consider $\omega \in H(AB)$. If *S* is either LO, LO*, LOCC, LOCC*, SEPC or SEPC*, then

$$\gamma_{S(\mathbb{C};\mathbb{C}\to A;B)}(\omega) = \gamma_{SEP(A;B)}(\omega). \tag{5.16}$$

If instead S is either PPTC or PPTC * , then

$$\gamma_{S(C;C\to A;B)}(\omega) = \gamma_{PPT(A;B)}(\omega). \tag{5.17}$$

So to summarize, in the case of states Equation (5.14) collapses to

$$\gamma_{\rm SEP} \ge \gamma_{\rm PPT}$$
. (5.18)

Proof. By Equations (5.14) and (5.15), it remains to show $\gamma_{\text{LO}}(\omega) \leq \gamma_{\text{SEPC}^*}(\omega)$ to prove the first equation. For this purpose, consider a QPD $\omega = \sum_i a_i \sigma_i$ with $\sigma_i \in \text{SEPC}^*$. This means that we can write $\sigma_i = \sum_j \beta_j^{(i)} \tau_j^{(i)}$ where the $\tau_j^{(i)}$ are subnormalized separable states that add up to a normalized separable state. We write $\tau_j^{(i)} = \sum_k p_{ijk} |f_{ijk}\rangle\langle f_{ijk}| \otimes |g_{ijk}\rangle\langle g_{ijk}|$ where $\sum_{j,k} p_{ijk} = 1$. We obtain a valid QPD of ω w.r.t. LO (i.e. w.r.t. to separable states)

$$\omega = \sum_{i,j,k} a_i \beta_j^{(i)} p_{ijk} |f_{ijk}\rangle\langle f_{ijk}| \otimes |g_{ijk}\rangle\langle g_{ijk}|$$
 (5.19)

with 1-norm $\sum_{ijk} |a_i| \left| \beta_j^{(i)} \right| p_{ijk} \leq \sum_{ijk} |a_i| p_{ijk} = ||a||_1$.

For the PPTC part, a very similar proof applies. Alternatively, it also follows from the more general statement in Proposition 5.10 that we will prove later.

⁵We note that Lemma 3.29 cannot be invoked for LO, since it doesn't fulfill the axiom of free instruments. Therefore, this statement needs to be re-proven explicitly for LO.

5.1.2 Utility of classical side information

This section will focus on understanding the utility of classical side information from intermediate measurements for the task of QPS of nonlocal operations. More mathematically, this means that we will compare γ_{LO^*} with γ_{LO} , γ_{LOCC^*} with γ_{LOCC} , γ_{SEPC^*} with γ_{PPTC} . Recall that in the special case of states, we already saw previously in Lemma 5.5 that intermediate measurements don't impact the extent. We now study the more general case of superoperators.

In the LO setting, it turns out that classical side information is absolutely crucial to realize space-like cuts. In fact, the only operations that we can simulate without intermediate measurements are non-signalling channels, which are essentially useless for the purpose of circuit knitting.

Definition 5.6: A bipartite superoperator $\mathcal{E} \in \mathrm{HP}(AB \to A'B')$ is called non-signalling if $\mathrm{tr}_{B'}[\mathcal{E}(\rho)] = \mathrm{tr}_{B'}[\mathcal{E}(\sigma)]$ for all $\rho_{AB}, \sigma_{AB} \in \mathrm{D}(AB)$ with $\mathrm{tr}_{B}[\rho_{AB}] = \mathrm{tr}_{B}[\sigma_{AB}]$ and vice versa if we exchange of $A \leftrightarrow B$, $A' \leftrightarrow B'$. We denote the set of all non-signalling superoperators by $\mathrm{NS}(A; B \to A; B')$.

The idea is that if ρ and σ are indistinguishable to the party A, then the application of \mathcal{E} cannot change the statistics of any measurement that A performs on its side afterwards.

Lemma 5.7: The set of superoperators that exhibit a QPD under LO and LO* are

$$\operatorname{span}_{\mathbb{R}}(\operatorname{LO}(A; B \to A'; B')) = \mathbb{R}\operatorname{HPTP}(A \to A') \otimes \mathbb{R}\operatorname{HPTP}(B \to B')$$

$$= \operatorname{NS}(A; B \to A'; B')$$
(5.20)
$$= (5.21)$$

and

$$\operatorname{span}_{\mathbb{R}}(\operatorname{LO}^{\star}(A; B \to A'; B')) = \operatorname{HP}(AB \to A'B'). \tag{5.22}$$

Recall the definition of RHPTP in Lemma 4.1. Note that the tensor product here should be understood as the *real* tensor product of two subspaces of the *real* vector space of Hermitian-preserving superoperators.

Proof. To show Equation (5.20), notice on one hand that $LO(A; B \to A'B') \subset \mathbb{R}HPTP(A \to A') \otimes \mathbb{R}HPTP(B \to B')$. Conversely, by Lemma 4.1 any pure tensor in $\mathbb{R}HPTP(A \to A') \otimes \mathbb{R}HPTP(B \to B')$ can be written as a linear combination of LO channels. For the proof of Equation (5.21) we refer to [71, Theorem 13] or [37, Theorem 5.1].

Next, we consider Equation (5.22). Clearly, $\operatorname{span}_{\mathbb{R}}(\operatorname{LO}^{\star}) \subset \operatorname{HP}$, so it remains to show the converse. Because $\operatorname{HP}(AB \to A'B') = \operatorname{HP}(A \to A') \otimes \operatorname{HP}(B \to B')$, the desired statement follows from Lemma 4.1.

In contrast, classical side information is not required in the setting of LOCC to simulate all CPTP maps.

Lemma 5.8: The set of superoperators that exhibit a QPD w.r.t. LOCC and $LOCC^*$ are

$$\operatorname{span}_{\mathbb{R}}(\operatorname{LOCC}(A; B \to A'; B')) = \mathbb{R}\operatorname{HPTP}(AB \to A'B') \tag{5.23}$$

and

$$\operatorname{span}_{\mathbb{R}}(\operatorname{LOCC}^{\star}(A; B \to A'; B')) = \operatorname{HP}(AA' \to BB'). \tag{5.24}$$

Notice that classical side information is strictly necessary to simulate non-trace-preserving maps. This is unsurprising, and in fact already implied by the result in Lemma 4.1. In this context, one can interpret Lemma 5.8 as showing that LOCC and LOCC* span the "largest possible" spaces achievable by decomposition sets without and with side information.

Proof. By Lemma 4.1, we know that the inclusion \subset must hold for Equations (5.23) and (5.24). The converse side for Equation (5.24) follows from Equation (5.22) and $LO^* \subset LOCC^*$. It thus remains to show \supset in Equation (5.23).

By Lemma 4.1, this can be proven by showing that any $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ lies in $\operatorname{span}_{\mathbb{R}}(\operatorname{LOCC})$. We assume without loss of generality that A = A' and B = B' are multi-qubit systems, since any quantum system can be embedded into a large enough multi-qubit system. By the Stinespring dilation theorem, we can thus write \mathcal{E} as a unitary evolution on some enlarged Hilbert space

$$\mathcal{E}(\rho_{AB}) = \operatorname{tr}_{E} \left[U_{ABE}(\rho_{AB} \otimes |0\rangle\langle 0|_{E}) U_{ABE}^{\dagger} \right]$$
 (5.25)

where $U_{ABE} \in U(ABE)$. It is well-known that any unitary can be decomposed into CNOT gates and single-qubit unitaries [7]. We have already seen that CNOT gates can be decomposed into LOCC operations in Figure 5.3. By decomposing all CNOT gates in the circuit of U_{ABE} that act nonlocally across A and BE, we can write $Ad[U_{ABE}]$ as a QPD of channels in $LOCC(A; BE \rightarrow A; BE)$. By tracing out E, these channels induce a QPD of \mathcal{E} into $LOCC(A; B \rightarrow A; B)$.

As a direct consequence of Equations (5.12) and (5.13) as well as Lemmas 4.1 and 5.8, the analogous result also holds in the SEPC and PPTC settings.

Corollary 5.9: The set of superoperators that exhibit a QPD w.r.t. SEPC and PPTC are

$$\operatorname{span}_{\mathbb{R}}(\operatorname{SEPC}(A;B\to A';B')) = \operatorname{span}_{\mathbb{R}}(\operatorname{PPTC}(A;B\to A';B')) = \mathbb{R}\operatorname{HPTP}(AB\to A'B')$$
 (5.26) and

$$\operatorname{span}_{\mathbb{R}}(\operatorname{SEPC}^{\star}(A;B\to A';B')) = \operatorname{span}_{\mathbb{R}}(\operatorname{PPTC}^{\star}(A;B\to A';B')) = \operatorname{HP}(AA'\to BB'). \tag{5.27}$$

In fact, for this case there is a simple argument that the quasiprobability extent of any quantum channel is identical with and without classical side information.

Proposition 5.10

For any $\mathcal{E} \in \mathbb{R}\mathrm{HPTP}(AB \to A'B')$ one has $\gamma_{\mathrm{PPTC}}(\mathcal{E}) = \gamma_{\mathrm{PPTC}^{\star}}(\mathcal{E})$ and $\gamma_{\mathrm{SEPC}}(\mathcal{E}) = \gamma_{\mathrm{SEPC}^{\star}}(\mathcal{E})$.

Proof. In the following, S may denote either SEPC or PPTC. Clearly, $\gamma_S \geq \gamma_{S^*}$ by Lemma 3.10, so we only need to show the converse. Consider a general QPD of \mathcal{E} w.r.t. S^* . Without loss of generality, we can assume it to only have one element by Lemma 3.14, i.e., $\mathcal{E} = \kappa \mathcal{F}$, $\mathcal{F} \in S^*$. By the characterization of S^* (either Equation (5.9) or Equation (5.11)), we can write $\mathcal{F} = \mathcal{F}^+ - \mathcal{F}^-$. If either $\mathcal{F}^+ = 0$ or $\mathcal{F}^- = 0$, then we already have a QPD into S. Otherwise, observe that both $J_{\mathcal{F}^+} + J_{\mathcal{F}^-}$ and $J_{\mathcal{F}^+} - J_{\mathcal{F}^-}$ are in RHPTP. Clearly, this is only possible if both $J_{\mathcal{F}^+}$ and $J_{\mathcal{F}^-}$ are themselves in RHPTP. The following is therefore a valid QPD of \mathcal{E} w.r.t. S

$$\mathcal{E} = \kappa \operatorname{tr}[J_{\mathcal{F}^{+}}] \frac{\mathcal{F}^{+}}{\operatorname{tr}[J_{\mathcal{F}^{+}}]} - \kappa \operatorname{tr}[J_{\mathcal{F}^{-}}] \frac{\mathcal{F}^{-}}{\operatorname{tr}[J_{\mathcal{F}^{-}}]}$$
(5.28)

and the 1-norm of its coefficients is given by $\kappa(\operatorname{tr}[J_{\mathcal{F}^+}] + \operatorname{tr}[J_{\mathcal{F}^-}]) = \kappa$.

In summary, we have already established some interesting observations from rather simple mathematical arguments. While classical side information is absolutely crucial for space-like cuts in the LO setting, it is completely unnecessary for SEPC and PPTC (unless one wants to simulate non-trace-preserving maps). The setting LOCC lies somewhere in the middle, and we know that intermediate measurements are not required in order to enable QPS of general CPTP maps. However, it remains an open question whether there exists some CPTP map $\mathcal E$ for which $\gamma_{\text{LOCC}^*}(\mathcal E)$ is strictly smaller than $\gamma_{\text{LOCC}}(\mathcal E)$.

5.1.3 Quasiprobability extent of states

Before we delve into the most general setting by studying the quasiprobability extent of arbitrary channels, we first focus on the comparatively easier task of characterizing the quasiprobability extent $\gamma_{\text{SEP}(A;B)}(\rho_{AB})$ for bipartite states $\rho_{AB} \in D(AB)$ (recall from Lemma 5.5 that $\gamma_{\text{LOCC}^*}(\rho) = \gamma_{\text{LO}^*}(\rho) = \gamma_{\text{SEP}}(\rho)$). In many ways, SEP is a well-behaved set and many properties from Chapter 3 are directly applicable to γ_{SEP} . For instance, since SEP is convex and closed, we can write its quasiprobability extent as

$$\gamma_{\text{SEP}}(\rho_{AB}) = \min\{a^+ + a^- | \rho = a^+ \sigma^+ - a^- \sigma^-, \sigma^{\pm} \in \text{SEP}(A; B), a^{\pm} \ge 0\}$$
(5.29)

(see Lemmas 3.11 and 3.25). Furthermore, γ_{SEP} is invariant under local unitaries (see Corollary 3.33) and it can be expressed in terms of a robustness measure

$$\gamma_{\text{SEP}}(\rho_{AB}) = 1 + 2R_{\text{SEP}}(\rho_{AB}) \tag{5.30}$$

(see Lemma 3.17) where

$$R_{\text{SEP}}(\rho_{AB}) := \min\{t \ge 0 | \frac{\rho_{AB} + t\sigma_{AB}}{1 + t} \in \text{SEP}(A; B) \text{ for some } \sigma_{AB} \in \text{SEP}(A; B)\}$$
 (5.31)

The quantity R_{SEP} is commonly referred to as the *robustness of entanglement* and it was first introduced by Vidal and Tarrach in Ref. [166]. The original motivation for the robustness of entanglement was to capture the amount of (separable) noise that a certain entangled state can tolerate before it becomes itself separable. Since then, it has found applications in various results on entanglement manipulation [19], [20], [100], [103].

Furthermore, we note that γ_{SEP} is Lipschitz continuous.

Proposition 5.11

Let ω , $\tau \in H(AB)$ and $d_A := \dim(A)$, $d_B := \dim(B)$. Then

$$\forall \omega, \tau \in \mathcal{H}(AB) : |\gamma_{\text{SEP}}(\omega) - \gamma_{\text{SEP}}(\tau)| \le 2^{\frac{3}{2}(\lceil \log(d_A) \rceil + \lceil \log(d_B) \rceil)} \|\omega - \tau\|_2. \quad (5.32)$$

Proof. We first consider the case where A and B are n-qubit and m-qubit system. By Lemma 3.15 we can consider $\gamma_{\text{aconv}(\text{SEP})}$ instead of γ_{SEP} . Denote the n+m-qubit Pauli strings by $\{Q_0,\ldots,Q_{4^{n+m}-1}\}$. Notice that the normalized Pauli matrices $\frac{1}{2^{n+m}}Q_i$ all lie in aconv(SEP(A;B)), since they each have an eigenbasis of product states and eigenvalues -1 and 1. Furthermore, they form an orthogonal basis of H(AB), so we can apply Lemma 3.23. The Gram matrix is diagonal and its smallest eigenvalues is $2^{-2(n+m)}$.

We now turn our attention to the general case. Consider embedding A and B into large enough n-qubit and m-qubit systems $\bar{A} = A \oplus A^{\perp}$ and $\bar{B} = B \oplus B^{\perp}$. Define by $\operatorname{Emb}_{A \to \bar{A}}$ and $\operatorname{Emb}_{B \to \bar{B}}$ the CPTP maps that naturally embed a state into the larger space. Correspondingly, denote by $\operatorname{Proj}_{\bar{A} \to A}$ and $\operatorname{Proj}_{\bar{B} \to B}$ the CPTN maps $\rho \mapsto \Pi \rho \Pi$ where Π is the projection onto A or B respectively. Define the extended state $\omega \in \operatorname{H}(\bar{A}\bar{B})$ by

$$\bar{\omega} := (\operatorname{Emb}_{A \to \bar{A}} \otimes \operatorname{Emb}_{B \to \bar{B}})(\omega) \tag{5.33}$$

which clearly fulfills

$$\omega = (\operatorname{Proj}_{\bar{A} \to A} \otimes \operatorname{Emb}_{\bar{B} \to B})(\bar{\omega}). \tag{5.34}$$

Since both $(\operatorname{Emb}_{A\to \bar{A}}\otimes\operatorname{Emb}_{B\to \bar{B}})(\omega)$ and $(\operatorname{Proj}_{\bar{A}\to A}\otimes\operatorname{Emb}_{\bar{B}\to B})$ lie in $\operatorname{LOCC}^{\star}$, the chaining property in Corollary 3.32 implies that $\gamma_{\operatorname{SEP}}(\omega)=\gamma_{\operatorname{SEP}}(\bar{\omega})$. Finally, notice that that by appropriately choosing a basis, the embedding does not change the non-zero matrix entries of a matrix. Hence, for two extended states one has

$$\|\bar{\omega} - \bar{\tau}\|_2 = \|\omega - \tau\|_2 \ . \tag{5.35}$$

Hence,

$$|\gamma_{\text{SEP}}(\omega) - \gamma_{\text{SEP}}(\tau)| = |\gamma_{\text{SEP}}(\bar{\omega}) - \gamma_{\text{SEP}}(\bar{\tau})|$$
 (5.36)

$$\leq 2^{\frac{3}{2}(n+m)} \|\bar{\omega} - \bar{\tau}\|_{2}$$
 (5.37)

$$=2^{\frac{3}{2}(n+m)}\|\omega-\tau\|_{2}. \tag{5.38}$$

While $\gamma_{\rm SEP}$ is the quantity that we aim to characterize due to its operational importance, we will also study the quantity $\gamma_{\rm PPT}$. It provides a lower bound to $\gamma_{\rm SEP}$ and it is efficiently computable: Since PPT is itself also convex and closed, we can analogously write

$$\gamma_{\text{PPT}}(\rho_{AB}) = \min\{a^+ + a^- | \rho = a^+ \sigma^+ - a^- \sigma^-, \sigma^{\pm} \in \text{PPT}(A; B), a^{\pm} \ge 0\}$$
 (5.39)

which through a simple substitution can be expressed in the form of a SDP

$$\gamma_{\text{PPT}}(\rho) = \begin{cases}
\min_{\sigma^{\pm} \in P(AB)} & \text{tr}[\sigma^{+}] + \text{tr}[\sigma^{-}] \\
\text{s.t.} & \rho = \sigma^{+} - \sigma^{-} \\
& (\text{id}_{A} \otimes T_{B})(\sigma^{\pm}) \succcurlyeq 0
\end{cases} (5.40)$$

This SDP characterization enables the efficient numerical evaluation of γ_{PPT} and also serves as a valuable analytical tool for proving the optimality of an QPD.

Lemma 5.12: The following is a valid dual SDP formulation of Equation (5.40).

$$\gamma_{\text{PPT}}(\rho) = \begin{cases} \max_{X \in \mathcal{H}(AB), Z^{\pm} \in \mathcal{P}(AB)} & \text{tr}[X\rho] \\ \text{s.t.} & -\mathbb{1}_{AB} + (\text{id}_A \otimes T_B)(Z^{-}) \preccurlyeq X \\ & X \preccurlyeq \mathbb{1}_{AB} - (\text{id}_A \otimes T_B)(Z^{+}) \end{cases}$$
(5.41)

Proof. In the following, we will denote the unnormalized Hilbert-Schmidt inner product by $\langle A, B \rangle := \text{tr}[A^{\dagger}B]$. The Lagrangian is given by

$$L(\sigma^{\pm}; X, Z^{\pm}) = \operatorname{tr}[\sigma^{+}] + \operatorname{tr}[\sigma^{-}] - \langle X, \rho - \sigma^{+} + \sigma^{-} \rangle$$

$$- \langle Z^{+}, (\operatorname{id}_{A} \otimes T_{B})(\sigma^{+}) \rangle - \langle Z^{-}, (\operatorname{id}_{A} \otimes T_{B})(\sigma^{-}) \rangle$$

$$= -\langle X, \rho \rangle + \langle \sigma^{+}, \mathbb{1}_{AB} + X - (\operatorname{id}_{A} \otimes T_{B})(Z^{+}) \rangle$$

$$+ \langle \sigma^{-}, \mathbb{1}_{AB} - X - (\operatorname{id}_{A} \otimes T_{B})(Z^{-}) \rangle. \tag{5.42}$$

The dual function $g(X, Z^{\pm}) := \inf_{\sigma^{\pm}} L(\sigma^{\pm}; X, Z^{\pm})$ can only take values greater than $-\infty$ if $\mathbb{1}_{AB} \pm X - (\mathrm{id}_A \otimes T_B)(Z^{\pm}) \ge 0$. We thus get the dual SDP

$$\gamma_{\mathrm{PPT}}(\rho) = \begin{cases} \max_{X, \in \mathrm{H}(AB)Z^{\pm} \in \mathrm{P}(AB)} & -\operatorname{tr}[X\rho] \\ \mathrm{s.t.} & -\mathbb{1}_{AB} + (\mathrm{id}_A \otimes T_B)(Z^+) \preccurlyeq X \\ & X \preccurlyeq \mathbb{1}_{AB} - (\mathrm{id}_A \otimes T_B)(Z^-) \,. \end{cases}$$
(5.43)

A simple variable substitution $X \to -X$ yields the desired form of the dual SDP. \square

The rest of this section is structured as follows. First, we will consider the quasiprobability extent of pure entangled states. We will see that in this case, it is possible to find a closed-form analytical expression of $\gamma_{\rm SEP}$ in terms of the Schmidt coefficients. Then, we will turn our attention to mixed states, which are generally more complicated to handle. We will deal with the one-shot and asymptotic regimes separately form each other, as they will entail some different techniques.

5.1.3.1 Pure states

Since the quasiprobability extent w.r.t. SEP is invariant under local unitaries, $\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|_{AB})$ can clearly only depend on the Schmidt coefficients of a bipartite state $|\psi\rangle_{AB}$. In fact, γ_{SEP} takes a simple form and one can even find an expression for an achieving QPD.

Theorem 5.13: [166]

Let $|\psi\rangle_{AB}$ be a bipartite pure state with Schmidt decomposition

$$|\psi\rangle_{AB} = \sum_{j=1}^{r} u_j |f_j\rangle_A \otimes |g_j\rangle_B . \qquad (5.44)$$

Then

$$\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|_{AB}) = 2\left(\sum_{j} u_{j}\right)^{2} - 1 \tag{5.45}$$

and the QPD $|\psi\rangle\langle\psi|_{AB}=a^+\sigma^+-a^-\sigma^-$ achieves the optimal extent, where

$$a^{-} := (\sum_{j} u_{j})^{2} - 1, \quad a^{+} := (\sum_{j} u_{j})^{2},$$
 (5.46)

$$\sigma^{-} := \frac{1}{(\sum_{j} u_{j})^{2} - 1} \sum_{i \neq j} u_{i} u_{j} |f_{i}\rangle\langle f_{i}|_{A} \otimes |g_{j}\rangle\langle g_{j}|_{B} , \qquad (5.47)$$

$$\sigma^{+} := \frac{1}{(\sum_{j} u_{j})^{2}} \left(|\psi\rangle\langle\psi|_{AB} + ((\sum_{j} u_{j})^{2} - 1)\sigma^{-} \right). \tag{5.48}$$

The proof follows from two observations. First, it can be shown that $|\psi\rangle\langle\psi|=a^+\sigma^+-a^-\sigma^-$ is a valid QPD w.r.t. SEP. Secondly, using the SDP formulation of $\gamma_{\rm PPT}$, one can verify that this QPD also achieves the quasiprobability extent w.r.t. PPT. Combining these two insights gives us matching lower and upper bounds for $\gamma_{\rm SEP}(|\psi\rangle\langle\psi|)$, hence finishing the proof. In the following, we show both of these steps separately.

We briefly note that the proof of the converse statement differs from the original one introduced in Ref. [166]. Our approach based on convex programming simplifies the proof, and to our knowledge has not been proposed in literature before.

Lemma 5.14: The states σ^{\pm} in Equations (5.47) and (5.48) are separable.

Proof. Clearly, σ^- is separable, so we only need to consider σ^+ . We will show that

$$\sigma^{+} = \frac{1}{2^{r} - 1} \sum_{j=1}^{2^{r} - 1} |\phi_{j}\rangle\langle\phi_{j}| \otimes |\tau_{j}\rangle\langle\tau_{j}|$$
 (5.49)

where

$$\left|\phi_{j}\right\rangle := \frac{1}{\sqrt{\sum_{l} u_{l}}} \sum_{k=1}^{r} \sqrt{u_{k}} \exp\left(2\pi i j \frac{2^{k-1} - 1}{2^{r} - 1}\right) \left|f_{k}\right\rangle , \qquad (5.50)$$

$$\left|\tau_{j}\right\rangle := \frac{1}{\sqrt{\sum_{l} u_{l}}} \sum_{k=1}^{r} \sqrt{u_{k}} \exp\left(-2\pi i j \frac{2^{k-1}-1}{2^{r}-1}\right) \left|g_{k}\right\rangle. \tag{5.51}$$

To show this, we consider the matrix elements of $\sum\limits_{j=1}^{2^r-1}|\phi_j\rangle\langle\phi_j|\otimes|\tau_j\rangle\langle\tau_j|$ in the Schmidt basis

$$(\langle f_a | \otimes \langle g_b |) \sum_{j=1}^{2^r - 1} |\phi_j \rangle \langle \phi_j | \otimes |\tau_j \rangle \langle \tau_j | (|f_c \rangle \otimes |g_d \rangle)$$

$$= \frac{\sqrt{u_a u_b u_c u_d}}{(\sum_l u_l)^2} \sum_{j=1}^{2^r - 1} \exp\left(\frac{2\pi i j}{2^r - 1} (2^{a-1} + 2^{d-1} - 2^{b-1} - 2^{c-1})\right). \tag{5.52}$$

Notice that $|2^{a-1} + 2^{d-1} - 2^{b-1} - 2^{c-1}|$ is always smaller than $2^r - 1$. Due to a standard geometric series argument, the matrix elements are zero unless $2^{a-1} + 2^{d-1} - 2^{b-1} - 2^{c-1} = 0$. This in turn can only occur when either a = b and c = d or otherwise when a = c and b = d. In summary, the only non-vanishing matrix elements are

$$(\langle f_a | \otimes \langle g_a |) \sum_{j=1}^{2^r - 1} |\phi_j \rangle \langle \phi_j | \otimes |\tau_j \rangle \langle \tau_j | (|f_b \rangle \otimes |g_b \rangle)$$
 (5.53)

$$= (\langle f_a | \otimes \langle g_b |) \sum_{j=1}^{2^r - 1} |\phi_j \rangle \langle \phi_j | \otimes |\tau_j \rangle \langle \tau_j | (|f_a \rangle \otimes |g_b \rangle)$$
 (5.54)

$$=\frac{u_a u_b}{(\sum_l u_l)^2} \tag{5.55}$$

which corresponds precisely to the matrix representation of σ^+ .

Lemma 5.15: Let $|\psi\rangle_{AB}$ be a bipartite pure state with Schmidt coefficients $(u_j)_j$. Then

$$\gamma_{\text{PPT}}(|\psi\rangle\langle\psi|_{AB}) = 2\left(\sum_{j} u_{j}\right)^{2} - 1 \tag{5.56}$$

and the QPD $\rho = a^+ \sigma^+ - a^- \sigma^-$ in Theorem 5.13 achieves $\gamma_{\rm PPT}(|\psi\rangle\langle\psi|_{AB})$.

Proof. Since the QPD $\rho = a^+\sigma^+ - a^-\sigma^-$ is a valid QPD w.r.t. SEP (see Lemma 5.14), it is automatically also a valid QPD w.r.t. PPT. It remains to show its optimality. We accomplish this by finding a feasible solution of the dual SDP given in Lemma 5.12 which achieves the same value of its objective.

With a slight abuse of notation, we write the Schmidt decomposition of $|\psi\rangle_{AB}$ as

$$|\psi\rangle_{AB} = \sum_{j=1}^{r} u_j |j\rangle_A \otimes |j\rangle_B .$$
 (5.57)

where $|i\rangle_A$ and $|i\rangle_B$ are orthonormal basis states on A and B respectively. We now pick

$$X = 2\sum_{i,j} |ii\rangle \langle jj|_{AB} - \mathbb{1}_{AB}$$
 (5.58)

$$Z^{-} = 0 (5.59)$$

and

$$Z^{+} = \sum_{i < j} (|ij\rangle_{AB} - |ji\rangle_{AB}) (\langle ij|_{AB} - \langle ji|_{AB}).$$
 (5.60)

Clearly, $Z^- \geq 0$ and $X \geq -1_{AB}$. The final constraint can be shown as follows

$$\mathbb{1}_{AB} - (\mathrm{id}_A \otimes T_B)(Z^+) = \mathbb{1}_{AB} - \sum_{i < j} (\mathrm{id}_A \otimes T_B) \left((|ij\rangle - |ji\rangle) (\langle ij| - \langle ji|) \right) \tag{5.61}$$

$$= \mathbb{1}_{AB} - \sum_{i < j} (|ij\rangle\langle ij| + |ji\rangle\langle ji| - |ii\rangle\langle jj| - |jj\rangle\langle ii|)$$
 (5.62)

$$= \mathbb{1}_{AB} - 2\sum_{i \neq j} (|ij\rangle\langle ij| - |ii\rangle\langle jj|)$$
 (5.63)

$$= \mathbb{1}_{AB} - 2\mathbb{1}_{AB} + 2\sum_{i} |ii\rangle\langle ii| + 2\sum_{i\neq j} |ii\rangle\langle jj|$$
 (5.64)

$$= -\mathbb{1}_{AB} + 2\sum_{i,j} |ii\rangle \langle jj| \tag{5.65}$$

$$=X \tag{5.66}$$

It remains to evaluate the objective function for this feasible point:

$$tr[X\rho] = 2\sum_{i,j} tr[|ii\rangle \langle jj|_{AB} \rho] - tr[\rho]$$
(5.67)

$$=2\sum_{i,j}\langle ii|\rho|jj\rangle-1\tag{5.68}$$

$$=2\sum_{i,j,k,l}u_{k}u_{l}\left\langle ii\right\vert \left\vert kk\right\rangle \left\langle ll\right\vert \left\vert jj\right\rangle -1\tag{5.69}$$

$$= 2\left(\sum_{i} u_{i}\right)^{2} - 1. \tag{5.70}$$

An immediate consequence of Theorem 5.13 is that we can also evaluate the regularized quasiprobability extent for pure states.

Corollary 5.16: Let $|\psi\rangle_{AB}$ be a bipartite pure state on AB with Schmidt coefficients $(u_i)_{i=1,\dots,r}$. Then

$$\overline{\gamma}_{\text{SEP}}(|\psi\rangle\langle\psi|_{AB}) = 1 + R_{\text{SEP}}(|\psi\rangle\langle\psi|_{AB}) = \left(\sum_{j} u_{j}\right)^{2} = 2^{H_{1/2}(\text{tr}_{B}[|\psi\rangle\langle\psi|_{AB}])}$$
(5.71)

where $H_{1/2}$ is the quantum Rényi-1/2 entropy.

Notice that this regularized quasiprobability extent has halved distance to 1 compared to $\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|) = 1 + 2R_{\text{SEP}}(|\psi\rangle\langle\psi|)$. The quantum Rényi entropy of order

 $\alpha \in (0,1) \cup (1,\infty)$ is defined as

$$H_{\alpha}(\rho) := \frac{1}{1-\alpha} \log \operatorname{tr}[\rho^{\alpha}]. \tag{5.72}$$

The specialization $H_{1/2}$ is commonly referred to as the *max-Entropy* (also denoted H_{max}), though older literature used to refer to H_0 by this name.

Proof. The Schmidt coefficients of $|\psi\rangle^{\otimes n}$ are given by the collection $\left(\prod_{i=1}^n u_{k_i}\right)_{k\in\mathbb{F}_r^n}$ where $\mathbb{F}_r:=\{1,\ldots,r\}$. By Theorem 5.13 the quasiprobability extent is thus

$$\gamma_{\text{SEP}}(|\psi\rangle\langle\psi|^{\otimes n}) = 2\left(\sum_{k\in\mathbb{F}_r^n}\prod_{i=1}^n u_{k_i}\right)^2 - 1 \tag{5.73}$$

$$=2\left(\sum_{i=1}^{r}u_{i}\right)^{2n}-1. \tag{5.74}$$

Observing that $\sum_i u_i \ge \sqrt{\sum_i u_i^2} = 1$, we see that taking the limit $\lim_{n \to \infty} (\cdot^{1/n})$ on both sides precisely gives us the desired statement.

Example 5.17. Consider a two-qudit system AB, dim(A) = dim(B) = d and the maximally entangled state

$$|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle_A \otimes |i\rangle_B$$
 (5.75)

for some choice of orthonormal bases on A and B. Then

$$\gamma_{\text{SEP}}(|\Psi\rangle\langle\Psi|) = 2\left(\sum_{i} \frac{1}{\sqrt{d}}\right)^{2} - 1 = 2d - 1 \tag{5.76}$$

and

$$\overline{\gamma}_{SEP}(|\Psi\rangle\langle\Psi|) = \lim_{n \to \infty} (2d^n - 1)^{1/n} = d.$$
 (5.77)

Since any state in D(AB) can be reached by applying some LOCC protocol on $|\Psi\rangle\langle\Psi|$, $\gamma_{\text{SEP}}(|\Psi\rangle\langle\Psi|) = 2^d - 1$ is the maximal value that the quasiprobability γ_{SEP} takes over all states in D(AB) (recall Corollary 3.32).

For d=2, the following QPD achieves the quasiprobability extent $\gamma_{\rm SEP}(|\Psi\rangle\!\langle\Psi|)=3$

$$|\Psi\rangle\langle\Psi| = \frac{1}{2} (|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1| - |+\rangle\langle +| \otimes |-\rangle\langle -| -|-\rangle\langle -| \otimes |+\rangle\langle +| + |i+\rangle\langle i+| \otimes |i-\rangle\langle i-| + |i-\rangle\langle i-| \otimes |i+\rangle\langle i+|)$$

$$(5.78)$$

where
$$|\pm\rangle \coloneqq (|0\rangle \pm |1\rangle)/\sqrt{2}$$
 and $|i\pm\rangle \coloneqq (|0\rangle \pm i |1\rangle)/\sqrt{2}$.

5.1.3.2 Mixed states in the single-shot regime

Previously, we have observed that the quasiprobability extent of an arbitrary pure state w.r.t. SEP can be characterized by a simple formula involving the corresponding Schmidt coefficients. Unfortunately, it is highly unlikely that an analogous result exists for general mixed states. It is known that detecting entanglement is generally NP-hard [61], [70], so the computation of any faithful entanglement measure (such as the $\gamma_{\rm SEP}$, recall Lemma 3.27) must be at least as difficult. As such, we are fundamentally restricted to finding computationally inefficient characterizations of the quasiprobability extent or computationally efficient lower and upper bounds.

We have already seen that γ_{PPT} is one such lower bound which reaches the optimal value for any pure state. As a direct consequence of Lemma 5.3, it is also optimal for small systems.

Corollary 5.18: Let
$$\dim(A)\dim(B) \le 6$$
 and $\omega \in \mathrm{H}(AB)$. Then
$$\gamma_{\mathrm{SEP}}(\omega) = \gamma_{\mathrm{PPT}}(\omega)\,. \tag{5.79}$$

Example 5.19. Consider the depolarized Bell state across two qubits A and B

$$\rho_p := (1 - p) |\Psi\rangle\langle\Psi|_{AB} + p\frac{1}{4}\mathbb{1}_{AB}$$

$$(5.80)$$

where $|\Psi\rangle_{AB}=(|00\rangle+|11\rangle)/\sqrt{2}$ and $p\in[0,1]$. Because of Corollary 5.18, we can numerically evaluate $\gamma_{\rm SEP}(\rho_p)$. Up to numerical precision, we observe that

$$\gamma_{\text{SEP}}(\rho_{p}) = \max(1, 3(1-p)).$$
 (5.81)

 \triangle

However, there are states for which the lower bound does not achieve γ_{SEP} . This is exemplified by any entangled PPT state ρ , such as the one introduced in Example 5.4. Clearly, one has $\gamma_{\text{PPT}}(\rho) = 1 < \gamma_{\text{SEP}}(\rho)$ in this situation.

Conversely, to find upper bounds to γ_{SEP} , one approach is to simply guess a finite decomposition set $S \subset \text{SEP}(A;B)$ and use the linear programming approach from Section 3.6 to evaluate $\gamma_S(\rho)$. By Lemma 3.10, $\gamma_S(\rho) \ge \gamma_{\text{SEP}}(\rho)$.

Example 5.20. Consider the maximally entangled state $|\Psi\rangle_{AB}$ across two systems A, B which are each two-qubit systems. We know by Theorem 5.13 that $\gamma_{\text{SEP}}(|\Psi\rangle\langle\Psi|)=7$. We now consider the depolarized version of $|\Psi\rangle\langle\Psi|$ with strength p

$$\rho_p := (1 - p) |\Psi\rangle\langle\Psi|_{AB} + p \frac{1}{16} \mathbb{1}_{AB}. \tag{5.82}$$

To obtain an upper bound to the quasiprobability extent, we choose the decomposition basis S which contains all product states $\sigma \otimes \tau$ where σ and τ are either one of the four 2-qubit maximally entangled states, or themselves product states of

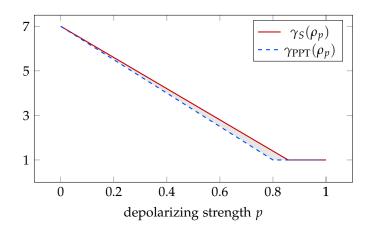


Figure 5.4: Numerical evaluation of γ_{PPT} and γ_S on a depolarized maximally entangled 4-qubit state ρ_p . The finite decomposition set S is heuristically chosen. γ_{SEP} must lie in the shaded region.

 $|0\rangle$, $|1\rangle$, $|+\rangle$, $|-\rangle$, $|i+\rangle$ or $|i-\rangle$. The numerical evaluation of $\gamma_S(\rho_p)$ and $\gamma_{PPT}(\rho_p)$ is depicted in Figure 5.4. Up to numerical precision, we observe

$$\gamma_{\text{PPT}}(\rho_p) = \max(1, 7 - \frac{15}{2}p) \text{ and } \gamma_S(\rho_p) = \max(1, 7(1-p)).$$
 (5.83)

 \triangle

In the following, we will introduce a more principled approach to characterize the quasiprobability extent γ_{SEP} by expressing it in terms of a converging hierarchy of SDPs. This approach is based on the well-known Doherty-Parrilo-Spedalieri (DPS) hierarchy of approximations for the set of separable states [51]. We briefly summarize its most important ideas below.

The main workhorse of the DPS hierarchy is the notion of extendibility. Let $k \in \mathbb{N}$ and $\rho_{AB} \in D(AB)$ be a bipartite state. Let B_1, \ldots, B_k be identical copies of the system B. We say ρ_{AB} is k-extendible w.r.t. B if there exists a state $\tilde{\rho}_{AB_1...B_k} \in D(AB_1...B_k)$ which fulfills following properties:

• $\tilde{\rho}$ is invariant under permutations of the subsystems B_i , i.e.

$$\forall i \neq j : SWAP_{B_i,B_i} \tilde{\rho} SWAP_{B_i,B_i} = \tilde{\rho}$$
 (5.84)

where for any states $|\psi\rangle$, $|\phi\rangle$ the operation SWAP_{B_i,B_i} acts as

$$SWAP_{B_i,B_j} |\psi\rangle_{B_i} \otimes |\phi\rangle_{B_j} = |\phi\rangle_{B_i} \otimes |\psi\rangle_{B_j} . \qquad (5.85)$$

• $\rho_{AB} = \tilde{\rho}_{AB_1} := \operatorname{tr}_{B_2...B_k}[\tilde{\rho}_{AB_1...B_k}]$ where the equality should be understood by identifying B and B_1 .

Furthermore, we say ρ_{AB} is *k-PPT-extendible* w.r.t. *B* if additionally

• $\tilde{\rho}$ remains positive under any partial transposition w.r.t. to its n+1 sub-systems. Because of the permutation invariance, this amounts to

$$(\mathrm{id}_A \otimes T_{B_1} \otimes \mathrm{id}_{B_2} \otimes \mathrm{id}_{B_3} \otimes \cdots \otimes \mathrm{id}_{B_k}) (\tilde{\rho}) \geq 0 \tag{5.86}$$

$$\left(\mathrm{id}_{A}\otimes T_{B_{1}}\otimes T_{B_{2}}\otimes \mathrm{id}_{B_{3}}\otimes \cdots \otimes \mathrm{id}_{B_{k}}\right)\left(\tilde{\rho}\right)\succcurlyeq0\tag{5.87}$$

:

$$(\mathrm{id}_A \otimes T_{B_1} \otimes T_{B_2} \otimes T_{B_3} \otimes \cdots \otimes T_{B_k}) (\tilde{\rho}) \succcurlyeq 0 \tag{5.88}$$

For $k \ge 1$ we define

$$SEP_k(A; B) := \{ \rho \in D(AB) | \rho \text{ is } k\text{-PPT-extendible w.r.t. } B \}.$$
 (5.89)

Evidently, we have $SEP_1(A; B) = PPT(A; B)$. It can also be verified that for any $k \ge 1$ one has $SEP_{k+1} \subset SEP_k$. We leave the proof as an exercise for the reader and refer to Ref. [51] for more details. Clearly, any separable state $\sum_i p_i \rho_i \otimes \sigma_i$ is k-PPT-extendible for any $k \in \mathbb{N}$ as we can pick the extension $\sum_i p_i \rho_i \otimes \sigma_i^{\otimes k}$. Furthermore, it turns out that

$$SEP(A;B) = \bigcap_{k=1}^{\infty} SEP_k(A;B).$$
 (5.90)

This is a consequence of the completeness of the k-extendibility criterion for separability.

Proposition 5.21: [51, Theorem 1]

A bipartite state $\rho \in D(AB)$ is separable if and only if it is k-extendible w.r.t. B for all $k \in \mathbb{N}$.

The proof is based on techniques related to the quantum de Finetti theorem [38]. In fact, quantum de Finetti theorems can even provide an upper bound on the convergence speed of the hierarchy.

Proposition 5.22: [46, Theorem II.7']

Let $\rho \in D(AB)$ be a k-extendible state w.r.t. B. There exists a state $\sigma \in SEP(A;B)$ such that

$$\|\rho - \sigma\|_1 \le \frac{2\dim(B)^2}{k}$$
 (5.91)

We note that the PPT constraint in our definition of SEP_k is clearly not necessary for the convergence of the hierarchy. The PPT constraint makes the hierarchy strictly tighter and as such is often included to get better results for small k. We also note that there is some variation in literature w.r.t. how to the PPT constraint on the extension is defined. For instance, in Ref. [113], [114] the PPT criterion is not applied to every

possible bipartition of $AB_1 \dots B_k$, but instead only to one. In this thesis, we chose to follow the convention from [51].

By definition, the inclusion in the set $SEP_k(A; B)$ exhibits a semidefinite representation. As such, we can write $\gamma_{SEP_k}(\rho)$ of a state $\rho \in D(AB)$ as an SDP

$$\gamma_{\text{SEP}_{k}}(\rho) = \begin{cases} \min_{\sigma^{\pm} \in P(AB_{1}...B_{k})} & \text{tr}[\sigma^{+}] + \text{tr}[\sigma^{-}] \\ \text{s.t.} & \rho = \text{tr}_{B_{2}...B_{k}}[\sigma^{+}] - \text{tr}_{B_{2}...B_{k}}[\sigma^{-}], \\ & \text{SWAP}_{B_{i},B_{j}} \sigma^{\pm} \text{SWAP}_{B_{i},B_{j}} = \sigma^{\pm}, \\ & (\text{id}_{A} \otimes T_{B_{1}} \otimes \cdots \otimes \text{id}_{B_{k}}) (\sigma^{\pm}) \succcurlyeq 0, \end{cases}$$

$$\vdots$$

$$(\text{id}_{A} \otimes T_{B_{1}} \otimes \cdots \otimes T_{B_{k}}) (\sigma^{\pm}) \succcurlyeq 0$$

$$(5.92)$$

where we used the convexity of SEP_k , Lemma 3.11 and the same substitution as for Equation (5.40). We highlight again that the first level of this SDP hierarchy precisely retrieves the PPT lower bound, i.e., $\gamma_{SEP_1}(\rho) = \gamma_{PPT}(\rho)$. As such, the SDP hierarchy can be considered a refinement of the PPT lower bound, that provably converges in the limit of large k.

Theorem 5.23

For a bipartite operator $\rho \in D(AB)$ the quasiprobability extent $\gamma_{SEP_k}(\rho)$ monotonically approaches $\gamma_{SEP}(\rho)$ from below. Furthermore, the convergence speed is bounded by

$$\gamma_{\text{SEP}}(\rho) - \gamma_{\text{SEP}_k}(\rho) \le \frac{64 \dim(A)^{3/2} \dim(B)^{7/2} \min(\dim(A), \dim(B))}{k}.$$
 (5.93)

Proof. The set SEP_k is convex and compact (because it is the image of a compact set under the partial trace, which is continuous). Thus, by Lemmas 3.11 and 3.25, the quasiprobability extent of ρ w.r.t. $SEP_k(A; B)$ is achieved by some QPD of the form

$$\rho = (1+t)\rho^{+} - t\rho^{-} \tag{5.94}$$

for some $\rho^{\pm} \in SEP_k(A; B)$. By Proposition 5.22, there exist $\sigma^+, \sigma^- \in SEP(A; B)$ such that $\|\rho^{\pm} - \sigma^{\pm}\|_1 \le 2\dim(B)^2/k$. We can write

$$\rho - \Delta = (1+t)\sigma^+ - t\sigma^- \tag{5.95}$$

where $\Delta := (1+t)(\rho^+ - \sigma^+) - t(\rho^- - \sigma^-)$. Using the Lipschitz continuity of the quasiprobability extent (see Proposition 5.11), we have

$$\gamma_{\text{SEP}}(\rho) \le \gamma_{\text{SEP}}(\rho - \Delta) + L \|\Delta\|_{2} \tag{5.96}$$

$$\leq (1+2t) + L \|\Delta\|_2$$
 (5.97)

$$= \gamma_{\text{SEP}_k}(\rho) + L \left\| \Delta \right\|_2 \tag{5.98}$$

where $L \leq 2^{\frac{3}{2}(\log \dim(A) + \log \dim(B) + 2)} = 8(\dim(A) \dim(B))^{3/2}$. As such, the desired statement follows from

$$\|\Delta\|_2 \le \|\Delta\|_1 \tag{5.99}$$

$$\leq (1+2t) \left(\|\rho^{+} - \sigma^{+}\|_{1} + \|\rho^{-} - \sigma^{-}\|_{1} \right)$$
 (5.100)

$$\leq \gamma_{\text{SEP}_k}(\rho) \frac{4 \dim(B)^2}{k}$$

$$\leq \gamma_{\text{SEP}}(\rho) \frac{4 \dim(B)^2}{k}$$
(5.101)

$$\leq \gamma_{\text{SEP}}(\rho) \frac{4\dim(B)^2}{k} \tag{5.102}$$

$$\leq 2\min(\dim(A),\dim(B))\frac{4\dim(B)^2}{k} \tag{5.103}$$

where we used that the maximal quasiprobability extent γ_{SEP} is achieved by the maximally entangled state across *A* and *B* (recall Example 5.17).

We briefly illustrate an example where the SDP hierarchy provides an improvement over the PPT lower bound.

Example 5.24. We have already seen previously that entangled PPT states provide an example where $\gamma_{\rm SEP}$ is strictly larger than $\gamma_{\rm PPT}$. Taking ρ to be the two-qutrit state from Example 5.4, we numerically evaluate the SDP hierarchy $\gamma_{\text{SEP}_k}(\rho)$ for the first levels k = 1, 2, 3, 4.

As expected, the first level recovers the trivial bound $\gamma_{PPT}(\rho) = 1$. Interestingly, the hierarchy seemingly converges already at the second level.

5.1.3.3 Mixed states in the asymptotic regime

All previously introduced techniques to find bounds for γ_{SEP} rely on linear or semidefinite programming methods where the runtime scales polynomially with the dimension of the involved systems. Clearly, these approaches are ineffective in computing the regularized and asymptotic quasiprobability extent $\overline{\gamma}_{SEP}$, γ_{SEP}^{∞} as the involved system size tends to infinity. In the following, we explore some more analytical strategies to address this issue.

A natural starting point for such an endeavor are conversion bounds based on the so-called entanglement cost and distillable entanglement.

Definition 5.25: Let $\epsilon \geq 0$ and $\rho \in D(AB)$. The *single-shot entanglement cost* and *single-shot distillable entanglement* of ρ w.r.t. some set $S \subset CPTP$ are defined as

$$E_{C,S}^{(1),\epsilon}(\rho) := \min_{m \in \mathbb{N}} \{ \log m | \exists \mathcal{E} \in S : \frac{1}{2} \| \rho - \mathcal{E}(|\Psi_m\rangle\langle\Psi_m|) \|_1 \le \epsilon \}$$
 (5.104)

$$E_{D,S}^{(1),\epsilon}(\rho) := \max_{m \in \mathbb{N}} \{ \log m | \exists \mathcal{E} \in S : \frac{1}{2} \| |\Psi_m\rangle\langle\Psi_m| - \mathcal{E}(\rho) \|_1 \le \epsilon \}$$
 (5.105)

where $|\Psi_m\rangle$ denotes the maximally entangled state with local dimension m.

Recall from Corollary 3.32 that if two bipartite states $\rho, \sigma \in D(AB)$ are related by an LOCC map $\mathcal E$ in the sense that $\rho = \mathcal E(\sigma)$, then we know that $\gamma_{\rm SEP}(\rho) \leq \gamma_{\rm SEP}(\sigma)$. The quantities $E^\epsilon_{C,{\rm LOCC}}(\rho)$ and $E^\epsilon_{D,{\rm LOCC}}(\rho)$ characterize (up to an error ϵ) how many Bell pairs are required to create ρ , respectively, how many Bell pairs can be extracted from ρ . Combined with the aforementioned state conversion bound, this yields following inequality.

Lemma 5.26: Let $\epsilon \geq 0$. For any bipartite state $\rho \in D(AB)$ one has

$$2^{E_{D,LOCC}^{(1),\epsilon}(\rho)+1} - 1 - 4L\epsilon \le \gamma_{SEP}^{\epsilon}(\rho) \le 2^{E_{C,LOCC}^{(1),\epsilon}(\rho)+1} - 1. \tag{5.106}$$

where *L* is the Lipschitz constant of γ_{SEP} from Proposition 5.11.

Recall the definition of the smoothed quasiprobability extent in Definition 3.46.

Proof. Let us denote $c := E_{C,S}^{(1),\epsilon}(\rho)$ and $d := E_{D,S}^{(1),\epsilon}(\rho)$. There exist LOCC maps $\mathcal E$ and $\mathcal F$ as well as states ρ', ρ'' such that

$$\mathcal{E}(\rho) = \rho' \quad \text{and} \quad \rho' \approx_{\epsilon} |\Psi_d\rangle\langle\Psi_d| ,$$
 (5.107)

$$\mathcal{F}(|\Psi_c\rangle\langle\Psi_c|) = \rho'' \quad \text{and} \quad \rho'' \approx_{\epsilon} \rho.$$
 (5.108)

On one hand, we have

$$\gamma_{\text{SEP}}^{\epsilon}(\rho) \le \gamma_{\text{SEP}}(\rho'') \le \gamma_{\text{SEP}}(|\Psi_c\rangle\langle\Psi_c|),$$
(5.109)

and conversely,

$$\gamma_{\text{SEP}}^{\epsilon}(\rho) \ge \gamma_{\text{SEP}}^{\epsilon}(\rho') \tag{5.110}$$

$$= \gamma_{\text{SEP}}(\tau) \tag{5.111}$$

$$\geq \gamma_{\text{SEP}}(|\Psi_d\rangle\langle\Psi_d|) - L \|\tau - |\Psi_d\rangle\langle\Psi_d|\|_2 \tag{5.112}$$

$$\geq \gamma_{\text{SEP}}(|\Psi_d\rangle\langle\Psi_d|) - L(\|\tau - \rho'\|_1 + \|\rho' - |\Psi_d\rangle\langle\Psi_d|\|_1) \tag{5.113}$$

$$\geq \gamma_{\text{SEP}}(|\Psi_d\rangle\langle\Psi_d|) - 4L\epsilon \tag{5.114}$$

(5.115)

where τ is defined to be the achieving state of $\gamma_{\text{SEP}}^{\epsilon}(\rho')$. We used $\|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1 \le \|\rho - \sigma\|_1$ and the Lipschitz continuity of γ_{SEP} .

106

One of the most central quantities in entanglement theory are the optimal asymptotic conversion rates, which characterize the optimal number of Bell pairs that a state can be converted to/from in the asymptotic regime.

Definition 5.27: The asymptotic entanglement cost and asymptotic distillable entanglement of a bipartite state $\rho \in D(AB)$ w.r.t. $S \subset CPTP$ are defined as

$$E_{C,S}(\rho) := \lim_{\epsilon \to 0^+} \liminf_{n \to \infty} \frac{1}{n} E_{C,S}^{(1),\epsilon}(\rho^{\otimes n}) \quad \text{and} \quad E_{D,S}(\rho) := \lim_{\epsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} E_{D,S}^{(1),\epsilon}(\rho^{\otimes n}).$$

$$(5.116)$$

Notice that in the above definitions, we allow for approximate state conversion as long as the approximation error can be made to vanish asymptotically. The reasoning here is that exact conversion might not be possible in the first place. If we do not allow for such an approximation error, we obtain a slightly different quantity, which is arguably less interesting from an operational point of view.

Definition 5.28: The *exact* asymptotic entanglement cost and *exact* asymptotic distillable entanglement of a bipartite state $\rho \in D(AB)$ w.r.t. $S \subset CPTP$ are defined as

$$E_{C,S}^{\text{exact}}(\rho) := \lim_{n \to \infty} \frac{1}{n} E_{C,S}^{(1),0}(\rho^{\otimes n}) \quad \text{and} \quad E_{D,S}^{\text{exact}}(\rho) := \limsup_{n \to \infty} \frac{1}{n} E_{D,S}^{(1),0}(\rho^{\otimes n}) \,. \quad (5.117)$$

By applying the proper regularization on both of the inequalities in Lemma 5.26, we obtain the bounds

$$E_{D,\text{LOCC}}^{\text{exact}}(\rho) \le \log \overline{\gamma}_{\text{SEP}}(\rho) \le E_{C,\text{LOCC}}^{\text{exact}}(\rho)$$
 (5.118)

and

$$\log \gamma_{\text{SEP}}^{\infty}(\rho) \le E_{C,\text{LOCC}}(\rho). \tag{5.119}$$

Note that the Lipschitz constant L grows too quickly in the system size to allow for a lower bound of $\log \gamma_{\rm SEP}^{\infty}(\rho)$ by the asymptotic distillable entanglement. We will see later that this shortcoming can be fixed.

For arbitrary mixed states, the asymptotic entanglement cost can generally be extremely hard to evaluate. It can be written as the regularized *entanglement of* formation $E_{C,LOCC}(\rho) = \lim_{n \to \infty} \frac{1}{n} E_F(\rho^{\otimes n})$ [77]. We omit the exact definition of E_F , but only comment that it can be readily be verified to be sub-additive, hence showing $E_{C,LOCC}(\rho) \leq E_F(\rho)$. If the support of a bipartite state is known to form a so-called *entanglement breaking subspace* [163], [181], then the entanglement of formation is even additive, implying that $E_{C,LOCC}(\rho) = E_F(\rho)$. For two-qubit states σ_{AB} , an explicit expression for E_F is known [177], [178]

$$E_F(\sigma_{AB}) = h\left(\frac{1 + \sqrt{1 - C^2(\sigma_{AB})}}{2}\right) \tag{5.120}$$

where $h(x) := -x \log(x) - (1-x) \log(1-x)$ and

$$C(\sigma_{AB}) := \max\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\} \tag{5.121}$$

where $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are the eigenvalues in decreasing order of the matrix $\left| \sqrt{\sigma} \sqrt{(Y \otimes Y) \bar{\sigma}(Y \otimes Y)} \right|$ and $\bar{\rho}$ is the complex conjugate of ρ .

While the bounds provided by Equations (5.118) and (5.119) did give us some interesting insights, they ultimately fall a bit short, because they are generally not tight and the involved quantities are often quite difficult to compute. In fact, the *extremality theorem* states that the regularization of any "sufficiently nice" entanglement measure lies between the distillable entanglement and the entanglement cost [52], [82]. Remarkably, as we will see below, the approach of relating the regularized quasiprobability extent to an entanglement cost is still extremely fruitful — however we have to shift our attention to the entanglement cost w.r.t. to a different set of channels *S* instead of LOCC.

Definition 5.29: A quantum channel $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ is said to be *separability preserving* if

$$\forall \rho \in \text{SEP}(A; B) : \mathcal{E}(\rho) \in \text{SEP}(A'; B').$$
 (5.122)

We denote by SEPP($A; B \rightarrow A'; B'$) the set of all separability preserving channels from AB to A'B'.

Note that SEPP is a strictly larger set than SEPC, since the latter only contains *completely* separability preserving channels, i.e., channels that cannot generate any entanglement even when applied to a subsystem. The SWAP operation is an example of a channel that lies in SEPP but not SEPC. Thanks to its simpler structure, the finite-shot entanglement cost under SEPP has a much simpler characterization. For this purpose, consider a smoothed version of the robustness of entanglement of a state $\rho \in D(AB)$

$$R_{\text{SEP}}^{\epsilon}(\rho) := \min_{\sigma \in B_{\epsilon}(\rho)} R_{\text{SEP}}(\sigma)$$
 (5.123)

where $\epsilon \geq 0$ and $B_{\epsilon}(\rho) \coloneqq \{\sigma \in \mathrm{D}(AB) | \frac{1}{2} \|\rho - \sigma\|_1 \leq \epsilon \}$. It is clearly equivalent to the smoothed quasiprobability extent $\gamma_{\mathrm{SEP}}^{\epsilon}(\rho_{AB}) = 1 + 2R_{\mathrm{SEP}}^{\epsilon}(\rho_{AB})$ by Lemma 3.17.

Lemma 5.30 ([18, Theorem 2], see also [65, Equation (13.363)]): Let $\rho \in D(AB)$ be a bipartite state and $\epsilon \geq 0$. Then

$$\log(1 + R_{\text{SEP}}^{\epsilon}(\rho)) \le E_{C,\text{SEPP}}^{(1),\epsilon}(\rho) \le \log(1 + R_{\text{SEP}}^{\epsilon}(\rho)) + 1. \tag{5.124}$$

Proof. The following proof works independently of the distance measure that is utilized for the smoothing of the robustness and the error tolerance of the entanglement cost, as long as the same distance is used in both cases. To reflect this, we denote the distance measure by $\delta(\rho_1, \rho_2)$.

We start by proving the lower bound of $E_{C, \text{SEPP}}^{(1), \varepsilon}(\rho)$. Consider a SEPP channel $\mathcal E$ that fulfills $\delta(\mathcal E(|\Psi_M\rangle\langle\Psi_M|), \rho) \leq \varepsilon$ where $|\Psi_M\rangle$ denotes the maximally entangled state with local dimension $M=2^{E_{C, \text{SEPP}}^{(1), \varepsilon}(\rho)}$. Using the monotonicity of the robustness and the formula for the robustness of a maximally entangled state (see Example 5.17), we obtain

$$\log(1 + R_{\text{SEP}}^{\epsilon}(\rho)) \le \log(1 + R_{\text{SEP}}(\mathcal{E}(|\Psi_M\rangle\langle\Psi_M|)))$$
 (5.125)

$$\leq \log(1 + R_{\text{SEP}}(|\Psi_M\rangle\langle\Psi_M|)) \tag{5.126}$$

$$= \log(1 + M - 1) \tag{5.127}$$

$$=E_{\text{C.SEPP}}^{(1),\varepsilon}(\rho) \tag{5.128}$$

We now turn to the upper bound of $E_{C,\text{SEPP}}^{(1),\epsilon}(\rho)$. Let σ be the achieving state in the minimization of $R_{\text{SEP}}^{\epsilon}(\rho)$. Then, let τ be the achieving state in the minimization of $R_{\text{SEP}}(\sigma)$ such that

$$\frac{\sigma + R(\sigma)\tau}{1 + R(\sigma)} \in SEP(A; B). \tag{5.129}$$

We define $M := 1 + \lceil R(\sigma) \rceil$ and

$$\kappa := \frac{\sigma + (M - 1)\tau}{M} = \frac{\sigma + \lceil R(\sigma) \rceil \tau}{1 + \lceil R(\sigma) \rceil}$$
 (5.130)

which is also in SEP(A;B), since $M \ge 1 + R(\sigma)$. Finally, we define the quantum channel

$$\mathcal{E}(X) := \operatorname{tr}[|\Psi_M\rangle\langle\Psi_M|X]\sigma + (1 - \operatorname{tr}[|\Psi_M\rangle\langle\Psi_M|X])\tau \tag{5.131}$$

where again, $|\Psi_M\rangle$ denotes the maximally entangled state with local dimension M. First, notice that clearly $\mathcal{E}(|\Psi_M\rangle\langle\Psi_M|) = \sigma$. Next, we show that $\mathcal{E} \in \text{SEPP}$. For this purpose, let X be a separable state, and define $q := \text{tr}[|\Psi_M\rangle\langle\Psi_M|X] \cdot M$. It holds that $q \in [0,1]$, because $\max_{X \in \text{SEP}} \text{tr}[|\Psi_M\rangle\langle\Psi_M|X]$ is achieved by a pure product state and

$$\operatorname{tr}[|\Psi_{M}\rangle\langle\Psi_{M}|\cdot|f\rangle\langle f|\otimes|g\rangle\langle g|] = \frac{1}{M}\left|\sum_{i}\langle i|f\rangle\langle i|g\rangle\right| \le \frac{1}{M}$$
 (5.132)

where $|i\rangle$ is an orthonormal basis and we used the Cauchy-Schwartz inequality. With this in mind, we now observe that

$$\mathcal{E}(X) = q\frac{1}{M}\sigma + (1 - \frac{q}{M})\tau = q\kappa + (1 - q)\tau \tag{5.133}$$

is a convex mixture of two separable states and thus itself separable. Since ${\mathcal E}$ is in SEPP, we now conclude that

$$E_{C,\text{SEPP}}^{(1),\epsilon}(\rho) \le \log M = \log(1 + \lceil R_{\text{SEP}}(\sigma) \rceil)$$
(5.134)

$$\leq \log(1 + R_{\text{SEP}}(\sigma)) + 1$$
 (5.135)

$$= \log(1 + R_{\text{SFP}}^{\epsilon}(\rho)) + 1. \tag{5.136}$$

This result allows us to directly relate the regularized quasiprobability extent with the asymptotic entanglement cost under SEPP.

Theorem 5.31

For any bipartite state $\rho \in D(AB)$, the regularized and asymptotic quasiprobability extent w.r.t. SEP can be expressed as

$$\log \overline{\gamma}_{\text{SEP}}(\rho) = E_{C,\text{SEPP}}^{\text{exact}}(\rho) \quad \text{and} \quad \log \gamma_{\text{SEP}}^{\infty}(\rho) = E_{C,\text{SEPP}}(\rho).$$
 (5.137)

Proof. We can write the regularized quasiprobability extent as

$$\log \overline{\gamma}_{\text{SEP}}(\rho) = \lim_{n \to \infty} \frac{1}{n} \log \gamma_{\text{SEP}}(\rho^{\otimes n}). \tag{5.138}$$

Note that the sequence $\frac{1}{n}\log\gamma_{\rm SEP}(\rho^{\otimes n})$ converges as $\log\gamma_{\rm SEP}$ is sub-additive. The sequence $\frac{1}{n}\log(1+R_{\rm SEP}(\rho^{\otimes n}))$ must hence also converge to $\log\overline{\gamma}_{\rm SEP}(\rho)$ as

$$\left| \frac{1}{n} \log \left(1 + R_{\text{SEP}}(\rho^{\otimes n}) \right) - \log \overline{\gamma}_{\text{SEP}}(\rho) \right| \leq \frac{\left| \log \left(1 + R_{\text{SEP}}(\rho^{\otimes n}) \right) - \log \left(1 + 2R_{\text{SEP}}(\rho^{\otimes n}) \right) \right|}{n} + \left| \frac{1}{n} \log \gamma_{\text{SEP}}(\rho^{\otimes n}) - \log \overline{\gamma}_{\text{SEP}}(\rho) \right|$$

$$\leq \frac{1}{n} \frac{R_{\text{SEP}}(\rho^{\otimes n})}{\ln(2)(1 + R_{\text{SEP}}(\rho^{\otimes n}))} + \left| \frac{1}{n} \log \gamma_{\text{SEP}}(\rho^{\otimes n}) - \log \overline{\gamma}_{\text{SEP}}(\rho) \right|$$

$$+ \left| \frac{1}{n} \log \gamma_{\text{SEP}}(\rho^{\otimes n}) - \log \overline{\gamma}_{\text{SEP}}(\rho) \right|$$

$$(5.140)$$

$$\xrightarrow{n \to \infty} 0.$$

$$(5.141)$$

The statement $\log \overline{\gamma}_{SEP}(\rho) = E_{C,SEPP}^{exact}(\rho)$ follows by taking the appropriate limit (with $\epsilon = 0$) on both sides of Lemma 5.30.

A similar argument can be used to show

$$\liminf_{n \to \infty} \frac{\log \left(1 + 2R_{\text{SEP}}^{\epsilon}(\rho^{\otimes n})\right)}{n} = \liminf_{n \to \infty} \frac{\log \left(1 + R_{\text{SEP}}^{\epsilon}(\rho^{\otimes n})\right)}{n} \tag{5.142}$$

$$= \liminf_{n \to \infty} \frac{\log \left(1 + R_{\text{SEP}}^{\epsilon}(\rho^{\otimes n})\right) + 1}{n} \tag{5.143}$$

which then implies $\log \gamma_{\rm SEP}^{\infty}(\rho) = E_{C,{\rm SEPP}}(\rho)$.

One direct consequence of this result is that the upper bound $\log \gamma_{\rm SEP}^{\infty} \leq E_{C,LOCC}$ from Equation (5.119) is actually tight for pure states. This is a direct consequence of the chain of inequalities

$$E_{\text{DLOCC}}(\rho) \le E_{\text{DSEPP}}(\rho) \le E_{\text{CSEPP}}(\rho) \le E_{\text{CLOCC}}(\rho)$$
. (5.144)

together with the fact that for pure states $E_{D,LOCC}$ and $E_{C,LOCC}$ coincide and are equal to the entanglement entropy [10].

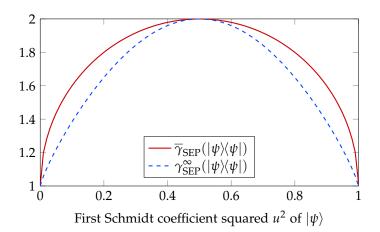


Figure 5.5: Comparison of $\overline{\gamma}_{\text{SEP}}(|\psi\rangle\langle\psi|)$ and $\gamma_{\text{SEP}}^{\infty}(|\psi\rangle\langle\psi|)$ for a pure two-qubit state $|\psi\rangle_{AB}$ with Schmidt coefficients u and $\sqrt{1-u^2}$. The two quasiprobability extents are only identical for separable states ($u^2=0,1$) and the Bell state ($u^2=0.5$).

Corollary 5.32: For any pure state
$$\rho_{AB} \in D(AB)$$
, one has
$$\log \gamma_{\text{SEP}}^{\infty}(\rho_{AB}) = H(\rho_A) \tag{5.145}$$
 where $\rho_A \coloneqq \operatorname{tr}_B[|\psi\rangle\langle\psi|_{AB}]$ and H denotes the von Neumann entropy.

This is especially remarkable, as we previously showed in Corollary 5.16 that the (exact) regularized quasiprobability extent $\overline{\gamma}_{\rm SEP}$ is given in terms of the Rényi-1/2 entanglement entropy instead. We have thus proven that $\overline{\gamma}_{\rm SEP}$ and $\gamma_{\rm SEP}^{\infty}$ must strictly differ for almost all pure states. The difference between the two is depicted in Figure 5.5 for general two-qubit pure states

Theorem 5.31 provides a direct connection between the quasiprobability extent and the SEPP entanglement cost. However, beyond pure states it is unclear whether $E_{C,\text{SEPP}}^{\text{exact}}(\rho)$ or $E_{C,\text{SEPP}}(\rho)$ can be efficiently computed for interesting states ρ . To address this, we will now relax the free channels even further and turn our attention to computing $\overline{\gamma}_{\text{PPT}}$, which is a lower bound for $\overline{\gamma}_{\text{SEP}}$. To achieve this goal, we first need to introduce another entanglement measure called κ -entanglement [171], [172] which turns out to characterize the one-shot entanglement cost under PPTC.

Definition 5.33: The κ-entanglement of a bipartite state
$$\rho \in D(AB)$$
 is defined as
$$E_{\kappa}(\rho) := \log \inf_{S \in P(AB)} \{ tr[S] | - (id_A \otimes T_B)(S) \le (id_A \otimes T_B)(\rho) \le (id_A \otimes T_B)(S) \}.$$
 (5.146)

The κ -entanglement is a special case of the so-called α -logarithmic negativity in the limit of $\alpha \to \infty$ [170]. Note that by definition, the κ entanglement can be expressed as a semidefinite program. Analogous to other quantities, we defined the smoothed

 κ -entanglement as

$$E_{\kappa}^{\epsilon}(\rho) := \min_{\sigma \in B_{\epsilon}(\rho)} E_{\kappa}(\sigma). \tag{5.147}$$

Lemma 5.34: For any bipartite state $\rho \in D(AB)$ one has

$$\gamma_{\text{PPT}}^{\epsilon}(\rho) \ge 2^{E_{\kappa}^{\epsilon}(\rho)} - 1.$$
 (5.148)

Proof. We first prove the non-smoothed case $\epsilon = 0$. Let $\rho = (1+t)\sigma^+ - t\sigma^-$ be the QPD which achieves $\gamma_{\rm PPT}(\rho)$. Due to the positive partial transpose of σ^\pm , we have

$$-(1+t)(\mathrm{id}_A\otimes T_B)(\sigma^++\sigma^-) \preccurlyeq (\mathrm{id}_A\otimes T_B)(\rho) \preccurlyeq (1+t)(\mathrm{id}_A\otimes T_B)(\sigma^++\sigma^-). \tag{5.149}$$

As such, $S := (1+t)(\sigma^+ + \sigma^-)$ is a feasible point for $E_{\kappa}(\rho)$, which implies

$$2^{E_{\kappa}(\rho)} \le (1+t)\operatorname{tr}[\sigma^{+} + \sigma^{-}] = (1+2t) + 1 = \gamma_{PPT}(\rho) + 1. \tag{5.150}$$

The smoothed case follows immediately

$$\gamma_{\text{PPT}}^{\epsilon}(\rho) = \min_{\sigma \in B_{\epsilon}(\rho)} \gamma_{\text{PPT}}(\sigma)$$
 (5.151)

$$\geq \min_{\sigma \in B_{\epsilon}(\rho)} 2^{E_{\kappa}(\sigma)} - 1 \tag{5.152}$$

$$=2^{E_{\kappa}^{\varepsilon}(\rho)}-1. \tag{5.153}$$

The next lemma relates the one-shot entanglement cost under PPTC with the κ -entanglement, analogously to how the log-robustness is related to the entanglement cost under SEPP.

Lemma 5.35 ([65, Theorem 13.25], see also [171, Proposition 2]): For any bipartite state $\rho \in D(AB)$ we have

$$2^{E_{\kappa}^{\epsilon}(\rho)} - 1 \le 2^{E_{\mathsf{C},\mathsf{PPTC}}^{(1),\epsilon}(\rho)} \le 2^{E_{\kappa}^{\epsilon}(\rho)} + 1. \tag{5.154}$$

Theorem 5.36

The regularized and asymptotic quasiprobability extent of a bipartite state $\rho \in D(AB)$ w.r.t. PPT are given by

$$\log \overline{\gamma}_{PPT}(\rho) = \lim_{n \to \infty} \frac{1}{n} E_{\kappa}(\rho^{\otimes n}) = E_{C,PPTC}^{\text{exact}}(\rho)$$
 (5.155)

and

$$\log \gamma_{\rm PPT}^{\infty}(\rho) = \lim_{\epsilon \to 0^+} \lim_{n \to \infty} \frac{1}{n} E_{\kappa}^{\epsilon}(\rho^{\otimes n}) = E_{C,\rm PPTC}(\rho). \tag{5.156}$$

Proof. We first note that the sequence $\frac{1}{n}E_{\kappa}^{\epsilon}(\rho^{\otimes n})$ converges for any $\epsilon \in [0,1)$, see e.g. [65, Theorem 13.26]. As such, Lemma 5.35 directly implies the second equalities in Equations (5.155) and (5.156). The first inequalities follow by taking $\liminf_{n\to\infty}\frac{1}{n}(\cdot)$ on each term of

$$\log(2^{E_{\kappa}^{\varepsilon}(\rho)} - 1) \le \log \gamma_{\text{PPT}}^{\varepsilon}(\rho) \le \log(2^{E_{C,\text{PPTC}}^{(1),\varepsilon}(\rho) + 1} - 1). \tag{5.157}$$

where the first inequality follows from Lemma 5.34 and the second one can be derived analogously to the proof of Lemma 5.26 (recall that the quasiprobability extent of the maximally entangled state is equal under SEP and PPT, see Lemma 5.15). \Box

It was originally stated in Ref. [171] that the κ -entanglement is additive, and thus directly equal to the entanglement cost. However, this was later shown to be wrong [99], [169], as E_{κ} is only sub-additive, and in some cases strictly sub-additive. In our quest to find efficiently computable lower bounds of $\overline{\gamma}_{\text{SEP}}$, it thus remains to show that $E_{\text{C,PPTC}}^{\text{exact}}(\rho)$ is an efficiently computable quantity. For some classes of states, it has a closed-form expression.

Proposition 5.37: [5]

Let $\rho \in D(AB)$ be a bipartite state with zero bi-negativity, i.e.,

$$(\mathrm{id}_A \otimes T_B) \left(|(\mathrm{id}_A \otimes T_B)(\rho)| \right) \geq 0. \tag{5.158}$$

Then $E_{C,PPTC}^{\text{exact}}$ can be written as

$$E_{C,PTC}^{\text{exact}}(\rho) = \log \|(\text{id}_A \otimes T_B)(\rho)\|_1.$$
 (5.159)

Note that the right-hand side of Equation (5.159) is the definition of the entanglement measure called *entanglement negativity* [132], [164].

For general states with non-zero bi-negativity, the situation is more complicated. Here, the entanglement negativity only provides a lower bound to $E_{C,PPTC}^{\rm exact}$ [5]. Recent work has shown that $E_{C,PPTC}^{\rm exact}$ can still be computed through the evaluation of two SDP hierarchies that converge from above and below respectively. We refrain from delving into the details here and simply summarize the main result.

Proposition 5.38: [99, Theorem 3]

Let $\rho \in D(AB)$ be a bipartite state. There exists a classical algorithm that computes $E_{C,PPTC}^{exact}(\rho)$ up to additive error ϵ with runtime

$$\mathcal{O}\left((dD)^{6+o(1)} \cdot \text{polylog}(1/\epsilon)\right)$$
 (5.160)

where $D := \dim(A) \dim(B)$ and $d := \min\{\dim(A), \dim(B)\}$.

Example 5.39. Consider the two noisy 2-qubit states ρ_p and σ_p which are defined to be the depolarized version of the Bell state $|\Psi\rangle$

$$\rho_p := (1 - p) |\Psi\rangle\langle\Psi|_{AB} + p \frac{1}{4} \mathbb{1}_{AB}$$
 (5.161)

and of the pure entangled state $|\phi\rangle := \frac{1}{3} |00\rangle + \frac{2\sqrt{2}}{3} |11\rangle$

$$\sigma_p := (1 - p) |\phi\rangle\langle\phi|_{AB} + p\frac{1}{4}\mathbb{1}_{AB}. \tag{5.162}$$

In Figure 5.6 we depict for ρ_p and σ_p

- the quasiprobability extent γ_{SEP} which is numerically computed through the SDP for γ_{PPT} (recall Corollary 5.18).
- the exact PPTC entanglement cost $E_{C,PPTC}^{\text{exact}}$, which is numerically computed through Proposition 5.37. We note that both ρ_p and σ_p have zero bi-negativity for all values of $p \in [0,1]$.
- the entanglement of formation E_F which is computed through Equation (5.120).
- the regularized and asymptotic quasiprobability extents $\overline{\gamma}_{SEP}$ and γ_{SEP}^{∞} at p=0, which are analytically known through Corollaries 5.16 and 5.32.

The exact PPTC entanglement cost provides a lower bound for $\overline{\gamma}_{SEP}$. This bound coincides with the analytic expression of $\overline{\gamma}_{SEP}$ at p=0 and with γ_{SEP} at $p=p^*$, where p^* is the smallest probability p at which γ_{SEP} reaches the value 1 (p=2/3 for ρ_p and $p\approx 0.56$ for σ_p). Furthermore, we observe (up to numerical precision) that $E_{C,PPTC}^{\text{exact}}$ linearly interpolates between these two points. Remarkably this indicates that $\overline{\gamma}_{SEP}=E_{C,PPTC}^{\text{exact}}$ due to the convexity of the quasiprobability extent.

Finally, we note that the entanglement of formation provides an upper bound to $\gamma_{\rm SEP}^{\infty}$ as

$$\gamma_{\text{SEP}}^{\infty} = E_{C,\text{SEPP}} \le E_{C,\text{LOCC}} = \lim_{n \to \infty} \frac{1}{n} E_F(\cdot^{\otimes n}) \le E_F \tag{5.163}$$

due to Theorem 5.31. Interestingly, for p=0, this upper bound coincides with the analytical expression of $\gamma_{\rm SEP}^{\infty}$ up to numerical precision (note that for the Bell state $|\Psi\rangle$ it also coincides with $\overline{\gamma}_{\rm SEP}$ since the von Neumann and Rényi-1/2 entropies take the same value). For p>0, $\gamma_{\rm SEP}^{\infty}$ must lie somewhere in the shaded region. \triangle

5.1.4 Quasiprobability extent of channels

In this section, we study the quasiprobability extent $\gamma_{LO^*}(\mathcal{E})$ and $\gamma_{LOCC^*}(\mathcal{E})$ of a quantum channel $\mathcal{E} \in \text{CPTP}(AB \to A'B')$. This is a generalization of the quasiprobability extent of states, which can be considered to be the specialization with trivial input space $A = B = \mathbb{C}$. At first, one might hope that the problem could be reduced back to evaluating γ_{SEP} through the state-channel duality of the Choi isomorphism. However, the complicated structure of LOCC* and LO* Choi states as well as the

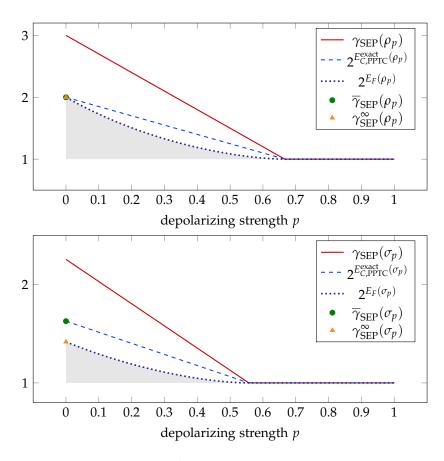


Figure 5.6: Numerical evaluation of various entanglement quantities on ρ_p (top) and σ_p (bottom) which characterize $\gamma_{\rm SEP}$, $\overline{\gamma}_{\rm SEP}$ and $\gamma_{\rm SEP}^{\infty}$. See main text for explanation. We argue that $\overline{\gamma}_{\rm SEP}$ must be equal to $2^{E_{\rm C,PPTC}^{\rm exact}}$ and that $\gamma_{\rm SEP}^{\infty}$ must lie in the shaded region.

partial trace constraint make this endeavor impossible. These difficulties persist even for unitary channels \mathcal{U} : As illustrated by Example 5.50 further below, $\gamma_{\text{SEP}}(J_{\mathcal{U}})$ can generally be strictly smaller than $\gamma_{\text{SEPC}}(\mathcal{U})$, $\gamma_{\text{LOCC}^*}(\mathcal{U})$ and $\gamma_{\text{LO}^*}(\mathcal{U})$.

As a result of these difficulties, we don't know an explicit expression for $\gamma_{\text{LO}^*}(\mathcal{U})$ and $\gamma_{\text{LOCC}^*}(\mathcal{U})$ for all unitary channels \mathcal{U} , which is somewhat disappointing considering we have a complete understanding of γ_{SEP} for pure states. Still, many ideas and techniques that we utilized in the state setting can be (at least partly) generalized to the channel setting. In this section will explore these generalizations. We will start off with two specific sub-classes of unitaries for which we do have a characterization of the quasiprobability extent (Sections 5.1.4.1 and 5.1.4.2). Then, we will turn our attention to more general techniques to find lower and upper bounds for both unitary and non-unitary channels.

5.1.4.1 Clifford gates

In the introduction of this chapter, we have already discussed that a CNOT gate can be realized with an LOCC protocol that consumes a pre-shared Bell pair. Conversely, a Bell pair can clearly be realized from a product state using a CNOT gate. In this sense, a Bell pair and a CNOT gate are equally powerful resources under the QRT of LOCC. This insight can be straightforwardly generalized to arbitrary Clifford gates through the generalized gate teleportation protocol.

Proposition 5.40: Gate teleportation

Let A, A', A'' be n-qubit systems and B, B', B'' m-qubit systems. For any bipartite Clifford unitary $U \in U(AB)$ with induced channel $U \in CPTP(AB)$, there exists an LOCC protocol $\mathcal{E} \in LOCC(AA'A''; BB'B'' \to A; B)$ such that

$$\forall \rho_{AB} \in D(AB) : \mathcal{U}(\rho_{AB}) = \mathcal{E}(\rho_{AB} \otimes (J_{\mathcal{U}})_{A'B'A''B''}). \tag{5.164}$$

Let us denote the *n*-qubit Pauli operators by $\{Q_0^n, Q_1^n, \dots, Q_{4^n-1}^n\}$. The generalized *n*-qubit Bell basis is defined as

$$|\Psi_i^n\rangle := (Q_i^n \otimes \mathbb{1}) \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n - 1} |i\rangle \otimes |i\rangle . \tag{5.165}$$

It is an orthonormal basis because $\langle \Psi_0^n | O \otimes \mathbb{1} | \Psi_0^n \rangle = 2^{-n} \operatorname{tr}[O]$ for any n-qubit operator O. The gate teleportation protocol \mathcal{E} that achieves Proposition 5.40 is depicted in Figure 5.7. Both parties perform a measurement of the generalized Bell basis on the input systems A, B and the corresponding environment systems A'', B'' of the Choi state. They then apply some correction operation that depends on these two measurement outcomes.

This protocol is essentially an extension of the well-known quantum teleportation protocol (which we will also encounter again later, see Figure 5.16) with the correction operation being "pushed through" the unitary U. Whenever U is Clifford, the resulting correction $U(Q_i^n \otimes Q_j^m)U^{\dagger}$ is Pauli and hence locally implementable. Therefore, \mathcal{E} is in LOCC.

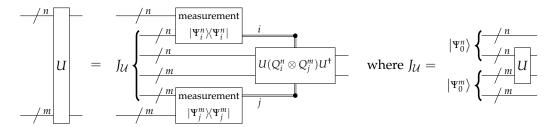


Figure 5.7: Gate teleportation for a general unitary U with associated channel U. If U is a Clifford gate, then the correction operation $U(Q_i^n \otimes Q_j^m)U^{\dagger}$ is a Pauli gate and hence locally implementable.

Corollary 5.41: Let A and B be n-qubit and m-qubit systems. For any bipartite Clifford unitary $U \in U(AB)$ with induced channel $U \in CPTP(AB)$, one has

$$\gamma_{LOCC^*}(\mathcal{U}) = \gamma_{LOCC}(\mathcal{U}) = \gamma_{SEP(AA';BB')}(J_{\mathcal{U}}). \tag{5.166}$$

Proof. A priori, we know by Corollary 3.32 that

$$\gamma_{\text{SEP}}(J_{\mathcal{U}}) \le \gamma_{\text{LOCC}^*}(\mathcal{U}) \le \gamma_{\text{LOCC}}(\mathcal{U})$$
(5.167)

since the Choi state can be prepared from a product state and an invocation of \mathcal{U} . The converse inequality $\gamma_{\text{SEP}}(J_{\mathcal{U}}) \geq \gamma_{\text{LOCC}}(\mathcal{U})$ follows from Proposition 5.40: Any QPD of $J_{\mathcal{U}}$ into separable states automatically induces a QPD of \mathcal{U} into LOCC protocols. \square

Using the closed-form formula for $\gamma_{\text{SEP}}(J_{\mathcal{U}})$ in Theorem 5.13, we can thus efficiently compute the quasiprobability extent for various gates. Similarly, $\overline{\gamma}_{\text{LOCC}^{\star}}(\mathcal{U}) = \overline{\gamma}_{\text{LOCC}}(\mathcal{U}) = \overline{\gamma}_{\text{SEP}}(J_{\mathcal{U}})$ can be evaluated through Corollary 5.16.

Example 5.42. Every two-qubit Clifford gate is equivalent to either the CNOT, SWAP or iSWAP gate up to local Clifford gates. The Schmidt coefficients of J_{CNOT} are $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and those of J_{SWAP} and J_{iSWAP} are $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Hence, we get

$$\gamma_{\text{LOCC}^*}(\text{CNOT}) = 3 \text{ and } \overline{\gamma}_{\text{LOCC}^*}(\text{CNOT}) = 2$$
 (5.168)

and similarly

$$\gamma_{\mathrm{LOCC^{\star}}}(\mathrm{SWAP}) = \gamma_{\mathrm{LOCC^{\star}}}(\mathrm{iSWAP}) = 7 \text{ and } \overline{\gamma}_{\mathrm{LOCC^{\star}}}(\mathrm{SWAP}) = \overline{\gamma}_{\mathrm{LOCC^{\star}}}(\mathrm{iSWAP}) = 4 \,. \tag{5.169}$$

Example 5.43. Consider the general permutation gate $\operatorname{Perm}_{\pi}$ across n+m qubits for some $\pi \in S_{n+m}$ where S_k denotes the permutation group. Define the number of "crossings" $\operatorname{cross}(\pi)$ to be the number of elements $i \in \{1, \ldots, n\}$ such that $\sigma(i) > n$. The Choi state $J_{\operatorname{Perm}_{\pi}}$ is equivalent to $2\operatorname{cross}(\pi)$ Bell pairs, so we get by Example 5.17

$$\gamma_{\text{LOCC}^{\star}}(\text{Perm}_{\pi}) = \gamma_{\text{LOCC}}(\text{Perm}_{\pi}) = 2^{2\text{cross}(\sigma)+1} - 1. \tag{5.170}$$

 \triangle

5.1.4.2 Gates with unitary operator Schmidt decomposition

While the result in Corollary 5.41 is elegant, the approach cannot be applied to channels without an equivalence to a state, and more generally, it also cannot make any statement about γ_{LO^*} . Here, we provide another approach that is applicable to a class of unitaries which exhibit a so-called unitary operator Schmidt decomposition. Any operator $X \in L(AB)$ can be written in terms of its operator Schmidt decomposition [115]

$$X = \sum_{j} u_j A_j \otimes B_j \tag{5.171}$$

where $u_j > 0$ and $A_j \in L(A)$, $B_j \in L(B)$ are orthonormal operators w.r.t. the Hilbert-Schmidt norm, i.e., $\langle A_j, A_k \rangle = \delta_{jk}$ and $\langle B_j, B_k \rangle = \delta_{jk}$.

Theorem 5.44

Let $U \in L(AB)$ be a bipartite unitary with an operator Schmidt decomposition

$$U_{AB} = \sum_{j} u_j(L_j)_A \otimes (R_j)_B \tag{5.172}$$

such that the L_i and R_i are unitary operators on A and B respectively. Then, the quasiprobability extent of the induced channel $\mathcal{U} \in \operatorname{CPTP}(AB)$ is

$$\gamma_{\text{LOCC}^*}(\mathcal{U}) = \gamma_{\text{LO}^*}(\mathcal{U}) = 2\left(\sum_j u_j\right)^2 - 1.$$
(5.173)

This result is applicable to any two-qubit unitary U, as Cartan's KAK decomposition implies that it can always be written in the form

$$U = (V_1 \otimes V_2) \exp(i\theta_X X \otimes X + i\theta_Y Y \otimes Y + i\theta_Z Z \otimes Z) (V_3 \otimes V_4)$$
 (5.174)

for some single-qubit unitaries V_1 , V_2 , V_3 , V_4 and angles θ_X , θ_Y , $\theta_Z \in \mathbb{R}$ fulfilling $|\theta_Z| \le \theta_Y \le \theta_X \le \pi/4$ [92], [157]. Evaluating the exponential function, this directly provides us (up to phases) with a unitary operator Schmidt decomposition

$$U = (V_1 \otimes V_2) (u_0 \mathbb{1} \otimes \mathbb{1} + u_1 X \otimes X + u_2 Y \otimes Y + u_3 Z \otimes Z) (V_3 \otimes V_4)$$
 (5.175)

for some $u_j \in \mathbb{C}$ fulfilling $\sum_{j=0}^3 |u_j|^2 = 1$. By extension, any tensor product of two-qubit unitaries has a unitary operator Schmidt decomposition allowing us to apply Theorem 5.44. This insight enables us to evaluate the regularized quasiprobability extent.

Corollary 5.45: Let $U \in L(AB)$ be a bipartite unitary as in Theorem 5.44. Then,

$$\overline{\gamma}_{LOCC^*}(\mathcal{U}) = \overline{\gamma}_{LO^*}(\mathcal{U}) = (\sum_j u_j)^2.$$
 (5.176)

The proof is completely analogous to the one of Corollary 5.16. Note that the regularization halves the distance to 1 as

$$\gamma_{\text{LOCC}^*}(\mathcal{U}) = \gamma_{\text{LO}^*}(\mathcal{U}) = 1 + 2R_{\text{LOCC}^*}(\mathcal{U}) \tag{5.177}$$

and

$$\overline{\gamma}_{LOCC^{\star}}(\mathcal{U}) = \overline{\gamma}_{LO^{\star}}(\mathcal{U}) = 1 + R_{LOCC^{\star}}(\mathcal{U}),$$
 (5.178)

where $R_{\text{LOCC}^*}(\mathcal{U}) = (\sum_{i} u_i)^2 - 1$ is the robustness measure as in Lemma 3.18.

We now focus our attention on the proof of Theorem 5.44. The general strategy consists of finding matching upper and lower bounds for the quasiprobability extents. The upper bound is provided by explicitly constructing an achieving QPD. This construction is applicable in a slightly more general setting, so we state it as a separate lemma

Lemma 5.46: Let $U \in L(AB)$ be a bipartite unitary of the form

$$U_{AB} = \sum_{j} u_j(L_j)_A \otimes (R_j)_B \tag{5.179}$$

with induced channel \mathcal{U} , where $u_i > 0$, $L_i \in U(A)$ and $R_i \in U(B)$. Then,

$$\gamma_{\text{LO}^*}(\mathcal{U}) \le 2(\sum_{j} u_j)^2 - \sum_{j} u_j^2.$$
(5.180)

We refer to Appendix B for an explicit circuit implementation of the achieving QPD that only requires one ancilla qubit on each bipartition.

Proof. In the following, we construct an explicit LO* QPD of \mathcal{U} that achieves a 1-norm of $2(\sum_i u_i)^2 - \sum_i u_i^2$. The action of the bipartite unitary can be written as

$$\mathcal{U}(\rho) = \sum_{j} u_{j}^{2} (L_{j} \otimes R_{j}) \rho(L_{j}^{\dagger} \otimes R_{j}^{\dagger})$$

$$+ \sum_{j \leq k} u_{j} u_{k} \left[(L_{j} \otimes R_{j}) \rho(L_{k}^{\dagger} \otimes R_{k}^{\dagger}) + (L_{k} \otimes R_{k}) \rho(L_{j}^{\dagger} \otimes R_{j}^{\dagger}) \right].$$
 (5.181)

The first term is already a mixture of local channels, so we only need to concentrate our efforts to decompose the second term. For this purpose, we introduce the two linear superoperators

$$\mathcal{A}_{j,k}^{\phi}(\rho) := \frac{1}{2} \left(e^{i\phi} L_j \rho L_k^{\dagger} + e^{-i\phi} L_k \rho L_j^{\dagger} \right) \tag{5.182}$$

and

$$\mathcal{B}_{j,k}^{\phi}(\rho) := \frac{1}{2} \left(e^{i\phi} R_j \rho R_k^{\dagger} + e^{-i\phi} R_k \rho R_j^{\dagger} \right) \tag{5.183}$$

where $\phi \in [0, 2\pi)$ is some angle. A simple calculation shows that

$$2\left[\mathcal{A}_{j,k}^{0} \otimes \mathcal{B}_{j,k}^{0} - \mathcal{A}_{j,k}^{\pi/2} \otimes \mathcal{B}_{j,k}^{\pi/2}\right] (\rho)$$

$$= \frac{1}{2} \left[(L_{j} \otimes R_{j}) \rho (L_{k}^{\dagger} \otimes R_{k}^{\dagger}) + (L_{j} \otimes R_{k}) \rho (L_{k}^{\dagger} \otimes R_{j}^{\dagger}) + (L_{k} \otimes R_{j}) \rho (L_{j}^{\dagger} \otimes R_{k}^{\dagger}) + (L_{k} \otimes R_{k}) \rho (L_{j}^{\dagger} \otimes R_{j}^{\dagger}) \right]$$

$$- \frac{1}{2} \left[-(L_{j} \otimes R_{j}) \rho (L_{k}^{\dagger} \otimes R_{k}^{\dagger}) + (L_{j} \otimes R_{k}) \rho (L_{k}^{\dagger} \otimes R_{j}^{\dagger}) + (L_{k} \otimes R_{j}) \rho (L_{j}^{\dagger} \otimes R_{k}^{\dagger}) - (L_{k} \otimes R_{k}) \rho (L_{j}^{\dagger} \otimes R_{j}^{\dagger}) \right]$$

$$= (L_{j} \otimes R_{j}) \rho (L_{k}^{\dagger} \otimes R_{k}^{\dagger}) + (L_{k} \otimes R_{k}) \rho (L_{j}^{\dagger} \otimes R_{j}^{\dagger}). \tag{5.184}$$

As such, we have

$$\mathcal{U} = \sum_{j} u_j^2 \mathcal{L}_j \otimes \mathcal{R}_j + \sum_{j < k} 2u_j u_k \left(\mathcal{A}_{j,k}^0 \otimes \mathcal{B}_{j,k}^0 - \mathcal{A}_{j,k}^{\pi/2} \otimes \mathcal{B}_{j,k}^{\pi/2} \right)$$
(5.186)

where $\mathcal{L}_{j}(\rho) = L_{j}\rho L_{j}^{\dagger}$ and $\mathcal{R}_{j}(\rho) = R_{j}\rho R_{j}^{\dagger}$. To show that this is indeed a valid QPD, it remains to show that $\mathcal{A}_{j,k}^{\phi} \otimes \mathcal{B}_{j,k}^{\phi} \in \mathrm{LO}^{\star}(A;B)$. For this, it suffices to show $\mathcal{A}_{j,k}^{\phi} \in \mathrm{CPTP}^{\star}(A)$ and $\mathcal{B}_{j,k}^{\phi} \in \mathrm{CPTP}^{\star}(B)$ as $\mathrm{CPTP}^{\star}(A) \otimes \mathrm{CPTP}^{\star}(B) \subset \mathrm{LO}^{\star}(A;B)$ (recall the discussion of Equation (5.6)). Using the characterization of CPTP^{\star} in Equation (4.1), the desired statement follows from the observation that

$$\mathcal{A}_{i,k}^{\phi}(\rho) = \mathcal{A}_{i,k}^{+,\phi}(\rho) - \mathcal{A}_{i,k}^{-,\phi}(\rho) \tag{5.187}$$

$$\mathcal{B}_{j,k}^{\phi}(\rho) = \mathcal{B}_{j,k}^{+,\phi}(\rho) - \mathcal{B}_{j,k}^{-,\phi}(\rho)$$
 (5.188)

where

$$\mathcal{A}_{j,k}^{\pm,\phi}(\rho) := \frac{1}{4} (L_j \pm e^{-i\phi} L_k) \rho(L_j^{\dagger} \pm e^{i\phi} L_k^{\dagger})$$
 (5.189)

$$\mathcal{B}_{j,k}^{\pm,\phi}(\rho) := \frac{1}{4} (R_j \pm e^{-i\phi} R_k) \rho(R_j^{\dagger} \pm e^{i\phi} R_k^{\dagger}). \tag{5.190}$$

Clearly, $\mathcal{A}_{j,k}^{\pm,\phi}$ and $\mathcal{B}_{j,k}^{\pm,\phi}$ are completely positive maps. To see that $\mathcal{A}_{j,k}^{+,\phi}+\mathcal{A}_{j,k}^{-,\phi}$ is trace-preserving, observe that

$$\frac{1}{4}(L_j^{\dagger} + e^{i\phi}L_k^{\dagger})(L_j + e^{-i\phi}L_k) + \frac{1}{4}(L_j^{\dagger} - e^{i\phi}L_k^{\dagger})(L_j - e^{-i\phi}L_k)
= \frac{1}{4}(2L_j^{\dagger}L_j + 2L_k^{\dagger}L_k) = 1.$$
(5.191)

Analogously, $\mathcal{B}_{j,k}^{+,\phi}+\mathcal{B}_{j,k}^{-,\phi}$ is trace-preserving. Since we have a valid QPD, we conclude that

$$\gamma_{\text{LO}^*}(\mathcal{U}) \le \sum_{j} u_j^2 + 4 \sum_{i \le k} u_i u_k = 2(\sum_{j} u_j)^2 - \sum_{i} u_j^2.$$
(5.192)

Proof of Theorem 5.44. Notice that

$$\dim(AB) = \operatorname{tr}[UU^{\dagger}] \tag{5.193}$$

$$= \sum_{j,k} u_j u_k \operatorname{tr}[(L_j L_k^{\dagger}) \otimes (R_j R_k^{\dagger})]$$
 (5.194)

$$= \sum_{j,k} u_j u_k \operatorname{tr}[L_j L_k^{\dagger}] \operatorname{tr}[R_j R_k^{\dagger}]$$
 (5.195)

$$= \dim(AB) \sum_{j,k} u_j u_k \delta_{jk} \delta_{jk}$$
 (5.196)

$$= \dim(AB) \sum_{j,k} u_j u_k \delta_{jk} \delta_{jk}$$

$$= \dim(AB) \sum_j u_j^2$$
(5.196)
$$= (5.197)$$

and hence $\sum_i u_i^2 = 1$. Thanks to Lemma 5.46 and $\gamma_{LOCC^*} \leq \gamma_{LO^*}$, it remains to show that $\gamma_{\text{LOCC}^*} \geq 2(\sum_i u_i)^2 - 1$.

Recall that the Choi state of the unitary channel $\mathcal U$ is defined as

$$J_{\mathcal{U}} = (\mathcal{U} \otimes \mathrm{id}_{A'B'})(|\Psi\rangle\langle\Psi|_{AA'} \otimes |\Phi\rangle\langle\Phi|_{BB'}) \tag{5.198}$$

where A', B' are copies of A, B and $|\Psi\rangle_{AA'}, |\Phi\rangle_{BB'}$ are maximally entangled states. By Corollary 3.32, this implies that

$$\gamma_{\text{LOCC}^{\star}(A;B)}(\mathcal{U}) \ge \gamma_{\text{SEP}(AA';BB')}(J_{\mathcal{U}}).$$
 (5.199)

Since the Choi state is pure $J_{\mathcal{U}} = |J_{\mathcal{U}}\rangle\langle J_{\mathcal{U}}|$, we can express its quasiprobability extent in terms of its Schmidt coefficients v_j (see Theorem 5.13)

$$\gamma_{\text{SEP}(AA';BB')}(J_{\mathcal{U}}) = 2(\sum_{j} v_{j})^{2} - 1.$$
 (5.200)

It remains to explicitly evaluate the Schmidt coefficients for the Choi state. For some fixed orthonormal bases $\{|i\rangle\}_i$ and $\{|j\rangle\}_i$ of A and B we get

$$|J_{U}\rangle_{A'B'AB} = (U \otimes \mathrm{id}_{A'B'}) \frac{1}{\sqrt{\dim(A)\dim(B)}} \sum_{i=1}^{\dim(A)} \sum_{j=1}^{\dim(B)} |i\rangle_{A} |i\rangle_{A'} |j\rangle_{B} |j\rangle_{B'}$$
 (5.201)

$$= \sum_{k} u_{k} \frac{1}{\sqrt{\dim(A)\dim(B)}} \sum_{i=1}^{\dim(A)} \sum_{j=1}^{\dim(B)} (L_{k} |i\rangle_{A}) |i\rangle_{A'} (R_{k} |j\rangle_{B}) |j\rangle_{B'} \quad (5.202)$$

$$=\sum_{k}u_{k}\left|f_{k}\right\rangle _{AA^{\prime}}\left|g_{k}\right\rangle _{BB^{\prime}}\tag{5.203}$$

where

$$|f_k\rangle_{A'A} := \frac{1}{\sqrt{\dim(A)}} \sum_{i=1}^{\dim(A)} (L_k |i\rangle_A) |i\rangle_{A'}$$
 (5.204)

$$|g_k\rangle_{B'B} := \frac{1}{\sqrt{\dim(B)}} \sum_{j=1}^{\dim(B)} (R_k |j\rangle_B) |j\rangle_{B'}.$$
 (5.205)

It remains to verify that the set of states $\{|f_k\rangle\}_k$ and $\{|g_k\rangle\}_k$ are orthogonal, which then implies that we have found a valid Schmidt decomposition of $|J_U\rangle$ with Schmidt coefficients u_i . This follows directly from the orthogonality of the unitaries

$$\langle f_j | f_k \rangle = \frac{1}{\dim(A)} \sum_{i,i'} \langle i | i' \rangle \cdot \langle i | L_j^{\dagger} L_k | i' \rangle = \frac{1}{\dim(A)} \operatorname{tr}[L_j^{\dagger} L_k] = \delta_{jk}$$
 (5.206)

and analogously $\langle g_i | g_k \rangle = \delta_{ik}$.

Example 5.47. As an application of Theorem 5.44, we list a few important two-qubit gates and their associated quasiprobability extent.

Controlled rotation gate The controlled rotation gate w.r.t. the Bloch sphere vector $\vec{n} \in \mathbb{R}^3$, $||\vec{n}||_2 = 1$ and angle $\theta \in [0, 2\pi)$ is defined as

$$\operatorname{CR}_{\vec{n}}(\theta) := e^{-i\frac{\theta}{2}|1\rangle\langle 1|\otimes(n_1X + n_2Y + n_3Z)}. \tag{5.207}$$

Its Schmidt coefficients are given by $u = (\cos \frac{\theta}{4}, \sin \frac{\theta}{4}, 0, 0)$ and as such one has

$$\gamma_{\mathrm{LO}^{\star}}(\mathrm{CR}_{\vec{n}}(\theta)) = \gamma_{\mathrm{LOCC}^{\star}}(\mathrm{CR}_{\vec{n}}(\theta)) = 1 + 2\sin(\theta/2) \tag{5.208}$$

$$\overline{\gamma}_{\mathrm{LO}^*}(\mathrm{CR}_{\vec{n}}(\theta)) = \overline{\gamma}_{\mathrm{LOCC}^*}(\mathrm{CR}_{\vec{n}}(\theta)) = 1 + \sin(\theta/2). \tag{5.209}$$

CNOT gate We recover the result from Example 5.42 as a special case of the rotation gate with $\theta = \pi$.

Two-qubit Pauli rotation gate For $P \in \{X, Y, Z\}$ and an angle $\theta \in [0, 2\pi)$ we define

$$R_{PP}(\theta) := e^{-i\frac{\theta}{2}P \otimes P}. \tag{5.210}$$

The associated Schmidt coefficients are $u = (\left|\cos\frac{\theta}{2}\right|, \sin\frac{\theta}{2}, 0, 0)$ and as such one has

$$\gamma_{\text{LO}^*}(R_{PP}(\theta)) = \gamma_{\text{LOCC}^*}(R_{PP}(\theta)) = 1 + 2|\sin\theta|$$
 (5.211)

$$\overline{\gamma}_{\mathrm{LO}^{\star}}(\mathrm{R}_{PP}(\theta)) = \overline{\gamma}_{\mathrm{LOCC}^{\star}}(\mathrm{R}_{PP}(\theta)) = 1 + |\sin \theta| . \tag{5.212}$$

SWAP/iSWAP gate We recover the result from Example 5.42 by using that both the SWAP and iSWAP gates have the Schmidt coefficients $u = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Note that they have the maximal quasiprobability extent of any two-qubit unitary as the associated Schmidt vector $u \in \mathbb{R}^4$ must always fulfill

$$2(\sum_{j} u_{j})^{2} - 1 = 2 \|u\|_{1}^{2} - 1 \le 2(\sqrt{4} \|u\|_{2})^{2} - 1 = 7$$
 (5.213)

where we used the Cauchy-Schwarz inequality.

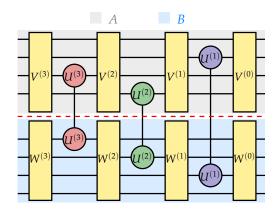


Figure 5.8: Cut of a large unitary that can be written as the composition of multiple two-qubit unitaries (here denoted by $U^{(1)}, U^{(2)}, U^{(3)}$) interleaved with local unitaries (denoted by $V^{(i)}, W^{(i)}$).

Example 5.48. We briefly illustrate the LO* QPD of the CNOT gate that achieves $\gamma_{\text{LO}^*}(\text{CNOT}) = 3$ throught the construction of Theorem 5.44. Up to local unitaries, we can equate the CNOT with the unitary

$$V := \frac{1}{\sqrt{2}} \mathbb{1} \otimes \mathbb{1} + \frac{1}{\sqrt{2}} Z \otimes (-iZ)$$
 (5.214)

as

$$CNOT = \left(S^{\dagger} \otimes (HS^{\dagger})\right) V \left(\mathbb{1} \otimes H\right) . \tag{5.215}$$

We can thus focus our considerations on finding the optimal LO * QPD of V. Following the construction in Lemma 5.46, the optimal QPD reads

$$Ad[V] = \frac{1}{2}id \otimes id + \frac{1}{2}Ad[Z] \otimes Ad[Z] + A_{1,2}^{0} \otimes B_{1,2}^{0} - A_{1,2}^{\pi/2} \otimes B_{1,2}^{\pi/2}$$
 (5.216)

where

$$A_{1,2}^{0} = \operatorname{Ad}[|0\rangle\langle 0|] - \operatorname{Ad}[|1\rangle\langle 1|], \qquad A_{1,2}^{\pi/2} = \frac{1}{2}\operatorname{Ad}[S^{\dagger}] - \frac{1}{2}\operatorname{Ad}[S], \qquad (5.217)$$

$$\mathcal{B}_{1,2}^{0} = \frac{1}{2} \operatorname{Ad}[S] - \frac{1}{2} \operatorname{Ad}[S^{\dagger}], \qquad \mathcal{B}_{1,2}^{\pi/2} = \operatorname{Ad}[|1\rangle\langle 1|] - \operatorname{Ad}[|0\rangle\langle 0|].$$
 (5.218)

Δ

We note that Lemma 5.46 provides us a technique to find upper bounds for the quasiprobability extent of large unitaries that are composed of two-qubit nonlocal gates interleaved with local unitaries. Consider for example the unitary U_{tot} depicted in Figure 5.8. From the chaining property (see Lemmas 3.31 and 3.35), one can obtain the trivial upper bound

$$\gamma_{\mathrm{LO}^{\star}}(\mathcal{U}_{\mathrm{tot}}) \le \gamma_{\mathrm{LO}^{\star}}(\mathcal{U}^{(1)})\gamma_{\mathrm{LO}^{\star}}(\mathcal{U}^{(2)})\gamma_{\mathrm{LO}^{\star}}(\mathcal{U}^{(3)})$$
(5.219)

However, following corollary shows that the cost of cutting U_{tot} is at most the cost of jointly cutting the $U^{(i)}$, instead of cutting them separately

$$\gamma_{\text{LO}^{\star}}(\mathcal{U}_{\text{tot}}) \le \gamma_{\text{LO}^{\star}}(\mathcal{U}^{(1)} \otimes \mathcal{U}^{(2)} \otimes \mathcal{U}^{(3)}).$$
 (5.220)

which is a better bound due to the strict sub-multiplicativity of γ_{LO^*} .

Corollary 5.49: Let $U_{\text{tot}} \in U(AB)$ be a unitary of the form

$$U_{\text{tot}} = (V_A^{(0)} \otimes W_B^{(0)}) \prod_{i=1}^n \left(U_{AB}^{(i)} (V_A^{(i)} \otimes W_B^{(i)}) \right)$$
 (5.221)

where $V_A^{(i)} \in \mathrm{U}(A)$, $W_B^{(i)} \in \mathrm{U}(B)$ and the $U_{AB}^{(i)} \in \mathrm{U}(AB)$ act as a two-qubit unitaries across A and B. We then have

$$\gamma_{\text{LO}^*}(\mathcal{U}_{\text{tot}}) \le \gamma_{\text{LO}^*} \left(\bigotimes_{i=1}^n \mathcal{U}^{(i)}\right)$$
(5.222)

where \mathcal{U}_{tot} and $\mathcal{U}^{(i)}$ denote the induced channels of U_{tot} and $U^{(i)}$ respectively.

We note that a better upper bound is not possible without more knowledge about the specific structure of the unitary, as the bound is clearly achieved when the $V^{(i)}$, $W^{(i)}$ are the identity and the $U^{(i)}$ act on different qubits. Conversely, a meaningful lower bound for $\gamma_{\mathrm{LO}^{\star}}(\mathcal{U}_{\mathrm{tot}})$ in terms of only the $\gamma_{\mathrm{LO}^{\star}}(\mathcal{U}^{(i)})$ is also impossible, as generally the entangling gates $U^{(i)}$ could undo the effects of each other leading to U being the identity.

Proof. Combining the KAK decompositions of the individual two-qubit unitaries $U^{(i)}$, we immediately obtain a decomposition of U_{tot} that is of the form as in Lemma 5.46.

5.1.4.3 State conversion bounds

Consider a quantum channel $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$. As a consequence of the chaining property (recall Corollary 3.32 and Lemma 3.35), we have

$$\gamma_{\text{SEPC}^*}(\mathcal{E}) \ge \gamma_{\text{SEP}(A'\bar{A};B'\bar{B})} \left((\text{id}_{\bar{A}\bar{B}} \otimes \mathcal{E})(\rho) \right)$$
(5.223)

for any systems \bar{A} , \bar{B} and for any separable state $\rho \in SEP(A\bar{A}; B\bar{B})$. This in turn provides us with a lower bound for $\gamma_{LOCC^*}(\mathcal{E})$ and $\gamma_{LO^*}(\mathcal{E})$. The power of this approach lies in the reduction to the state setting, as estimating γ_{SEP} is often much easier. The choice of ρ is critical: Ideally, $(id \otimes \mathcal{E})(\rho)$ should be as strongly entangled as possible.

We already used this lower bound in the proof of Corollary 5.41 and Theorem 5.44, where we chose ρ to be the product of two maximally entangled states on $A\bar{A}$ and $B\bar{B}$. In that case, $(\mathrm{id} \otimes \mathcal{E})(\rho)$ is precisely the Choi operator $J_{\mathcal{E}}$. It is natural to ask whether this choice of ρ is always the best. The example below shows that this is not the case, even for unitary channels \mathcal{E} .

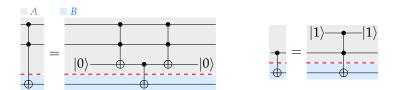


Figure 5.9: A Toffoli gate and a CNOT gate are equally powerful in the sense that each of the two can be realized through the other using only local operations. The red dotted line denotes the partition into two subsystems. The ancilla qubits are uncomputed and end up in a well-defined single-qubit state.

Example 5.50. Consider the three-qubit Toffoli gate acting on the bipartite system *AB* where *A* consists of the two control qubits and *B* of the target qubit. The Toffoli gate is equivalent to a CNOT gate across *A* and *B* up to some local operations, i.e., a Toffoli gate can be realized with one invocation of a CNOT gate and vice versa. The corresponding circuit identities are depicted in Figure 5.9. As such, it follows from the chaining property (recall Lemma 3.31) that

$$\gamma_{\text{LOCC}^*}(\text{Toffoli}) = \gamma_{\text{LO}^*}(\text{Toffoli}) = \gamma_{\text{LOCC}^*}(\text{CNOT}) = 3.$$
 (5.224)

However, the Choi-based lower bound does not achieve this value. A simple calculation reveals that the Schmidt coefficients of J_{Toffoli} are given by $\frac{1}{2}$ and $\frac{\sqrt{3}}{2}$, hence showing that

$$\gamma_{\rm SEP}(J_{\rm Toffoli}) = 1 + \sqrt{3} \approx 2.732 \tag{5.225}$$

using Theorem 5.13. Yet, clearly we can pick a simple state like $|1\rangle \otimes |+\rangle \otimes |0\rangle$ to achieve the quasiprobability extent using the state conversion bound. Since the state

Toffoli
$$|1\rangle |+\rangle |0\rangle = \frac{1}{\sqrt{2}} (|10\rangle \otimes |0\rangle + |11\rangle \otimes |1\rangle)$$
 (5.226)

has Schmidt coefficients $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, it achieves a quasiprobability extent of 3. \triangle

Coming back to the general case, the strongest version of the state-based lower bound takes the form

$$\gamma_{\text{SEPC}^*}(\mathcal{E}) \ge \sup_{\bar{A}, \bar{B}, \rho \in \text{SEP}(A\bar{A}; B\bar{B})} \gamma_{\text{SEP}}((\text{id} \otimes \mathcal{E})(\rho)). \tag{5.227}$$

We can restrict the optimization to pure states and systems with equal dimension $\dim(\bar{A}) = \dim(A)$, $\dim(B) = \dim(\bar{B})$, since

$$\sup_{\bar{A},\bar{B},\rho\in SEP(A\bar{A};B\bar{B})} \gamma_{SEP}((id\otimes\mathcal{E})(\rho)) = \sup_{\bar{A},\bar{B}} \max_{\rho\in SEP(A\bar{A};B\bar{B})} \gamma_{SEP}((id\otimes\mathcal{E})(\rho))$$

$$= \sup_{\bar{A},\bar{B}} \max_{\substack{|\psi\rangle\in\mathcal{H}_{A\bar{A}}\\|\phi\rangle\in\mathcal{H}_{B\bar{B}}}} \gamma_{SEP}((id\otimes\mathcal{E})(|\psi\rangle\langle\psi|\otimes|\phi\rangle\langle\phi|))$$
(5.228)

where we used the convexity of γ_{SEP} . For any given $|\psi\rangle$, $|\phi\rangle$, we can restrict \bar{A} , \bar{B} to be the spaces that are spanned by the corresponding Schmidt basis.

The quantity on the right-hand side of Equation (5.227) is essentially the *unas-sisted entangling capacity* introduced in Ref. [33]. The authors also showed that the lower bound cannot be improved by picking an already entangled state ρ and using $\gamma_{\text{SEPC}^*}(\mathcal{E}) \geq \frac{\gamma_{\text{SEP}}(\mathcal{E}(\rho))}{\gamma_{\text{SEP}}(\rho)}$, see [33, Corollary 1].

5.1.4.4 PPTC lower bound

One useful strategy to find lower bounds of γ_{LO^*} and γ_{LOCC^*} is to relax the decomposition set to PPTC*. In full analogy to the SDP of γ_{PPT} in Equation (5.40), we can express $\gamma_{PPTC^*}(\mathcal{E})$ for a channel $\mathcal{E} \in CPTP(AB \to A'B')$ as an SDP

$$\gamma_{\mathrm{PPTC}^{\star}}(\mathcal{E}) = \gamma_{\mathrm{PPTC}}(\mathcal{E}) = \begin{cases} \min_{\Lambda^{\pm} \in \mathrm{P}(ABA'B')} & \mathrm{tr}[\Lambda^{+}] + \mathrm{tr}[\Lambda^{-}] \\ \mathrm{s.t.} & J_{\mathcal{E}} = \Lambda^{+} - \Lambda^{-}, \\ & \mathrm{tr}_{A'B'}[\Lambda^{\pm}] = \mathrm{tr}[\Lambda^{\pm}] \frac{1}{\dim(AB)} \mathbb{1}_{AB}, \\ & (\mathrm{id}_{AA'} \otimes T_{BB'})(\Lambda^{\pm}) \succcurlyeq 0 \end{cases} . (5.230)$$

where we used Proposition 5.10 and Lemma 3.11.

Lemma 5.51: The following is a valid dual SDP of Equation (5.230)
$$\max_{\substack{X \in H(ABA'B') \\ Y^{\pm} \in H(AB) \\ Z^{\pm} \in P(ABA'B')}} \operatorname{tr}[XJ_{\mathcal{E}}]$$
 s.t.
$$\mathbb{1}_{A'B'} \otimes Y^{-} + (\operatorname{id}_{AA'} \otimes T_{BB'})(Z^{-}) - \mathbb{1}_{AA'BB'}(1 + \frac{\operatorname{tr}[Y^{-}]}{\dim(AB)}) \leq X,$$

$$X \leq -\mathbb{1}_{A'B'} \otimes Y^{+} - (\operatorname{id}_{AA'} \otimes T_{BB'})(Z^{+}) + \mathbb{1}_{AA'BB'}(1 + \frac{\operatorname{tr}[Y^{+}]}{\dim(AB)})$$
 (5.231)

The proof is very similar to the one of Lemma 5.12.

Proof. In the following, we will denote the unnormalized Hilbert-Schmidt inner prod-

uct by $\langle A, B \rangle := \operatorname{tr}[A^{\dagger}B]$. The Lagrangian is given by

$$\begin{split} L(\Lambda^{\pm}; X, Y^{\pm}, Z^{\pm}) &= \operatorname{tr}[\Lambda^{+}] + \operatorname{tr}[\Lambda^{-}] - \langle X, J_{\mathcal{E}} - \Lambda^{+} + \Lambda^{-} \rangle \\ &- \langle Y^{+}, \operatorname{tr}_{A'B'}[\Lambda^{+}] - \frac{\operatorname{tr}[\Lambda^{+}]}{\dim(AB)} \mathbb{1}_{AB} \rangle \\ &- \langle Y^{-}, \operatorname{tr}_{A'B'}[\Lambda^{-}] - \frac{\operatorname{tr}[\Lambda^{-}]}{\dim(AB)} \mathbb{1}_{AB} \rangle \\ &- \langle Z^{+}, (\operatorname{id}_{AA'} \otimes T_{BB'})(\Lambda^{+}) \rangle - \langle Z^{-}, (\operatorname{id}_{AA'} \otimes T_{BB'})(\Lambda^{-}) \rangle \\ &= - \langle X, J_{\mathcal{E}} \rangle \\ &+ \langle \Lambda^{+}, \mathbb{1}_{ABA'B'}(1 + \frac{\operatorname{tr}[Y^{+}]}{\dim(AB)}) + X - \mathbb{1}_{A'B'} \otimes Y^{+} - (\operatorname{id}_{AA'} \otimes T_{BB'})(Z^{+}) \rangle \\ &+ \langle \Lambda^{-}, \mathbb{1}_{ABA'B'}(1 + \frac{\operatorname{tr}[Y^{-}]}{\dim(AB)}) - X - \mathbb{1}_{A'B'} \otimes Y^{-} - (\operatorname{id}_{AA'} \otimes T_{BB'})(Z^{-}) \rangle \,. \end{split}$$
 (5.233)

where we used the relation $\langle R_A, \operatorname{tr}_B[S_{AB}] \rangle = \langle R_A \otimes \mathbb{1}_B, S_{AB} \rangle$. The constraints of the dual SDP thus read

$$\mathbb{1}_{ABA'B'}(1 + \frac{\text{tr}[Y^{\pm}]}{\dim(AB)}) \pm X - \mathbb{1}_{A'B'} \otimes Y^{\pm} - (\text{id}_{AA'} \otimes T_{BB'})(Z^{\pm}) \geq 0$$
 (5.234)

and the desired form of the dual SDP is obtained with the substitution $X \rightarrow -X$.

For a unitary channel $\mathcal{U} \in \operatorname{CPTP}(AB)$, the PPTC lower bound is provably at least as good as the state conversion bound discussed in Section 5.1.4.3

$$\gamma_{\text{PPTC}}(\mathcal{U}) \ge \max_{|\psi\rangle \text{ product state}} \gamma_{\text{PPT}}((\mathcal{U} \otimes \text{id}_{\bar{A}\bar{B}})(|\psi\rangle\langle\psi|))
= \max_{|\psi\rangle \text{ product state}} \gamma_{\text{SEP}}((\mathcal{U} \otimes \text{id}_{\bar{A}\bar{B}})(|\psi\rangle\langle\psi|))$$
(5.235)

$$= \max_{|\psi\rangle \text{ product state}} \gamma_{\text{SEP}}((\mathcal{U} \otimes \text{id}_{\bar{A}\bar{B}})(|\psi\rangle\langle\psi|)) \tag{5.236}$$

where we used that γ_{SEP} and γ_{PPT} coincide on pure states (recall Lemma 5.15). In the context of Example 5.50, this implies that $\gamma_{PPTC}(Toffoli) = 3$. This can readily be verified using a numerical solver.

Example 5.52. We consider the noisy CNOT channel

$$\mathcal{E}_p(\rho) := (1 - p) \operatorname{CNOT}(\rho) + p \mathcal{D}(\rho)$$
 (5.237)

where \mathcal{D} is the fully depolarizing two-qubit channel. In Figure 5.10, we depict the numerical evaluation of $\gamma_{PPTC^*}(\mathcal{E}_p)$ as well as $\gamma_S(\mathcal{E}_p)$ where S is a heuristically chosen finite decomposition set containing

- the elements $\mathcal{F} \in S_{\text{CNOT}}$ where S_{CNOT} is the decomposition set achieving the optimal LO* QPD of the CNOT gate discussed in Example 5.48,
- the two-qubit Pauli channels,
- the superoperators of the form $\mathcal{G} \circ \mathcal{F}$ where $\mathcal{F} \in S_{\text{CNOT}}$ and \mathcal{G} is a two-qubit Pauli channel.

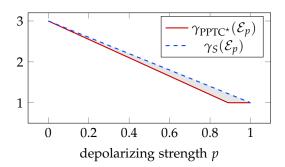


Figure 5.10: Numerical evaluation of $\gamma_{PPTC^*}(\mathcal{E}_p)$ and $\gamma_S(\mathcal{E}_p)$ for the depolarized CNOT gate \mathcal{E}_p . Here, S is a heuristically chosen decomposition set that is contained in LO*.

Since $S \subset LO^*$, we know that $\gamma_{LO^*}(\mathcal{E}_p)$ and $\gamma_{LOCC^*}(\mathcal{E}_p)$ must lie in the shaded region. We observe that γ_S does not perform better than simply taking the trivial bounds implied by the convexity of the quasiprobability extent. Interestingly, $\gamma_{LO^*}(\mathcal{E}_p)$ and $\gamma_{LOCC^*}(\mathcal{E}_p)$ are strictly larger than the corresponding extent of the noisy Bell state $\gamma_{SEP}(\rho_p)$ that we considered in Example 5.39.

5.1.4.5 SDP hierarchy for γ_{SEPC}

In the setting of states, we have previously seen in Section 5.1.3.2 that $\gamma_{\rm SEP}$ can be expressed in terms of an efficiently converging SDP hierarchy. Unfortunately, this approach cannot be directly applied to $\gamma_{\rm LOCC^*}$. Instead, we have to content ourselves with using the DPS hierarchy to evaluate the quasiprobability extent w.r.t. to the set of channels with separable Choi operators SEPC. This in turn at least provides a lower bound $\gamma_{\rm LO^*} \geq \gamma_{\rm LOCC^*} \geq \gamma_{\rm SEPC^*} = \gamma_{\rm SEPC}$ thanks to Proposition 5.10. The techniques will mirror closely the approach we previously introduced in Section 5.1.3.2. As such, we will not delve into as many details and rather focus on the differences and difficulties that specifically occur in this setting.

We define an outer approximation of the set SEPC($A; B \rightarrow A'; B'$) as

$$SEPC_k(A; B \to A'; B') := \{ \mathcal{E} \in CPTP(AB) | J_{\mathcal{E}} \in SEP_k(AA'; BB') \}. \tag{5.238}$$

As for SEP_k , we can optimize over elements of $SEPC_k$ by considering symmetric extensions in the larger space $AB_1 \dots B_k A'B'_1 \dots B'_k$ where the B_i and B'_i are copies of

B or B' respectively. Following the same steps in the derivation of γ_{SEP_k} , we obtain

or
$$B'$$
 respectively. Following the same steps in the derivation of γ_{SEP_k} , we obtain
$$\gamma_{\text{SEPC}_k}(\mathcal{E}) = \begin{cases} \min_{\Lambda^{\pm} \in P(AB_1...B_kA'B_1'...B_k')} & \text{tr}[\Lambda^{+}] + \text{tr}[\Lambda^{-}] \\ \text{s.t.} & J_{\mathcal{E}} = \text{tr}_{B_2B_2'...B_kB_k'}[\Lambda^{+}] - \text{tr}_{B_2B_2'...B_kB_k'}[\Lambda^{-}], \\ & \text{tr}_{A'B_2...B_kB_1'..B_k'}[\Lambda^{\pm}] = \frac{1}{\dim(AB)} \mathbb{1}_{AB_1}, \\ \text{SWAP}_{B_iB_i',B_jB_j'} & \Lambda^{\pm} \text{SWAP}_{B_iB_i',B_jB_j'} = \Lambda^{\pm}, \\ & \left(\text{id}_{AA'} \otimes T_{B_1B_1'} \otimes \cdots \otimes \text{id}_{B_kB_k'} \right) (\Lambda^{\pm}) \succcurlyeq 0, \\ & \vdots \\ & \left(\text{id}_{AA'} \otimes T_{B_1B_1'} \otimes \cdots \otimes T_{B_kB_k'} \right) (\Lambda^{\pm}) \succcurlyeq 0 \end{cases}$$

$$(5.239)$$

The first level of the hierarchy precisely corresponds to $SEPC_1(A; B \rightarrow A'; B') =$ $PPTC(A; B \to A'; B')$. The completeness of the DPS hierarchy (see Proposition 5.21) implies that this outer approximation will also converge, i.e.,

$$SEPC(A; B \to A'; B') = \bigcap_{k=1}^{\infty} SEPC_k(A; B \to A'; B').$$
 (5.240)

Unfortunately, the statement about convergence speed in Proposition 5.22 cannot be directly translated to this setting. Given the optimal QPD $\mathcal{E} = a^+ \mathcal{E}^+ - a^- \mathcal{E}^-$, the de Finetti theorem only guarantees that $J_{\mathcal{E}^+}$ and $J_{\mathcal{E}^-}$ are close to separable operators, but not necessarily to separable operators that fulfill the partial trace constraint of a tracepreserving channel. There exist modifications of the de Finetti theorem that allow for imposing linear constraints on the approximating state, such as the result by Berta et al. in Ref. [16, Theorem 2.3]. However, this modified de Finetti theorem can only incorporate linear constraints that separately act on AA' and BB', and is therefore not applicable to our case. A slightly different modification would be required, which to our knowledge has not been established in previous literature.

5.1.4.6 Overview of one-shot bounds

Here, we come back to the chain of inequalities we previously introduced in Equation (5.14) and try to complete that picture with a revised understanding of which of the inequalities are strict and which ones are not. We summarize our results in Figure 5.11, which also includes the Choi lower bounds. The relations \geq and \leq indicate that one quantity is generally at least/at most as large as the other and that there exists a specific counter example of a CPTP map $\mathcal E$ for which the two quantities do not coincide. The following relations are directly implied by previously shown results.

- $\gamma_{LO} \ge \gamma_{LO^*}$ and $\gamma_{LO} \ge \gamma_{LOCC}$ due to Lemmas 5.7 and 5.8.
- $\gamma_{\text{SEPC}^*} = \gamma_{\text{SEPC}}$ and $\gamma_{\text{PPTC}^*} = \gamma_{\text{PPTC}}$ due to Proposition 5.10.
- $\gamma_{\text{SEPC}}(\mathcal{E}) \ngeq \gamma_{\text{SEP}}(J_{\mathcal{E}})$ and $\gamma_{\text{PPTC}}(\mathcal{E}) \trianglerighteq \gamma_{\text{PPT}}(J_{\mathcal{E}})$ by the example of the Toffoli gate. Recall from Example 5.50 that $\gamma_{\rm SEP}(J_{\rm Toffoli}) = \gamma_{\rm PPT}(J_{\rm Toffoli}) \approx 2.732$. At the same time, we also saw that $\gamma_{PPTC}(Toffoli) = \gamma_{SEPC}(Toffoli) = 3$.

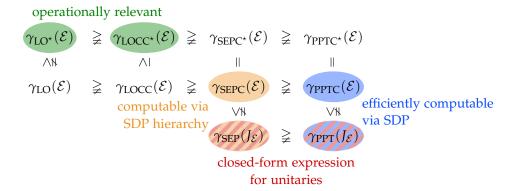


Figure 5.11: Relations between various quasiprobability extents for some CPTP map \mathcal{E} . The relations \supseteq and \supseteq indicate that one quantity is generally at least/at most as large as the other, and there exists a specific counter example of a quantum channel \mathcal{E} for which the two quantities do not coincide.

For two decomposition sets $S \subsetneq S' \subset \text{CPTP}$ such that S is convex and compact, we have $1 = \gamma_{S'}(\mathcal{F}) < \gamma_S(\mathcal{F})$ for any $\mathcal{F} \in S' \setminus S$ due to Lemma 3.27. This implies following insights.

- $\gamma_{\text{SEPC}} \ngeq \gamma_{\text{PPTC}}$ because SEPC is convex and compact and PPTC is strictly larger than SEPC, due to the existence of PPT entangled states (recall that we view states as a special case of channels).
- For the same reason, $\gamma_{\text{SEP}}(J_{\mathcal{E}}) \ngeq \gamma_{\text{PPT}}(J_{\mathcal{E}})$.
- To see that $\gamma_{LOCC} \ge \gamma_{SEPC}$, notice that $LOCC \subset \overline{LOCC} \subset SEPC$ where \overline{LOCC} is the closure of LOCC and we used that SEPC is closed. Since there exist channels in SEPC which do not lie in \overline{LOCC} (this has been demonstrated in the context of state discrimination [12], [40], [97] and random distillation [42], [48]), one thus finds that

$$\gamma_{\text{LOCC}} \ge \gamma_{\overline{\text{LOCC}}} \not\ge \gamma_{\text{SEPC}}$$
. (5.241)

The inequality $\gamma_{LO^*} \ngeq \gamma_{LOCC^*}$ is a bit more involved and operationally very important, since it implies that classical communication is advantageous for the simulation of certain channels.

Lemma 5.53: There exist quantum channels
$$\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$$
 such that $\gamma_{\operatorname{LO}^\star}(\mathcal{E}) > \gamma_{\operatorname{LOCC}^\star}(\mathcal{E}) = 1$.

We briefly note that we will later even encounter an explicit quantum channel \mathcal{E} for which $1 < \gamma_{LOCC^*}(\mathcal{E}) < \gamma_{LO^*}(\mathcal{E})$ in Section 5.2.3.

Proof. Let us consider some \mathcal{E} that lies in LOCC($A; B \to A'; B'$) but not in conv(LO($A; B \to A'; B'$)). As, such, we clearly have $\gamma_{\text{LOCC}^*}(\mathcal{E}) = 1$. Assume that it would also hold that $\gamma_{\text{LO}^*}(\mathcal{E}) = 1$. This means that for any $\epsilon > 0$ there exists a QPD $\mathcal{E} = \sum_i a_i \mathcal{F}_i$ with $\mathcal{F}_i \in \text{LO}^*$ such that $\|a\|_1 \leq 1 + \epsilon$. For every \mathcal{F}_i , let

 $(\mathcal{G}_{j}^{(i)})_{j} \in \mathfrak{I}_{LO}(A; B \to A'; B')$ be the associated instrument s.t. $\mathcal{F}_{i} = \sum_{j} \beta_{j} \mathcal{G}_{j}^{(i)}$. Define the superoperator $\tilde{\mathcal{E}} := \sum_{i,j} |a_{i}| \mathcal{G}_{j}^{(i)}$. We clearly have

$$J_{\mathcal{E}} = \sum_{i,i} a_i \beta_j^{(i)} J_{\mathcal{G}_j^{(i)}} \preccurlyeq \sum_{i,j} |a_i| J_{\mathcal{G}_j^{(i)}} = J_{\tilde{\mathcal{E}}}$$
 (5.242)

and hence

$$||J_{\mathcal{E}} - J_{\tilde{\mathcal{E}}}||_1 = \operatorname{tr}\left[J_{\tilde{\mathcal{E}}} - J_{\mathcal{E}}\right]$$
(5.243)

$$= \sum_{i,j} |a_i| \operatorname{tr}[J_{\mathcal{G}_j^{(i)}}] - 1$$
 (5.244)

$$= \|a\|_1 - 1 \tag{5.245}$$

$$\leq \epsilon$$
. (5.246)

The channel $\frac{1}{\|a\|_1}\tilde{\mathcal{E}}$ lies in conv(LO) and approximates \mathcal{E} up to error

$$\left\| J_{\mathcal{E}} - \frac{1}{\|a\|_{1}} J_{\tilde{\mathcal{E}}} \right\|_{1} \le \|J_{\mathcal{E}} - J_{\tilde{\mathcal{E}}}\|_{1} + \left\| J_{\tilde{\mathcal{E}}} - \frac{1}{\|a\|_{1}} J_{\tilde{\mathcal{E}}} \right\|_{1} \le \epsilon + \frac{\epsilon}{1 + \epsilon} \le 2\epsilon.$$
 (5.247)

In summary, we have shown that any such \mathcal{E} can be approximated by an element of conv(LO) to arbitrary small error. But since conv(LO) is compact, this would also imply that $\mathcal{E} \in \text{conv}(\text{LO})$, which is a contradiction.

The argument in the proof can be recycled to show a separation between γ_{LOCC^*} and γ_{SEPC^*} since there exist separable channels which can't be arbitrary well approximated by LOCC channels (recall the separation between $\overline{\text{LOCC}}$ and SEPC). More precisely, for any $\mathcal{E} \in \text{SEPC}(AB \to A'B') \setminus \overline{\text{LOCC}}(AB \to A'B')$ we have

$$\gamma_{\text{LOCC}^{\star}}(\mathcal{E}) > \gamma_{\text{SEPC}^{\star}}(\mathcal{E}) = 1.$$
 (5.248)

It remains unclear, whether there exist quantum channels \mathcal{E} for which $\gamma_{LOCC}(\mathcal{E}) \neq \gamma_{LOCC^*}(\mathcal{E})$. This question amounts to whether classical side information is useful for circuit knitting with classical communication.

5.1.4.7 Asymptotic characterization with entanglement cost

The main achievement of Section 5.1.3.3 was to establish a connection between the quasiprobability extent of states and various entanglement cost measures. In this section, we will replicate this result as much as possible in the setting of channels. Unfortunately, its unknown if this can be achieved for the decomposition sets LO* and LOCC*. As such, we will restrict ourselves to studying $\overline{\gamma}_{\text{SEPC}}$, $\gamma_{\text{SEPC}}^{\infty}$, $\overline{\gamma}_{\text{PPTC}}$ and $\gamma_{\text{PPTC}}^{\infty}$, which at least provide interesting lower bounds.

Since the results in this section are completely analogous to the previously discussed case for states, we will introduce and summarize them in more brevity. Furthermore, we note that the dynamic resource theory of entanglement is much less developed for channels compared to states, so not all aspects are equally well understood.

We require some notion of the entanglement cost of a quantum channel, i.e., a characterization of how many Bell pairs are required to realize it. This is achieved by adapting Definition 5.25 by replacing states with channels, channels with superchannels and the trace distance with the diamond distance [66].

Definition 5.54: Let $\epsilon \geq 0$ and $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$. The *single-shot entangle-ment cost* of \mathcal{E} w.r.t. some set of superchannels S is defined as

$$E_{C,S}^{(1),\epsilon}(\mathcal{E}) := \min_{m \in \mathbb{N}} \{ \log m | \exists \mathcal{F} \in S : \frac{1}{2} \| \mathcal{E} - \mathcal{F}(|\Psi_m\rangle\langle\Psi_m|) \|_{\Diamond} \le \epsilon \}$$
 (5.249)

where $|\Psi_m\rangle\langle\Psi_m|$ denotes the maximally entangled state of Schmidt rank m and should be understood as a quantum channel with trivial input space.

The asymptotic entanglement cost and exact asymptotic entanglement cost are then given by

$$E_{C,S}(\mathcal{E}) := \lim_{\epsilon \to 0^+} \liminf_{n \to \infty} \frac{1}{n} E_{C,S}^{(1),\epsilon}(\mathcal{E}^{\otimes n}) \text{ and } E_{C,S}^{\text{exact}}(\mathcal{E}) := \lim_{n \to \infty} \frac{1}{n} E_{C,S}^{(1),0}(\mathcal{E}^{\otimes n}).$$
 (5.250)

We will study the entanglement cost under following two sets of superchannels. The separability-preserving superchannels SEPPSC($AB, A'B' \rightarrow CD, C'D'$) contain all superchannels $\Theta \in SC(AB, A'B' \rightarrow CD, C'D')$ such that

$$\forall \mathcal{E} \in \text{SEPC}(A; B \to A'; B') : \Theta(\mathcal{E}) \in \text{SEPC}(C; D \to C'; D'). \tag{5.251}$$

Furthermore, the set of *PPT superchannels* PPTSC($AB, A'B' \rightarrow CD, C'D'$) contains all $\Theta \in SC(AB, A'B' \rightarrow CD, C'D')$ which have a Choi matrix J_{Θ} with positive partial transpose $(id_{A,A',C,C'} \otimes T_{B,B',D,D'})(J_{\Theta}) \succcurlyeq 0$. It can be shown that PPTSC precisely constitutes the set of superchannels which completely preserve PPTC [67, Theorem 5].

The one-shot entanglement cost of $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ w.r.t. SEPPSC can be characterized by the smooth robustness

$$R_{\text{SEPC}}^{\epsilon}(\mathcal{E}) := \min_{\mathcal{F} \in B_{\epsilon}(\mathcal{E})} R_{\text{SEPC}}(\mathcal{F})$$
 (5.252)

where $\epsilon \geq 0$ and $B_{\epsilon}(\mathcal{E}) := \{ \mathcal{F} \in \operatorname{CPTP}(AB \to A'B') | \frac{1}{2} \| \mathcal{E} - \mathcal{F} \|_{\Diamond} \leq \epsilon \}.$

Lemma 5.55 ([93, Theorem III.1]): Let $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ and $\epsilon \geq 0$. Then

$$\log(1 + R_{\text{SEPC}}^{\epsilon}(\mathcal{E})) \le E_{C,\text{SEPPSC}}^{(1),\epsilon}(\mathcal{E}) \le \log(1 + R_{\text{SEPC}}^{\epsilon}(\mathcal{E})) + 2. \tag{5.253}$$

As a consequence, we get the statement corresponding to Theorem 5.31.

Theorem 5.56

For any bipartite channel $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$, the regularized and asymptotic quasiprobability extent w.r.t. SEPC can be expressed as

$$\log \overline{\gamma}_{\text{SEPC}}(\mathcal{E}) = E_{C,\text{SEPPSC}}^{\text{exact}}(\mathcal{E}) \quad \text{and} \quad \log \gamma_{\text{SEPC}}^{\infty}(\mathcal{E}) = E_{C,\text{SEPPSC}}(\mathcal{E}). \tag{5.254}$$

We omit the proof because it is completely analogous. To characterize the oneshot entanglement cost of a channel w.r.t. PPTSC, we need to define a dynamical resource measure called *max-logarithmic negativity* which was introduced in Ref. [67].

Definition 5.57: The max-logarithmic negativity of $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ is defined as $LN_{\max}(\mathcal{E}) := \max\{LN_{\max}^{(0)}(\mathcal{E}), LN_{\max}^{(1)}(\mathcal{E})\}$ where

$$2^{LN_{\max}^{(0)}(\mathcal{E})} := \begin{cases} \inf_{\Lambda \in P(ABA'B')} & \|\operatorname{tr}_{A'B'}[\Lambda]\|_{\infty} \operatorname{dim}(AB) \\ \text{s.t.} & -(\operatorname{id}_{AA'} \otimes T_{BB'})(\Lambda) \preccurlyeq (\operatorname{id}_{AA'} \otimes T_{BB'})(J_{\mathcal{E}}) \\ & (\operatorname{id}_{AA'} \otimes T_{BB'})(J_{\mathcal{E}}) \preccurlyeq (\operatorname{id}_{AA'} \otimes T_{BB'})(\Lambda) \end{cases} , (5.255)$$

$$2^{LN_{\max}^{(1)}(\mathcal{E})} := \begin{cases} \inf_{\Lambda \in P(ABA'B')} & \|(\mathrm{id}_A \otimes T_B)(\mathrm{tr}_{A'B'}[\Lambda])\|_{\infty} \dim(AB) \\ \text{s.t.} & -(\mathrm{id}_{AA'} \otimes T_{BB'})(\Lambda) \preccurlyeq (\mathrm{id}_{AA'} \otimes T_{BB'})(J_{\mathcal{E}}) \\ & (\mathrm{id}_{AA'} \otimes T_{BB'})(J_{\mathcal{E}}) \preccurlyeq (\mathrm{id}_{AA'} \otimes T_{BB'})(\Lambda) \end{cases} . (5.256)$$

Note that $LN_{\max}^{(0)}$ and $LN_{\max}^{(1)}$ are both defined in terms of an SDP and can hence be efficiently computed. It can be shown that $LN_{\max}(\mathcal{E})$ reduces to the κ -entanglement when $\dim(A) = \dim(B) = 1$ [67].

Lemma 5.58 ([67, Lemma 9]): Let $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ and $\epsilon \geq 0$. Then

$$2^{LN_{\max}(\mathcal{E})} - 1 \le 2^{E_{C,PPTSC}^{(1),0}(\mathcal{E})} \le 2^{LN_{\max}(\mathcal{E})} + 2.$$
 (5.257)

Lemma 5.59 ([89, Lemma 5]): For any $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ one has

$$\gamma_{\text{PPTC}}(\mathcal{E}) \ge 2^{LN_{\text{max}}(\mathcal{E})} - 1$$
(5.258)

Proof. Consider the achieving QPD $\mathcal{E}=(1+t)\mathcal{F}^+-t\mathcal{F}^-$ for $\mathcal{F}^\pm\in PPTC$. Since $J_{\mathcal{F}^\pm}\in PPT$ we have

$$-(1+t)(\mathbb{1}_{AA'} \otimes T_{BB'})(J_{\mathcal{F}^{+}} + J_{\mathcal{F}^{-}}) \le (\mathbb{1}_{AA'} \otimes T_{BB'})(J_{\mathcal{E}})$$
(5.259)

$$\leq (1+t)(\mathbb{1}_{AA'} \otimes T_{BB'})(J_{\mathcal{F}^+} + J_{\mathcal{F}^-}).$$
 (5.260)

Hence, $\tilde{\Lambda} := (1+t)(J_{\mathcal{F}^+} + J_{\mathcal{F}^-})$ is a feasible point for $LN_{\max}(\mathcal{E})$. Notice that

$$\frac{2(1+t)}{\dim(AB)}\mathbb{1}_{AB} = \operatorname{tr}_{A'B'}[\tilde{\Lambda}] = (\operatorname{id}_A \otimes T_B)(\operatorname{tr}_{A'B'}[\tilde{\Lambda}])$$
 (5.261)

which implies

$$2^{LN_{\max}(\mathcal{E})} \leq \max\left(\left\|\operatorname{tr}_{A'B'}[\tilde{\Lambda}]\right\|_{\infty}, \left\|(\operatorname{id}_{A} \otimes T_{B})(\operatorname{tr}_{A'B'}[\tilde{\Lambda}]\right\|_{\infty}\right) \dim(AB)$$
 (5.262)

$$\leq \max\left(\frac{2(1+t)}{\dim(AB)}, \frac{2(1+t)}{\dim(AB)}\right) \dim(AB) \tag{5.263}$$

$$= (1+2t)+1 \tag{5.264}$$

$$= \gamma_{\text{PPTC}}(\mathcal{E}) + 1. \tag{5.265}$$

Theorem 5.60

The regularized quasiprobability extent of $\mathcal{E} \in \operatorname{CPTP}(AB \to A'B')$ w.r.t. PPTC is given by

$$\log \overline{\gamma}_{PPTC}(\mathcal{E}) = \lim_{n \to \infty} \frac{1}{n} L N_{\text{max}}(\mathcal{E}^{\otimes n}) = E_{C,PPTSC}^{\text{exact}}(\mathcal{E}).$$
 (5.266)

Proof. The second equality follows directly from regularizing the statement in Lemma 5.58. The first equality follows by regularizing all terms

$$\log(2^{LN_{\max}(\mathcal{E})} - 1) \le \log \gamma_{\text{PPTC}}(\mathcal{E}) \le \log(2^{E_{\text{C,PPTC}}^{(1),0}(\mathcal{E}) + 1} - 1)$$
 (5.267)

where the second inequality follows from the monotonicity of γ_{PPTC} in the sense of a dynamical resource monotone (recall Proposition 3.43). More precisely, if we have a PPT superchannel Θ that realizes $\mathcal E$ from the maximally entangled state $|\Psi_d\rangle\langle\Psi_d|$, then $\gamma_{PPTC}(\mathcal E)$ can be at most equal to $\gamma_{PPTC}(|\Psi_d\rangle\langle\Psi_d|) = \gamma_{PPT}(|\Psi_d\rangle\langle\Psi_d|) = 2d-1$ where we used Lemma 5.15.

Previous work has claimed that LN_{\max} is an additive quantity [67], which would imply that $\lim_{n\to\infty}\frac{1}{n}LN_{\max}(\mathcal{E}^{\otimes n})=LN_{\max}(\mathcal{E})$ is efficiently computable. However, this is wrong by the same counter-example that the κ -entanglement is not additive. To our knowledge, there is no known method at this time to efficiently evaluate $\log\overline{\gamma}_{\mathrm{PPTC}}(\mathcal{E})$. It remains an open question whether the techniques from Proposition 5.38 can be adapted from the state setting to the channel setting.

Example 5.61. We again consider the noisy CNOT channel

$$\mathcal{E}_{p}(\rho) := (1 - p) \operatorname{CNOT}(\rho) + p \mathcal{D}(\rho)$$
 (5.268)

where \mathcal{D} is the fully depolarizing two-qubit channel. In Figure 5.12 we depict the numerical evaluation of $\gamma_{\mathrm{PPTC}}(\mathcal{E}_p)$ and $2^{LN_{\mathrm{max}}(\mathcal{E}_p)}$. The latter provides an upper bound for the regularized quasiprobability extent $\overline{\gamma}_{\mathrm{PPTC}}(\mathcal{E}_p)$ which must hence lie in the gray region. Remarkably, the max-logarithmic negativity coincides with the exact result for $\overline{\gamma}_{\mathrm{PPTC}}(\mathcal{E}_p)$ known at p=0 (from Example 5.42), hinting at the possibility that it might be additive for \mathcal{E}_p .

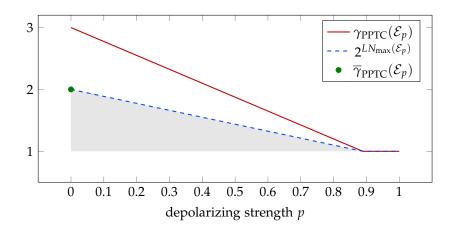


Figure 5.12: Numerical evaluation of $\gamma_{PPTC}(\mathcal{E}_p)$ and $2^{LN_{max}(\mathcal{E}_p)}$ for the depolarized CNOT gates \mathcal{E}_p

5.1.5 Exploiting strict sub-multiplicativity in the black box setting

The simplest approach for circuit knitting is to separately find space-like cuts for each individual nonlocal operation in a circuit, and then combine these into a QPD of the total circuit. However, treating all the nonlocal components individually is generally not optimal, and in some instances it is possible to find better QPDs of the circuit. For instance, we observed that the quasiprobability extents γ_{SEP} , γ_{LO^*} and γ_{LOCC^*} exhibit strictly sub-multiplicative behavior. Therefore, cutting multiple *parallel* instances of some operation (i.e. operations that occur in the same time slice of the circuit) is strictly cheaper than cutting them individually. Unfortunately, in practical situations, the nonlocal gates/states may not occur in the same time slice, but rather at possibly distant locations in the circuit. In this section, we will explore more refined QPS protocols that operate in this more realistic regime and how they can sometimes still achieve a sampling overhead that is the same as if the operations were all performed in parallel.

From now on, we will call this setup with non-parallel gates the *black box cut* setting to differentiate it from the *parallel cut setting*. The two settings, exemplified on three two-qubit gates, are depicted in Figure 5.13. The name black box cut is chosen to reflect that the circuit knitting protocol should work independently of what the operations \mathcal{B}_1 and \mathcal{B}_2 do, since we assume that we generally do not have any further information about them.

We now mathematically formalize the notion of the black box cut. Consider a quantum channel $\mathcal{G} \in \operatorname{CPTP}(AB)$ that is a composition of n nonlocal gates $\mathcal{E}_i \in \operatorname{CPTP}(A_iB_i)$, where A_i, B_i are subsystems of A, B, interleaved by some black box operations $\mathcal{B}_j \in \operatorname{CPTP}(AB)$

$$\mathcal{G} := \mathcal{E}_n \circ \mathcal{B}_{n-1} \circ \mathcal{E}_{n-1} \circ \mathcal{B}_{n-2} \circ \cdots \circ \mathcal{E}_2 \circ \mathcal{B}_1 \circ \mathcal{E}_1. \tag{5.269}$$

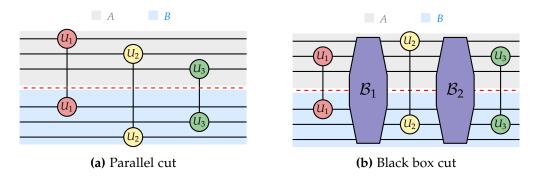


Figure 5.13: Graphical depiction of the two cutting scenarios for multiple space-like cuts. The dashed line represents the cut. The unitaries U_i are two-qubit gates acting on A and B. In the parallel cut setting, the unitaries act on distinct qubits and can be considered to occur in the same time slice (not depicted here for visual clarity). In the black box setting, B_1 and B_2 denote arbitrary unknown CPTP maps, which represent all operations in the circuit that occur between U_1 , U_2 and between U_2 , U_3 respectively.

We are interested in finding QPDs of G of the form

$$\mathcal{G}(\rho) = \sum_{i=1}^{m} a_i \operatorname{tr}_{E_A E_B} \left[\mathcal{F}_n^{(i)} \circ \mathcal{B}_{n-1} \circ \mathcal{F}_{n-1}^{(i)} \circ \mathcal{B}_{n-2} \circ \cdots \circ \mathcal{B}_1 \circ \mathcal{F}_1^{(i)} (\rho \otimes |0\rangle \langle 0|_{E_A E_B}) \right]$$
(5.270)

where the $\mathcal{F}_{j}^{(i)}$ are either in LO* $(A_{i}E_{A}; B_{i}E_{B})$ or in LOCC* $(A_{i}E_{A}; B_{i}E_{B})^{6}$. Here, the systems E_{A} and E_{B} capture the notion that one might use ancilla systems on both partitions to store information between the nonlocal gates \mathcal{E}_{i} . E_{A} and E_{B} are initialized in some arbitrary product state $|0\rangle\langle 0|_{E_{A}E_{B}}$. Graphically, this can be depicted as follows for n=3

$$\stackrel{AB}{\mathcal{E}_1} = \stackrel{B_1}{\mathcal{E}_2} = \stackrel{B_2}{\mathcal{E}_3} = \stackrel{B_3}{\mathcal{E}_3} = \stackrel{m}{\underset{i=1}{\sum}} a_i \qquad \stackrel{AB}{\underset{|0\rangle}{E_A E_B}} \stackrel{\mathcal{F}_1^{(i)}}{\mathcal{F}_1} = \stackrel{\mathcal{B}_1}{\mathcal{F}_2^{(i)}} = \stackrel{\mathcal{F}_2^{(i)}}{\mathcal{F}_3} = . \tag{5.271}$$

Crucially, we are interested in QPDs such that this equality holds for any choice of the black box channels \mathcal{B}_j . In summary, the optimal sampling overhead in the black box setting is captured by following quantity.

 $^{^6}$ We briefly note that choosing $\mathcal{F}_j^{(i)}$ to be either in LO* or LOCC* effectively means that the weighting of the classical side information is done independently for each gate. A priori, this is not the most general scheme imaginable, as the weighting could be done in a correlated fashion across the individual gates. However, this does not affect the lowest possible overhead, as such a correlated weighting scheme could also be effectively achieved by passing classical information between the operations $\mathcal{F}_j^{(i)}$ through the ancilla systems E_A, E_B .

Definition 5.62: Let $n \in \mathbb{N}$, $\mathcal{E}_i \in \operatorname{HP}(A_iB_i)$ for i = 1, ..., n and let S be either LO* or LOCC*. We define the *black box quasiprobability extent* $\gamma_S^{\operatorname{BB}}(\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_n)$ to be the infimum of the 1-norm $\|a\|_1$ over all valid black box QPDs with $\mathcal{F}_j^{(i)} \in S(A_iE_A; B_iE_B)$ for which Equation (5.270) holds for every system AB, for every embedding of A_iB_i into AB and for every choice of $\mathcal{B}_1, ..., \mathcal{B}_{n-1} \in \operatorname{CPTP}(AB)$.

We briefly note two trivial bounds for the black box quasiprobability extent.

Lemma 5.63: Let $n \in \mathbb{N}$, $\mathcal{E}_i \in HP(A_iB_i)$ for i = 1, ..., n and let S be either LO^* or $LOCC^*$. Then,

$$\gamma_S\left(\bigotimes_{i=1}^n \mathcal{E}_i\right) \le \gamma_S^{\mathrm{BB}}(\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n) \le \prod_{i=1}^n \gamma_S(\mathcal{E}_i).$$
(5.272)

Proof. Clearly, optimally cutting the individual gates separately of each other induces a valid black box cut. This implies the upper bound. For the lower bound, notice that by appropriately choosing the embedding and fixing the black box channels to be the identity, we can obtain $\mathcal{G} = \bigotimes_{i=1}^n \mathcal{E}_i$. Therefore, any black box cut directly induces a QPD for $\bigotimes_{i=1}^n \mathcal{U}_i$ with identical 1-norm.

To illustrate the simplest instance of how circuit knitting in the black box setting can operate, we start by considering that the nonlocal gates that we want to cut are all CNOT gates. Furthermore, we will for the moment consider the LOCC setting. In this case, the CNOT gates can be realized through the gate teleportation circuit (recall Figure 5.3) and the task reduces to the preparation of a shared Bell state, which can itself be simulated using QPS. In a circuit with multiple CNOT gates, these bell states could all be prepared at the same time at the very beginning of the circuit, as depicted in Figure 5.14. This in turns allows us to harvest the strictly sub-multiplicative behavior of $\gamma_{\rm SEP}$ to jointly prepare the Bell states more efficiently. This is fully analogous to the Clifford+T setting discussed in Example 3.45, where we used a magic state injection circuit and utilized the sub-multiplicative behavior of the quasiprobability extent to simulate the joint preparation of multiple magic states more efficiently.

Clearly, the strategy outlined above can be straightforwardly extended to any gate that can be teleported from a state, such as all Clifford gates. Together with Corollary 5.41 and Lemma 5.63, we obtain following result.

Proposition 5.64

The black-box quasiprobability extent of n bipartite Clifford gates U_1, \ldots, U_n is given by

$$\gamma_{\text{LOCC}^{\star}}^{\text{BB}}(\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n) = \gamma_{\text{LOCC}^{\star}}\left(\bigotimes_{i=1}^n \mathcal{U}_i\right).$$
(5.273)

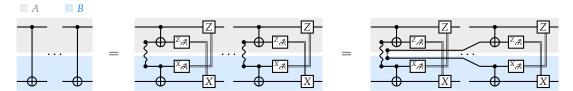


Figure 5.14: Two distant CNOT gates can be realized through gate teleportation under LOCC by each consuming a Bell pair (depicted as wavy line). By generating the two Bell pairs simultaneously (instead generating a single Bell pair twice), we can utilize the sub-multiplicativity of the quasiprobability extent under the tensor product to reduce the total sampling overhead.

Remark 5.65: The definition of the black box quasiprobability extent does not put any limit on the size of the ancilla systems E_A and E_B . This can become problematic for practical applications of circuit knitting, as the small number of qubits is the main motivation for circuit knitting in the first place. Indeed, we run into this issue with the black box cutting strategy for CNOT gates discussed above. The systems E_A and E_B contain the ancilla qubits that store the simulated Bell pairs, so their size grows linearly in the number of nonlocal CNOT gates. In practice, this problem can be addressed by picking a fixed number of ancilla qubits K on each side of the partition that are used to jointly generate K shared Bell pairs. Once all K Bell pairs are consumed by gate teleportations, the same qubits are re-used in another round of (simulated) preparation of K Bell pairs. The size K of such an "entanglement factory" allows for a trade-off: Increasing K means that more ancilla qubits are required, but also lowers the effective quasiprobability extent per CNOT to $(2^{K+1}-1)^{1/K}$. We note that a very small K is already sufficient to get very close to an effective extent of 2.

It is natural to ask at this point, whether optimal black box cutting (as given by Equation (5.273)) is also achievable for non-Clifford unitaries and without the usage of classical communication. At first, this might seem unlikely, given the crucial role of gate teleportation in the previous scheme. Somewhat surprisingly, the answer is positive for both questions, at least for general two-qubit unitaries.

Theorem 5.66

Let U_1, U_2, \dots, U_n be two-qubit unitary channels. Then

$$\gamma_{\mathrm{LO}^{\star}}^{\mathrm{BB}}(\mathcal{U}_{1},\mathcal{U}_{2},\ldots,\mathcal{U}_{n}) = \gamma_{\mathrm{LOCC}^{\star}}^{\mathrm{BB}}(\mathcal{U}_{1},\mathcal{U}_{2},\ldots,\mathcal{U}_{n}) = \gamma_{\mathrm{LO}^{\star}}(\mathcal{U}_{1}\otimes\mathcal{U}_{2}\otimes\ldots\mathcal{U}_{n}). \tag{5.274}$$

We note that the proof is constructive, i.e., we specify an explicit LO* protocol that achieves optimal black box cutting. The essential ingredient is the utilization of a gate teleportation-like protocol that works for non-Clifford gates and doesn't require any sort of classical communication. Instead, our protocol relies purely on the classical side information generated by mid-circuit measurements.

Proof. By Theorem 5.44 and Lemma 5.63, we directly have

$$\gamma_{\mathrm{LO}^{\star}}(\mathcal{U}_{1} \otimes \mathcal{U}_{2} \otimes \dots \mathcal{U}_{n}) \leq \gamma_{\mathrm{LOCC}^{\star}}^{\mathrm{BB}}(\mathcal{U}_{1}, \mathcal{U}_{2}, \dots, \mathcal{U}_{n}) \leq \gamma_{\mathrm{LO}^{\star}}^{\mathrm{BB}}(\mathcal{U}_{1}, \mathcal{U}_{2}, \dots, \mathcal{U}_{n}). \tag{5.275}$$

As such, it remains to show $\gamma_{LO^*}^{BB}(\mathcal{U}_1,\mathcal{U}_2,\ldots,\mathcal{U}_n) \leq \gamma_{LO^*}(\mathcal{U}_1\otimes\mathcal{U}_2\otimes\ldots\mathcal{U}_n)$, which we will do by explicitly constructing an LO* black box cut.

Let us denote the single-qubit Pauli strings by $Q_0 = 1$, $Q_1 = X$, $Q_2 = Y$, $Q_3 = Z$ and n-qubit Pauli strings by $Q_{i_1,i_2,...,i_n} := Q_{i_1} \otimes Q_{i_2} \otimes \cdots \otimes Q_{i_n}$. The unitary channel induced by a Pauli will be denoted $Q_i(\rho) := Q_i \rho Q_i$. The two-qubit Bell basis will be denoted by $|\Psi^i\rangle := (\sigma_i \otimes 1) |\Psi^0\rangle$ where $|\Psi^0\rangle := (|00\rangle + |11\rangle)/\sqrt{2}$. Let us furthermore define the two-qubit superoperator $\mathcal{D}_{\alpha,\beta}$ as follows

$$\mathcal{D}_{\alpha,\beta}(\rho) := Q_{\alpha,\alpha}\rho Q_{\beta,\beta}. \tag{5.276}$$

Recall the gate teleportation protocol that we discussed in Figure 5.7. This protocol allows us to realize any unitary channel from its Choi state, and in case the unitary is Clifford, then the protocol is LOCC. In the even more specialized case where the gate is a Pauli channel $\mathcal{Q}_{\alpha}\otimes\mathcal{Q}_{\beta}$, then the protocol is even in LO. We will extend this result by showing that $\mathcal{D}_{\alpha,\beta}$ can also be realized from its Choi state using an LO* protocol.

For this purpose, it is useful to first recapitulate the teleportation protocol for a Pauli channel $Q_{\alpha} \otimes Q_{\beta}$. We can write

$$(\mathcal{Q}_{\alpha} \otimes \mathcal{Q}_{\beta})(\rho_{\tilde{A}\tilde{B}}) = \sum_{i,j} T_{i,j} \left(\rho_{\tilde{A}\tilde{B}} \otimes (J_{\mathcal{Q}_{\alpha} \otimes \mathcal{Q}_{\beta}})_{ABA'B'} \right)$$
(5.277)

where $(T_{i,j})_{i,j} \in \mathfrak{I}_{LO}(AA'\tilde{A};BB'\tilde{B} \to A;B)$ is the quantum instrument describing the teleportation protocol which consists of

- 1. measuring AA' in the Bell basis, obtaining outcome $i \in \{0,1,2,3\}$,
- 2. measuring $\tilde{B}B'$ in the Bell basis, obtaining outcome $j \in \{0, 1, 2, 3\}$,
- 3. applying the operation $Q_{i,j}$ on AB.

Now, we turn our attention to the gate teleportation of $\mathcal{D}_{\alpha,\beta}$. Consider applying the same three steps as above, but now on the initial state $\rho_{\tilde{A}\tilde{B}} \otimes (J_{\mathcal{D}_{\alpha,\beta}})_{ABA'B'}$. By a standard state teleportation argument, the post-measurement state after obtaining the outcome i, j is given by $\mathcal{D}_{\alpha,\beta}(Q_{i,j}\rho Q_{i,j})$. After applying the correction operation $Q_{i,j}$ we hence have

$$Q_{i,j}\mathcal{D}_{\alpha,\beta}(Q_{i,j}\rho Q_{i,j})Q_{i,j} = Q_{i,j}Q_{\alpha,\alpha}Q_{i,j}\rho Q_{i,j}Q_{\beta,\beta}Q_{i,j}$$

$$= (-1)^{\llbracket Q_i,Q_{\alpha}\rrbracket + \llbracket Q_j,Q_{\alpha}\rrbracket + \llbracket Q_i,Q_{\beta}\rrbracket + \llbracket Q_j,Q_{\beta}\rrbracket}\mathcal{D}_{\alpha,\beta}(\rho).$$
(5.278)

$$= (-1)^{\|Q_i,Q_\alpha\| + \|Q_j,Q_\alpha\| + \|Q_i,Q_\beta\| + \|Q_j,Q_\beta\|} \mathcal{D}_{\alpha,\beta}(\rho). \tag{5.279}$$

The factor $f(i,j,\alpha,\beta) := (-1)^{[\![Q_i,Q_\alpha]\!]+[\![Q_j,Q_\alpha]\!]+[\![Q_i,Q_\beta]\!]+[\![Q_j,Q_\beta]\!]}$ can be compensated by appropriately weighting with ± 1 depending on i and j (recall that we are allowing LO^* operations, not just LO). In summary, we have shown that

$$\mathcal{D}_{\alpha,\beta}(\rho_{\tilde{A}\tilde{B}}) = \sum_{i,j} f(i,j,\alpha,\beta) T_{i,j} \left(\rho_{\tilde{A}\tilde{B}} \otimes (J_{\mathcal{D}_{\alpha,\beta}})_{ABA'B'} \right) . \tag{5.280}$$

Notice that $f(i, j, \alpha, \beta)$ is symmetric under exchange of α, β .

We now turn to proving the statement of the theorem. The KAK decomposition allows us to write the two-qubit unitaries U_m as

$$U_m = \sum_{k=0}^{3} v_k^{(m)} Q_k \otimes Q_k \tag{5.281}$$

for coefficients $v^{(m)} \in \mathbb{C}^4$. We separate the absolute value and the phase $v_k^{(m)} = u_k^{(m)} e^{i\phi_k^{(m)}}$ for $u_k^{(m)} \geq 0$, $\phi_k^{(m)} \in [0,2\pi)$. The induced channels are therefore

$$\mathcal{U}_m(\rho) = \sum_{k,l} u_k^{(m)} u_l^{(m)} e^{i(\phi_k^{(m)} - \phi_l^{(m)})} \mathcal{D}_{k,l}.$$
 (5.282)

Using the gate teleportation protocol in Equation (5.280), we can rewrite this as

$$\mathcal{U}_{m}(\rho) = \sum_{i,j,k,l} u_{k}^{(m)} u_{l}^{(m)} e^{i(\phi_{k}^{(m)} - \phi_{l}^{(m)})} f(i,j,k,l) T_{i,j} \left(\rho \otimes J_{\mathcal{D}_{k,l}} \right)$$
(5.283)

$$= \sum_{k,l} u_k^{(m)} u_l^{(m)} e^{i(\phi_k^{(m)} - \phi_l^{(m)})} \mathcal{T}^{(k,l)} \left(\rho \otimes J_{\mathcal{D}_{k,l}} \right) . \tag{5.284}$$

where $\mathcal{T}^{(k,l)} := \sum_{i,j} f(i,j,k,l) T_{i,j} \in LO^*(AA'\tilde{A};BB'\tilde{B} \to A;B)$.

The main idea for our black box cutting strategy is to apply the gate teleportation to every nonlocal gate U_m in the black box circuit

$$\mathcal{G} = \mathcal{U}_n \circ \mathcal{B}_{n-1} \circ \mathcal{U}_{n-1} \circ \mathcal{B}_{n-2} \circ \cdots \circ \mathcal{U}_2 \circ \mathcal{B}_1 \circ \mathcal{U}_1 \tag{5.285}$$

and then use the ancilla system to prepare all the Choi states of the form $J_{\mathcal{D}_{k,l}}$ together at the beginning of the circuit. This allows us to write

$$\mathcal{G} = \sum_{\boldsymbol{k},\boldsymbol{l} \in \mathbb{F}_{1}^{n}} \boldsymbol{u}_{\boldsymbol{k}} \boldsymbol{u}_{\boldsymbol{l}} e^{i\theta(\boldsymbol{k},\boldsymbol{l})} \Big(\mathcal{T}^{k_{n},l_{n}} \circ \mathcal{B}_{n-1} \circ \mathcal{T}^{k_{n-1},l_{n-1}} \circ \cdots \circ \mathcal{B}_{1} \circ \mathcal{T}^{k_{1},l_{1}} \Big) (\rho \otimes J_{\boldsymbol{D}_{\boldsymbol{k},\boldsymbol{l}}}) \quad (5.286)$$

where $\mathbb{F}_4 := \{0, 1, 2, 3\}$ and

$$u_{k} := u_{k_{1}}^{(1)} u_{k_{2}}^{(2)} \cdots u_{k_{n}}^{(n)},$$
 (5.287)

$$\theta(\mathbf{k}, \mathbf{l}) := \left(\phi_{k_1}^{(1)} - \phi_{l_1}^{(1)}\right) + \left(\phi_{k_2}^{(2)} - \phi_{l_2}^{(2)}\right) + \dots + \left(\phi_{k_n}^{(n)} - \phi_{l_n}^{(n)}\right), \tag{5.288}$$

$$D_{k,l} := D_{k_1,l_1} \otimes D_{k_2,l_2} \otimes \cdots \otimes D_{k_n,l_n}. \tag{5.289}$$

We re-arrange the sum to obtain

$$\mathcal{G} = \sum_{k \in \mathbb{F}_{4}^{n}} u_{k}^{2} \Big(\mathcal{T}^{k_{n},k_{n}} \circ \mathcal{B}_{n-1} \circ \mathcal{T}^{k_{n-1},k_{n-1}} \circ \cdots \circ \mathcal{B}_{1} \circ \mathcal{T}^{k_{1},k_{1}} \Big) (\rho \otimes J_{D_{k,k}})$$

$$+ \sum_{k < l} 4u_{k} u_{l} \Big(\mathcal{T}^{k_{n},l_{n}} \circ \mathcal{B}_{n-1} \circ \mathcal{T}^{k_{n-1},l_{n-1}} \circ \cdots \circ \mathcal{B}_{1} \circ \mathcal{T}^{k_{1},l_{1}} \Big) (\rho \otimes J_{\tilde{D}_{k,l}})$$
 (5.290)

where we used that $\mathcal{T}^{k,l} = \mathcal{T}^{l,k}$, k < l should be understood in lexicographic order and

$$\tilde{D}_{k,l} := \frac{1}{4} \left(e^{i\theta(k,l)} D_{k,l} + e^{-i\theta(k,l)} D_{l,k} \right). \tag{5.291}$$

Clearly, $D_{k,k} \in LO$, and further below, we will also show that $\tilde{D}_{k,l} \in LO^*$. This implies that Equation (5.290) provides a valid LO^* black box cut with a 1-norm of

$$\sum_{k} u_{k}^{2} + \sum_{k < l} 4u_{k}u_{l} = 2\left(\sum_{k} u_{k}\right)^{2} - \sum_{k} u_{k}^{2} = \gamma_{LO^{*}}(\mathcal{U}_{1} \otimes \mathcal{U}_{2} \otimes \dots \mathcal{U}_{n})$$
 (5.292)

where we used Theorem 5.44 and the fact that the Schmidt coefficients of $\bigotimes_{i=1}^{n} \mathcal{U}_i$ are precisely the u_k .

It now remains to show that $\tilde{D}_{k,l} \in LO^*$. Notice that we can write the operator Schmidt decomposition of $\bigotimes_{i=1}^n U_i$ as

$$\bigotimes_{i=1}^{n} U_{i} = \sum_{\mathbf{k}} \mathbf{u}_{\mathbf{k}} \prod_{j=1}^{n} e^{i\phi_{k_{j}}^{(j)}} Q_{\mathbf{k}} \otimes Q_{\mathbf{k}}.$$
 (5.293)

Following the construction in the proof of Lemma 5.46, the optimal LO* QPD of $\bigotimes_{i=1}^{n} \mathcal{U}_i$ is precisely

$$\bigotimes_{i=1}^{n} \mathcal{U}_{i} = \sum_{k} u_{k}^{2} \mathcal{Q}_{k} \otimes \mathcal{Q}_{k} + \sum_{k < l} 4u_{k} u_{l} \tilde{D}_{k,l}.$$
 (5.294)

Therefore, $\tilde{D}_{k,l} \in LO^*$ is a direct consequence from the proof of Lemma 5.46.

5.1.6 Applications

While our main focus has been the fundamental characterization of the overhead of circuit knitting, we will briefly touch upon some potential applications of space-like cuts in this section. It is not easy to find computational tasks that are well-suited for circuit knitting, as the involved quantum circuit needs to be separable into weakly interacting partitions. Asymptotically speaking, the number of space-like cuts needs to be logarithmic in the problem size, at least when the entangling strength of the cut gates is not vanishing.

Another consideration when trying to demonstrate some form of quantum advantage with circuit knitting is that the individual partitions of the circuit should be out of reach of classical simulation algorithms. Otherwise, the original computation can be classically simulated by simply simulating the circuit knitting procedure.

First, we will study the quantum Fourier transform, which is a central building block for some of the most important quantum algorithms such as phase estimation [95] and Shor's algorithms for prime factorization and the discrete logarithm [146]. Its action on an n-qubit computational basis state $|x\rangle$ for $x \in \{0, \ldots, 2^n - 1\}$ is defined as

$$|x\rangle \mapsto \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n - 1} \omega_{2^n}^{xy} |y\rangle$$
 (5.295)

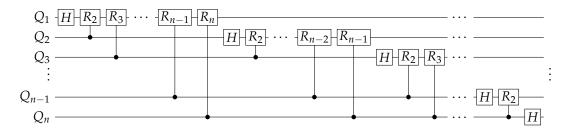


Figure 5.15: Circuit implementing quantum Fourier transform without qubit order inversion at the end. Here, R_i denotes the rotation gate $R_Z(\frac{2\pi}{2^i})$.

where $\omega_N := e^{\frac{2\pi i}{N}}$. It can be efficiently realized by the circuit in Figure 5.15 followed by layer of SWAP gates SWAP_{Q1Qn} SWAP_{Q2Qn-1} SWAP_{Q3Qn-3}... that inverts the order of the qubits. While the QFT is known to be very strongly entangling ⁷, it is entirely due to the qubit order reversal at the end [39]. In fact, if we take the circuit from Figure 5.15 and individually cut the various controlled rotation gates, we get a quasiprobability extent that does not diverge as a function of n.

Proposition 5.67

The quasiprobability extent of the n-qubit quantum Fourier transform without qubit order reversal QFT $_n$ is asymptotically constant in n

$$\gamma_{\mathrm{LO}^{\star}(A;B)}(\mathrm{QFT}_n) = \mathcal{O}(1) \tag{5.296}$$

w.r.t to the bipartition $A = Q_1 \dots Q_i$ and $B = Q_{i+1}, \dots, Q_n$ for any $1 \le i \le n-1$.

Proof. We prove this statement by taking the circuit in Figure 5.15 and separately cutting every controlled rotation gate which acts across the bipartition. This does most likely not produce the optimal QPD of QFT_n , but it provides us an upper bound to the quasiprobability extent.

A simple counting argument shows that no matter the value of n and the choice of the partition, the number of controlled R_{k+1} gates that we cut is at most k where

⁷The QFT is maximally entangling in the sense that its operator Schmidt decomposition has uniform singular values [117], [158].

 $k \ge 1$. Recalling the result in Example 5.47, we get

$$\gamma_{\mathrm{LO}^{\star}}(\mathrm{QFT}_n) \le \prod_{k=1}^{\infty} \gamma_{\mathrm{LO}^{\star}}(CR_{k+1})^k \tag{5.297}$$

$$= \prod_{k=1}^{\infty} \left(1 + 2\sin\frac{2\pi}{2^{k+1}} \right)^k \tag{5.298}$$

$$\leq \prod_{k=1}^{\infty} \left(1 + 2 \frac{2\pi}{2^{k+1}} \right)^k \tag{5.299}$$

$$= \prod_{k=1}^{\infty} \left(\frac{2^k + 2\pi}{2^k} \right)^k \tag{5.300}$$

$$= \exp_2 \left[\sum_{k=1}^{\infty} k \left(\log(2^k + 2\pi) - \log(2^k) \right) \right]$$
 (5.301)

where $\exp_2(x)$ denotes 2^x . Since log is concave, we can upper bound it with a firstorder Taylor approximation at any point, so

$$\gamma_{\mathrm{LO}^{\star}}(\mathrm{QFT}_n) \le \exp_2\left[\sum_{k=1}^{\infty} k2\pi \frac{1}{\ln(2)2^k}\right]$$
 (5.302)

$$= \exp_2\left[\frac{4\pi}{\ln 2}\right] \tag{5.303}$$
$$= e^{4\pi} \tag{5.304}$$

$$=e^{4\pi} \tag{5.304}$$

$$\approx 286751 \tag{5.305}$$

where we used
$$\sum_{k=1}^{\infty} \frac{k}{2^k} = 2$$
.

Beside the upper bound in the proof, we also numerically evaluated the quantity $\prod_{k=1}^{\infty} \gamma_{\text{LO}^*}(CR_{k+1})^k$ and obtained a value of roughly ≈ 26498 , which is more than an order of magnitude smaller than $e^{4\pi}$.

It is an interesting open question whether better QPDs of the quantum Fourier transform (or at least better estimates of its quasiprobability extent) can be obtained. Related results in literature have shown that the singular values of the operator Schmidt decomposition of the quantum Fourier transformation (without qubit reversal) are exponentially decaying [39]. Unfortunately, this does not directly provide us with some means of upper bounding the quasiprobability extent, as the operator Schmidt decomposition is not known to be of the unitary form required for Theorem 5.44.

Next, we briefly consider Hamiltonian simulation as a second application of space-like cuts, which has been thoroughly studied by Harrow and Lowe. We informally summarize their result as follows.

Theorem 5.68: [74, Theorem 3.12] (informal, simplified)

Consider a bipartite Hamiltonian $H_{tot} \in H(AB)$ that can be separated into two local parts and a weak interaction term

$$H_{\text{tot}} = H_A \otimes \mathbb{1}_B + \mathbb{1}_A \otimes H_B + H_{\text{int}} \tag{5.306}$$

where H_A , H_B and H_{int} are sums of Pauli terms acting on $\mathcal{O}(1)$ qubits. Circuit knitting can simulate the time evolution operator $\mathcal{T}(\rho) := e^{-iH_{\text{tot}}t}\rho e^{iH_{\text{tot}}t}$ up to error ϵ with efficient circuits and a sampling overhead $\mathcal{O}(e^{8\eta t}/\epsilon^2)$ where $\eta \leq \|H_{\text{int}}\|$.

For the full formal statement, we refer the reader to Ref. [74]. The main idea is to use the standard Trotterization approach to approximate \mathcal{T} with interleaved local and interaction operations. Trotterized Hamiltonian simulation has 2 freely choosable parameters: The simulation time t and the number of steps N. Remarkably, the quasiprobability extent of the Trotterized circuit only grows with t, but not with N. The reason for this is that, while the number of nonlocal gates increases linearly with N, their interaction strength also decreases with N.

As a simple model to illustrate this effect, consider the two-qubit Pauli operator $Z \otimes Z$ as our interaction term. The time evolution induced by this interaction is of the form of a rotation gate $R_{ZZ}(\theta) := e^{-i\frac{\theta}{2}Z\otimes Z}$. We previously showed in Example 5.47 that $\gamma_{LO^*}(R_{ZZ}(\theta)) = 1 + 2|\sin\theta|$. If we sub-divide the rotation into N smaller rotations which we cut individually, then the resulting sampling overhead converges to some fixed value as N grows to infinity

$$\gamma_{\mathrm{LO}^{\star}}(\mathrm{R}_{\mathrm{ZZ}}(\theta/N))^{N} = \left(1 + 2\sin\frac{|\theta|}{N}\right)^{N} \tag{5.307}$$

$$\approx \left(1 + \frac{2\theta}{N}\right)^N \tag{5.308}$$

$$\xrightarrow{N \to \infty} \exp(2\theta). \tag{5.309}$$

In this section, we have focused on applications of circuit knitting where we partitioned a large circuit into two or more smaller circuits. There is also a much more mundane, but perhaps more widely applicable usage of the technique. Many experimental quantum computing platforms exhibit strong locality constraints. For instance, superconducting quantum chips typically only allow for two-qubit interactions between neighboring qubits. As such, two-qubit gates between distant qubits require expensive swap chains that scale in depth with the distance. This difficulty remains in the fault-tolerant settings, at least when the logical qubits are encoded in a topological code like the surface code or color code. As such, it could be useful to locally simulate a few of these nonlocal operations, foregoing the need for swap chains at a cost of an increased sampling overhead. If quantum resources are more limited than the available time (i.e. number of shots), then this trade-off can potentially be advantageous.

5.2 Time-like cuts

We have previously seen that space-like cuts are deeply connected to the resource theory of entanglement. Remarkably, there is a similar connection for time-like cuts, with the lesser known *resource theory of quantum memories*. This is a dynamical resource theory where the non-free channels are the ones with the capability of preserving entanglement. Correspondingly, the free operations are called *entanglement breaking*.

Definition 5.69: A map $\mathcal{E} \in \operatorname{CPTP}(A \to A')$ is called entanglement breaking if $(\mathcal{E} \otimes \operatorname{id}_B)(\rho_{AB}) \in \operatorname{SEP}(A'; B)$ for any system B and density operator $\rho \in \operatorname{D}(AB)$. We denote the set of such entanglement breaking channels by $\operatorname{EBC}(A \to A')$.

In some ways, the resource theory of quantum memories can be considered the counterpart to the resource theory of entanglement as it studies quantum correlations in time rather than spatial correlations. It has found applications for the certification and benchmarking of quantum memories [137], [180].

In the context of circuit knitting, we will study decompositions of the identity channel

$$id = \sum_{i} a_i \mathcal{F}_i. {(5.310)}$$

A major question is what decomposition set we will allow for the channels \mathcal{F}_i to lie in. Clearly, the \mathcal{F}_i must be entanglement breaking, as otherwise they could generate entanglement across the circuit partitions. In fact, the set EBC is a very natural decomposition set, as it precisely contains the channels that can be realized with a measurement, followed by the preparation of a state that depends on the measurement outcome.

Proposition 5.70: [80, Theorem 2]

A channel $\mathcal{E} \in \operatorname{CPTP}(A \to A')$ is entanglement breaking if and only if it can be written in the form

$$\mathcal{E}(\rho) = \sum_{k} \operatorname{tr}[M_{k}\rho]\sigma_{k} \tag{5.311}$$

for some $(M_k)_k \in POVM(A)$ and $\sigma_k \in D(A')$.

As previously discussed in Figure 5.2, decomposing the identity into such channels allows us to realize circuit knitting. The choice of EBC as our decomposition set necessitates that our circuit partitions can exchange classical communication during the circuit execution. This is a requirement that we might not want to impose. Similarly to the study of the decomposition sets LO and LOCC for space-like cuts, we also introduce a second decomposition set that reflects the constraint that the two involved circuit partitions cannot exchange classical communication. It consists of measurement and prepare channels $\mathcal{E} \in \text{CPTP}(A \to A')$, which do not allow for the

prepared state to depend on the measurement result

$$\mathcal{E}(\rho) = \sum_{k} \text{tr}[M_k \rho] \sigma \tag{5.312}$$

for some $(M_k)_k \in \text{POVM}(A)$ and $\sigma \in D(A')$. We denote the set of such channels by MPC $(A \to A')$. At first glance, these channels might seem completely useless, as any information from their input is completely lost, as $\mathcal{E}(\rho) = \text{tr}[\rho]\sigma$. We will see in the next section that these channels can become useful if we allow for classical side information that is generated from the measurement. So while MPC $(A \to A')$ lacks the expressibility required for circuit knitting, the set MPC* $(A \to A')$ does not.

The main goal of this section is to identify the quasiprobability extent of the identity channel $\gamma_{EBC^*}(id)$ and $\gamma_{MPC^*}(id)$, which in turn characterizes the sampling overhead of circuit knitting with time-like cuts. We will also briefly touch upon the quasiprobability extent of other channels, but it will be a lesser focus since they are operationally not as important.

We note that time-like cuts can be considered a special case of space-like cuts by choosing the output system on one partition and the input system on the other partition to be trivial. As such, one can readily verify that $EBC(A \rightarrow B') = LOCC(A; B \rightarrow A'B')$, $MPC(A \rightarrow B') = LO(A; B \rightarrow A'; B')$, and correspondingly $EBC^*(A \rightarrow B') = LOCC^*(A; B \rightarrow A'; B')$, $MPC^*(A \rightarrow B') = LO^*(A; B \rightarrow A'; B')$, in the case that dim(B) = dim(A') = 1. This will however prove to be of limited use to us, as our results on space-like cuts are not powerful enough to cover the time-like cuts that we will study in this section.

Astute readers might also have noticed that MPC and EBC do not actually constitute valid QRTs as per Definition 2.8, as the identity channel is not free. This issue can be fixed by identifying the channels in MPC and EBC with their associated states through the Choi isomorphism. Under this new perspective, channels get promoted (or rather *demoted*) to states, and superchannels to channels. As such, the identity channel (the demoted identity superchannel) is now free, providing us with a valid quantum resource theory. This subtlety will be of no further importance in this section, besides the fact that we need to be a bit careful when applying results from Section 3.5.

5.2.1 Decomposition sets

Here, we introduce the decomposition sets of interest with full mathematical rigor. To capture the notion of classical side information through intermediate measurements, we start by defining the associated sets of quantum instruments. Note that in principle, both the sender and receiver have the capability of generating classical side information. In the case without classical communication, the most general setup consists of the sender doing a POVM measurement on the incoming state, and the receiver randomly preparing some state σ_j with probability p_j . The index of the chosen state j can serve as classical side information as well. We thus define

 $\mathfrak{I}_{\mathrm{MPC}}(A \to A') \subset \mathrm{QI}(A \to A')$ as the set of all instruments $(\mathcal{E}_{i,j})_{i,j}$ of the form⁸

$$\mathcal{E}_{i,j}(\rho) = \operatorname{tr}[M_i \rho] p_j \sigma_j \tag{5.313}$$

where $(M_i)_i \in \text{POVM}(A)$, $(p_j)_j$ is a probability vector and $\sigma_j \in D(A')$. In the setup with classical communication, the randomly prepared state of the receiver may additionally depend on the measurement outcome of the POVM. As such, we define $\mathfrak{I}_{EBC}(A \to A') \subset \text{QI}(A \to A')$ as all instruments $(\mathcal{E}_i)_i$ of the form

$$\mathcal{E}_i(\rho) = \operatorname{tr}[M_i \rho] \sigma_i \tag{5.314}$$

as well as all coarse-grainings of such instruments, where $(M_i)_i \in \text{POVM}(A)$ and $\sigma_i \in D(A')$. Note that without loss of generality, we could assume that the receiver prepares a deterministic state σ_i given the POVM outcome i, as the case of a randomly chosen state can be captured by fine-graining the POVM elements M_i .

We can directly recover EBC($A \rightarrow A'$) = chan[$\mathfrak{I}_{EBC}(A \rightarrow A')$] and MPC($A \rightarrow A'$) = chan[$\mathfrak{I}_{MPC}(A \rightarrow A')$]. For practical purposes, we are more interested in the decomposition sets

$$MPC^{*}(A \to A') := chan^{*}[\mathfrak{I}_{MPC}(A \to A')]$$

$$= \left\{ \rho \mapsto \sum_{i,j} \beta_{i,j} \operatorname{tr}[M_{i}\rho] p_{j}\sigma_{j} \middle| \begin{array}{l} |\beta_{i,j}| \leq 1, (M_{i})_{i} \in POVM(A), \\ p_{j} \text{ probabilities }, \sigma_{j} \in D(A') \end{array} \right\}$$
(5.316)

and

$$EBC^{*}(A \to A') := chan^{*}[\mathfrak{I}_{EBC}(A \to A')]$$

$$= \{ \rho \mapsto \sum_{i} \beta_{i} \operatorname{tr}[M_{i}\rho]\sigma_{i} | |\beta_{i}| \leq 1, (M_{i})_{i} \in POVM(A), \sigma_{i} \in D(A') \}$$

$$(5.318)$$

which capture the possibility of classical side information. Since $\mathfrak{I}_{EBC}(A \to A')$ is closed under mixture (the sender can randomly choose between two POVMs to perform and broadcast that choice to the receiver), EBC* is absolutely convex by Lemma 3.13.

We note that our definition of MPC* is slightly more general than the decomposition basis

$$\underline{\mathsf{MPC}}(A \to A') \coloneqq \{ \rho \mapsto \mathsf{tr}[O\rho]\sigma | O \in \mathsf{H}(A), \sigma \in \mathsf{H}(A'), \|O\|_{\infty} \le 1, \|\sigma\|_{1} \le 1 \}$$

$$(5.319)$$

introduced in Ref. [29]. Indeed, our construction of MPC* utilizes the full generality of classical side information by allowing the weighting to be jointly dependent on the classical side information of both the sender and receiver, whereas

$$\underline{MPC}(A \to A') = \{ \rho \mapsto \sum_{i,j} \operatorname{tr}[\alpha_i \rho] \beta_j p_j \sigma_j | |\alpha_i| \le 1, |\beta_j| \le 1 \}$$
 (5.320)

⁸More axiomatically, the sets $\mathfrak{I}_{MPC}(A \to B')$ and $\mathfrak{I}_{EBC}(A \to B')$ could equivalently also be defined as $\mathfrak{I}_{LO}(A; B \to A'; B')$ and $\mathfrak{I}_{LOCC}(A; B \to A'; B')$ where $\dim(A') = \dim(B) = 1$.

doesn't allow for correlated weighting. This discrepancy is essentially the same as the one between LO* and Equation (5.6) that we already mentioned previously. Clearly, $\underline{\mathrm{MPC}}(A \to A') \subset \mathrm{MPC}^*(A \to A')$, but the two sets are generally not equal. Fortunately, as we will see, the discrepancy is of no further consequence for the study of the quasiprobability extent. In fact, it will at times be more convenient to work with $\underline{\mathrm{MPC}}(A \to A')$ for technical reasons.

Working directly with the channel representations of these various sets can become somewhat unwieldy. It can be useful to consider the associated Choi representations instead.

Proposition 5.71

Let $\mathcal{E} \in HP(A \to A')$. Then

- $\mathcal{E} \in \mathrm{MPC}(A \to A') \iff J_{\mathcal{E}} = \frac{\mathbb{1}_A}{\dim(A)} \otimes \sigma \text{ for some } \sigma \in \mathrm{D}(A').$
- $\mathcal{E} \in \underline{\mathrm{MPC}}(A \to A') \Longleftrightarrow J_{\mathcal{E}} = \frac{O}{\dim(A)} \otimes \sigma$ for some $O \in \mathrm{H}(A)$, $\|O\|_{\infty} \le 1$ and $\sigma \in \mathrm{H}(A')$, $\|\sigma\|_1 \le 1$.

•
$$\mathcal{E} \in MPC^*(A \to A') \iff J_{\mathcal{E}} = \frac{1}{\dim(A)} \sum_{i} \left(M_i \otimes \left(\sum_{j} \beta_{i,j} p_j \sigma_j \right) \right)$$

$$= \frac{1}{\dim(A)} \sum_{j} p_j \left(\sum_{i} \beta_{i,j} M_i \right) \otimes \sigma_j$$

for some $(M_i)_i \in POVM(A)$, probabilities $(p_j)_j$, $|\beta_{i,j}| \leq 1$ and $\sigma_j \in D(A')$.

- $\mathcal{E} \in EBC(A \to A') \iff J_{\mathcal{E}} \in SEP(A; A') \text{ and } tr_{A'}[J_{\mathcal{E}}] = \frac{\mathbb{1}_A}{\dim(A)}$.
- $\mathcal{E} \in \mathrm{EBC}^*(A \to A') \iff J_{\mathcal{E}} = \rho^+ \rho^$ where $\rho^{\pm} \in [0,1] \cdot \mathrm{SEP}(A;A')$ are sub-normalized separable states and $\mathrm{tr}_{A'}[\rho^+ + \rho^-] = \frac{\mathbb{1}_A}{\dim(A)}$.

The statement about the Choi characterization of EBC is originally due to Horodecki, Shor and Ruskai [80, Theorem 4].

Proof. Consider a quantum instrument $(\mathcal{E}_{i,j})_{i,j} \in \mathfrak{I}_{MPC}(A \to A')$ of the form $\mathcal{E}_{i,j}(\rho) = \operatorname{tr}[M_i \rho] p_j \sigma_j$. Its Choi representation is as follows

$$(\mathrm{id}\otimes\mathcal{E}_{i,j})\left(\frac{1}{\dim(A)}\sum_{k,l}|kk\rangle\langle ll|\right)=\frac{1}{\dim(A)}\sum_{k,l}|k\rangle\langle l|\otimes\left(\mathrm{tr}[M_i|k\rangle\langle l|]p_j\sigma_j\right) \quad (5.321)$$

$$= \frac{p_j}{\dim(A)} \left(\sum_{k,l} \operatorname{tr}[M_i | k \rangle \langle l |] | k \rangle \langle l | \right) \otimes \sigma_j \qquad (5.322)$$

$$= \frac{p_j}{\dim(A)} M_i^T \otimes \sigma_j. \tag{5.323}$$

As such, the Choi representations of a map $\mathcal{E} \in MPC(A \to A')$ is given by

$$\sum_{i,j} \frac{p_j}{\dim(A)} M_i^T \otimes \sigma_j = \frac{\mathbb{1}_A}{\dim(A)} \otimes \left(\sum_j p_j \sigma_j\right). \tag{5.324}$$

Conversely, every state $\frac{\mathbb{1}_A}{\dim(A)} \otimes \sigma$ can be written in this form by appropriately choosing p_j, σ_j . Analogously, the Choi representation of a map $\mathcal{E} \in \underline{\mathrm{MPC}}(A \to A')$ is given by

$$\sum_{i,j} \alpha_i \beta_j \frac{p_j}{\dim(A)} M_i^T \otimes \sigma_j = \frac{1}{\dim(A)} \left(\sum_i \alpha_i M_i^T \right) \otimes \left(\sum_j p_j \beta_j \sigma_j \right). \tag{5.325}$$

This can be made equal to $\frac{1}{\dim(A)}O\otimes\sigma$ for any O,σ such that $\|O\|_{\infty}$, $\|\sigma\|_1\leq 1$ by appropriately choosing the M_i,α_i,p_j,β_j and σ_j . Finally, the Choi representation of a map $\mathcal{E}\in \mathsf{MPC}^*(A\to A')$ is precisely

$$\sum_{i,j} \beta_{i,j} \frac{p_j}{\dim(A)} M_i^T \otimes \sigma_j = \sum_i \left(\frac{M_i^T}{\dim(A)} \otimes \left(\sum_j \beta_{i,j} p_j \sigma_j \right) \right). \tag{5.326}$$

Next, we look at a quantum instrument $(\mathcal{E}_i)_i \in \mathfrak{I}_{EBC}(A \to A')$ of the form $\mathcal{E}(\rho) = \text{tr}[M_i \rho] \sigma_i$. By a similar calculation as above, its Choi representation is

$$(\mathrm{id} \otimes \mathcal{E}_i) \left(\frac{1}{\dim(A)} \sum_{k,l} |kk\rangle\langle ll| \right) = \frac{1}{\dim(A)} \sum_{k,l} |k\rangle\langle l| \otimes (\mathrm{tr}[M_i |k\rangle\langle l|]\sigma_i)$$
 (5.327)

$$=\frac{M_i^1}{\dim(A)}\otimes\sigma_i. \tag{5.328}$$

As such the maps $\mathcal{E} \in EBC(A \to A')$ are of the form

$$\sum_{i} \frac{M_i^T}{\dim(A)} \otimes \sigma_i = \sum_{i} \frac{\operatorname{tr}[M_i]}{\dim(A)} \frac{M_i^T}{\operatorname{tr}[M_i]} \otimes \sigma_i.$$
 (5.329)

These maps are clearly separable and have partial trace $\sum_i \frac{M_i^T}{\dim(A)} = \frac{\mathbb{1}_A}{\dim(A)}$. Conversely, a general separable state in SEP(A; A') can be written as $\sum_i p_i |f_i\rangle\langle f_i| \otimes |g_i\rangle\langle g_i|$. If the state additionally fulfills the partial trace constraint, then $\sum_i p_i |f_i\rangle\langle f_i| = \frac{\mathbb{1}_A}{\dim(A)}$. The above calculation shows that this state is the Choi representation of the map

$$\mathcal{E}'(\rho) := \sum_{i} \operatorname{tr}[\dim(A) p_i | f_i \rangle \langle f_i | \rho] | g_i \rangle \langle g_i | . \tag{5.330}$$

This map \mathcal{E}' lies in EBC($A \to A'$) since $\sum_i \dim(A) p_i |f_i\rangle\langle f_i| = \mathbb{1}_A$.

Finally, the characterization of EBC*($A \rightarrow A'$) is a direct consequence of Lemma 3.9: Any $\mathcal{E} \in EBC^*(A \rightarrow A')$ can be written as $\mathcal{E}^+ - \mathcal{E}^-$ for some instrument $(\mathcal{E}^+, \mathcal{E}^-) \in \mathfrak{I}_{EBC}$. That is, there exists a POVM $(M_i)_{i=1,\dots,m}$ and states $(\sigma_i)_{i=1,\dots,m}$ and some $k \in \{0,\dots,m-1\}$ such that

$$\mathcal{E}^{+}(\rho) = \sum_{i=1}^{k} \operatorname{tr}[M_{i}\rho]\sigma_{i} \quad \text{and} \quad \mathcal{E}^{-}(\rho) = \sum_{i=k+1}^{m} \operatorname{tr}[M_{i}\rho]\sigma_{i}.$$
 (5.331)

By the same argument as for the Choi characterization of EBC($A \to A'$), $J_{\mathcal{E}^+}$ and $J_{\mathcal{E}^-}$ are sub-normalized separable states and $J_{\mathcal{E}^++\mathcal{E}^-} \in SEP(A;A')$ fulfills the partial trace constraint. For the converse, consider some general sub-normalized states ρ^{\pm} which are separable

$$\rho^{+} = \sum_{i} p_{i}^{+} |f_{i}^{+}\rangle\langle f_{i}^{+}| \otimes |g_{i}^{-}\rangle\langle g_{i}^{-}| \quad \text{and} \quad \rho^{-} = \sum_{j} p_{j}^{-} |f_{j}^{-}\rangle\langle f_{j}^{-}| \otimes |g_{j}^{-}\rangle\langle g_{j}^{-}| \quad (5.332)$$

and fulfill the partial trace constraint

$$\sum_{i} p_{i}^{+} |f_{i}^{+}\rangle\langle f_{i}^{+}| + \sum_{j} p_{j}^{-} |f_{i}^{-}\rangle\langle f_{i}^{-}| = \frac{\mathbb{1}_{A}}{\dim(A)}.$$
 (5.333)

Then $\rho^+ - \rho^-$ is the Choi representation of the channel

$$\mathcal{E}'(\rho) := \sum_{i} \operatorname{tr}[\dim(A) p_{i}^{+} | f_{i}^{+} \rangle \langle f_{i}^{+} | \rho] | g_{i}^{+} \rangle \langle g_{i}^{+} |$$
 (5.334)

$$+\sum_{j}(-1)\operatorname{tr}[\dim(A)p_{j}^{-}|f_{j}^{-}\rangle\langle f_{j}^{-}|\rho]|g_{j}^{-}\rangle\langle g_{j}^{-}|$$
 (5.335)

which lies in EBC* since

$$\sum_{i} \dim(A) p_{i}^{+} |f_{i}^{+}\rangle\langle f_{i}^{+}| + \sum_{j} \dim(A) p_{j}^{-} |f_{j}^{-}\rangle\langle f_{j}^{-}| = \mathbb{1}_{A}.$$
 (5.336)

The Choi characterization of MPC * is more complicated than the other sets. Luckily, MPC * and MPC are equivalent in terms of their convex hulls.

Corollary 5.72: For any systems A, A' one has

$$\operatorname{conv}(\underline{\operatorname{MPC}}(A \to A')) = \operatorname{conv}(\operatorname{MPC}^{\star}(A \to A')). \tag{5.337}$$

Since the quasiprobability extent only depends on the convex hull of the decomposition set (recall Lemma 3.15), this corollary implies that it doesn't matter whether we work with γ_{MPC} or γ_{MPC^*} .

Proof. Since $\underline{MPC}(A \to A') \subset MPC^*(A \to A')$, it suffices to show that $MPC^*(A \to A') \subset \text{conv}(\underline{MPC}(A \to A'))$. This is evident from the Choi characterization in Proposition 5.71.

5.2.2 Utility of classical side information

In this section, we will study the importance of intermediate measurements and classical side information for the achievability of time-like cuts. This discussion will mirror the one about space-like cuts in Section 5.1.2. As alluded to previously, the channels $MPC(A \rightarrow A')$ without classical side information are useless for the task for circuit knitting.

Lemma 5.73: For any systems A, A' one has

$$\operatorname{span}_{\mathbb{R}}(\operatorname{MPC}(A \to A')) = \{ \rho \mapsto \operatorname{tr}[\rho]\sigma | \sigma \in \operatorname{H}(A') \}$$
 (5.338)

$$\operatorname{span}_{\mathbb{R}}(\operatorname{EBC}(A \to A')) = \mathbb{R}\operatorname{HPTP}(A \to A') \tag{5.339}$$

and

$$\operatorname{span}_{\mathbb{R}}(\operatorname{MPC}^{\star}(A \to A')) = \operatorname{span}_{\mathbb{R}}(\operatorname{EBC}^{\star}(A \to A')) = \operatorname{HP}(A \to A'). \quad (5.340)$$

The proof of this statement follows quite directly from the Choi characterizations in Proposition 5.71. We refrain from stating the full proof, as it essentially follows the identical arguments as Lemma 5.7 and Lemma 5.8. In fact, Lemma 5.73 can be considered a special instance of these two results.

To summarize, we have shown that classical side information through intermediate measurements is absolutely crucial for time-like cuts without classical communication. In contrast, when classical communication is available, circuit knitting remains possible even without classical side information. In fact, one can even show that in this case intermediate measurements provide no advantage in reducing the overhead.

Proposition 5.74

Let
$$\mathcal{E} \in \mathbb{R}HPTP(A \to A')$$
. Then $\gamma_{EBC}(\mathcal{E}) = \gamma_{EBC^*}(\mathcal{E})$.

The proof follows the same chain of arguments as the one of Proposition 5.10.

Proof. By Lemma 3.10 we know $\gamma_{EBC}(\mathcal{E}) \geq \gamma_{EBC^{\star}}(\mathcal{E})$, so it remains to show the converse. By the absolute convexity of EBC* and Lemma 3.14, we can without loss of generality assume all QPDs w.r.t. EBC* to only have one element. Let $\mathcal{E} = \kappa \mathcal{F}$, $\mathcal{F} \in EBC^{\star}$ be such a QPD. By the Choi characterization in Proposition 5.71 we can write $J_{\mathcal{F}} = \sigma^+ - \sigma^-$ for two sub-normalized separable states σ^+, σ^- fulfilling

$$\operatorname{tr}_{A'}[\sigma^+ + \sigma^-] = \frac{\mathbb{1}_A}{\dim(A)}.$$
 (5.341)

At the same time,

$$\operatorname{tr}_{A'}[\sigma^{+} - \sigma^{-}] = \frac{1}{\kappa} \operatorname{tr}_{A'}[J_{\mathcal{E}}] = \frac{\operatorname{tr}[J_{\mathcal{E}}]}{\kappa} \frac{\mathbb{1}_{A}}{\operatorname{dim}(A)}.$$
 (5.342)

This implies that both $\operatorname{tr}_{A'}[\sigma^+]$ and $\operatorname{tr}_{A'}[\sigma^-]$ are proportional to $\frac{\mathbb{1}_A}{\dim(A)}$. As such, $\frac{\sigma^+}{\operatorname{tr}[\sigma^+]}$ and $\frac{\sigma^-}{\operatorname{tr}[\sigma^-]}$ are the Choi representations of two channels in $\operatorname{EBC}(A \to A')$ (in the case $\operatorname{tr}[\sigma^\pm] = 0$, we define $\frac{\sigma^+}{\operatorname{tr}[\sigma^+]}$ to be zero) and we have a valid QPD w.r.t EBC

$$J_{\mathcal{E}} = \kappa \operatorname{tr}[\sigma^{+}] \frac{\sigma^{+}}{\operatorname{tr}[\sigma^{+}]} - \kappa \operatorname{tr}[\sigma^{-}] \frac{\sigma^{-}}{\operatorname{tr}[\sigma^{-}]}$$
 (5.343)

with 1-norm of its coefficients κ .

5.2.3 Quasiprobability extent of the identity channel

In this section, we study the quasiprobability extent of the most important channel in the context of circuit knitting with time-like cuts: the identity channel. We will study this quasiprobability extent both w.r.t. EBC* and w.r.t. MPC*.

Theorem 5.75

Let A and A' be n-qubit systems. Then

$$\gamma_{\text{MPC}^*(A \to A')}(\text{id}) = 4^n. \tag{5.344}$$

Proof. Due to Corollary 5.72 and Lemma 3.15, we can restrict ourselves to showing that $\gamma_{\underline{MPC}(A \to A')}(\mathrm{id}) = 4^n$. We denote by P_n the set of n-qubit Pauli strings $P_n := \{1, X, Y, Z\}^{\otimes n}$. We enumerate the elements of P_n as $P_n = \{Q_1, Q_2, \ldots, Q_{4^n}\}$. The set P_n forms an orthonormal basis of H(A) under the Hilbert-Schmidt inner product.

We first show $\gamma_{\underline{MPC}(A\to A')}(\mathrm{id}) \leq 4^n$ by finding an explicit QPD that fulfills the desired upper bound. We have

$$id(\rho) = \rho \tag{5.345}$$

$$= \sum_{Q \in P_n} \langle Q, \rho \rangle Q \tag{5.346}$$

$$= \sum_{Q \in P_n} \operatorname{tr}[Q\rho] \frac{Q}{2^n} \,. \tag{5.347}$$

Since for each $Q \in P_n$ one has $\left\| \frac{Q}{2^n} \right\|_1 = 1$ and $\|Q\|_{\infty} = 1$, the maps

$$\rho \mapsto \operatorname{tr}[Q\rho] \frac{Q}{2^n} \tag{5.348}$$

are elements of $MPC(A \to A')$. As such, we have found a valid QPD with 1-norm $|P_n| = 4^n$.

It remains to show the optimality of this decomposition. For this purpose, consider an arbitrary decomposition id $= \sum_i a_i \mathcal{F}_i$ where $\mathcal{F}_i \in \underline{MPC}(A \to A')$. Taking the trace on both sides, we get

$$4^{n} = \operatorname{tr}[\operatorname{id}] = \operatorname{tr}[\sum_{i} a_{i} \mathcal{F}_{i}] = \sum_{i} a_{i} \operatorname{tr}[\mathcal{F}_{i}]$$
 (5.349)

The operation \mathcal{F}_i is of the form $\rho \mapsto \operatorname{tr}[O_i \rho] \tau_i$ with $\|O_i\|_{\infty} \leq 1$ and $\|\tau_i\|_1 \leq 1$. The Pauli transfer matrix representation (i.e. the representation in the orthonormal basis of Pauli strings) $M^{(i)}$ of \mathcal{F}_i is given by

$$M_{\alpha,\beta}^{(i)} = \langle Q_{\alpha}, \mathcal{F}_i(Q_{\beta}) \rangle \tag{5.350}$$

$$= \frac{1}{2^n} \operatorname{tr}[Q_{\alpha} \operatorname{tr}[O_i Q_{\beta}] \tau_i]$$
 (5.351)

$$= \frac{1}{2^n} \operatorname{tr}[Q_{\alpha} \tau_i] \operatorname{tr}[O_i Q_{\beta}]. \tag{5.352}$$

Since $\operatorname{tr}[\mathcal{F}_i] = \operatorname{tr}[M^{(i)}]$, we get

$$4^n = \sum_i a_i \operatorname{tr}[\mathcal{F}_i] \tag{5.353}$$

$$= \sum_{i,\alpha} \frac{a_i}{2^n} \operatorname{tr}[Q_\alpha \tau_i] \operatorname{tr}[O_i Q_\alpha]$$
 (5.354)

$$= \sum_{i,\alpha} \frac{a_i}{2^n} \operatorname{tr}[(Q_\alpha \otimes Q_\alpha)(O_i \otimes \tau_i)]. \tag{5.355}$$

Let us next introduce the SWAP operator SWAP := $\frac{1}{2^n} \sum_{Q \in P_n} Q \otimes Q$. A well-known technique called the *SWAP trick* states that for any two operators A_1 , A_2 , we can write $\text{tr}[A_1A_2] = \text{tr}[(A_1 \otimes A_2) \text{SWAP}]$ [4], [35]. As such, we obtain

$$4^n = \sum_i a_i \operatorname{tr}[O_i \tau_i] \tag{5.356}$$

$$\leq \sum_{i} |a_i| |\operatorname{tr}[O_i \tau_i]| \tag{5.357}$$

$$\leq \sum_{i} |a_{i}| \|O_{i}\tau_{i}\|_{1} \tag{5.358}$$

$$\leq \sum_{i} |a_{i}| \|O_{i}\|_{\infty} \|\tau_{i}\|_{1} \tag{5.359}$$

$$\leq \sum_{i} |a_i| \tag{5.360}$$

where we used Hölder's inequality for Schatten norms. As such, no QPD can have a 1-norm smaller than 4^n .

Due to its practical importance, we briefly take the time to explicitly write down the achieving QPD in the n = 1 case.

Example 5.76. Let A, A' be single-qubit systems. By Theorem 5.75, we know that $\gamma_{\text{MPC}^*(A \to A')}(\text{id}) = 4$ which is achieved by the QPD id $= \sum_i a_i \mathcal{F}_i$ where

$$\mathcal{F}_i(\rho) = \operatorname{tr}[O_i \rho] \sigma_i \tag{5.361}$$

where the a_i , O_i and σ_i are given in Table 5.1. This decomposition can be attributed to Peng *et al.* [123].

Next, we consider the situation where classical communication is allowed for.

Theorem 5.77: [29, Proposition 4.2] and [180, Lemma 1 (supplementary materials)]

Let A, A' be d-dimensional systems. Then

$$\gamma_{\text{EBC}^*(A \to A')}(id) = 2d - 1.$$
 (5.362)

Proof. Thanks to Proposition 5.74, we can restrict our attention to γ_{EBC} instead of γ_{EBC^*} . The Choi representation of the identity channel is simply given by maximally

O_i	σ_i	$ a_i $
1	$ 0\rangle\langle 0 $	+1/2
1	$ 1\rangle\langle 1 $	+1/2
X	$ +\rangle\langle+ $	+1/2
X	$ -\rangle\langle - $	-1/2
Y	$ i\rangle\langle i $	+1/2
Y	$ -i\rangle\langle -i $	-1/2
Z	$ 0\rangle\langle 0 $	+1/2
\overline{Z}	$ 1\rangle\langle 1 $	-1/2

Table 5.1: Operators, states and weights required for a single time-like cut in a circuit.

entangled state $|\Psi\rangle_{AA'}$. We know from Theorem 5.13 that the optimal QPD of $|\Psi\rangle\langle\Psi|$ w.r.t. SEP(A; A') has 1-norm 2d-1 and is given by $|\Psi\rangle\langle\Psi| = d\sigma^+ - (d-1)\sigma^-$ where

$$\sigma^{-} = \frac{1}{d(d-1)} \sum_{i \neq j} |i\rangle\langle i| \otimes |j\rangle\langle j| \in SEP(A; A')$$
 (5.363)

and

$$\sigma^{+} = \frac{1}{d} \left(|\Psi\rangle\langle\Psi| + (d-1)\sigma^{-} \right) \in SEP(A; A'). \tag{5.364}$$

Fortunately, these two states happen to also fulfill the partial trace constraints

$$\operatorname{tr}_{A'}[\sigma^{-}] = \frac{1}{d(d-1)} \sum_{i \neq j} |i\rangle\langle i| = \frac{1}{d} \sum_{i} |i\rangle\langle i| = \frac{\mathbb{1}_{A}}{d}$$
 (5.365)

and

$$\operatorname{tr}_{A'}[\sigma^+] = \frac{1}{d} \left(\operatorname{tr}_{A'}[|\Psi\rangle\langle\Psi|] + (d-1)\operatorname{tr}_{A'}[\sigma^-] \right) = \frac{\mathbb{1}_A}{d}.$$
 (5.366)

Because of Proposition 5.71, σ^{\pm} are the Choi representations of two channels in EBC($A \to A'$), hence giving us a valid QPD of the identity channel. Clearly, $\gamma_{\text{EBC}}(\text{id})$ cannot be any lower than 2d-1, as that would contradict $\gamma_{\text{SEP}}(|\Psi\rangle\langle\Psi|) = 2d-1$. \square

The above proof hints at a deeper connection between time-like cuts of the identity channel and space-like cuts of maximally entangled states $|\Psi\rangle$. While the proof is quite succinct, it gives little insight into this connection. We explore this by providing a second argument that takes a much more operational approach to proving Theorem 5.77. For this, we will restrict ourselves to A and A' being n-qubit systems.

The argument for $\gamma_{\text{EBC}^*}(\text{id}) \leq \gamma_{\text{SEP}}(|\Psi\rangle\langle\Psi|)$ in the case of n=1 is graphically summarized in Figure 5.16. The first equality depicts the famous state teleportation protocol [11], which allows for the transfer of a qubit (i.e. the realization of the identity channel) from one party to another using a pre-shared Bell pair $|\Psi\rangle$ and one round of classical communication. By taking taking a QPD of $|\Psi\rangle\langle\Psi|$ into product states $|\Psi\rangle\langle\Psi| = \sum_i a_i \rho_i \otimes \sigma_i$, we can write the protocol as a mixture of entanglement breaking channels, which lie in EBC* $(A \to A')$. The quasiprobability extent $\gamma_{\text{SEP}}(|\Psi\rangle\langle\Psi|)$ precisely characterizes the lowest achievable 1-norm $\sum_i |a_i|$ through this means.

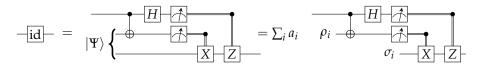


Figure 5.16: The state teleportation protocol (center) allows the transmission of one qubit through a round of classical communication by consuming one Bell pair $|\Psi\rangle$ of pre-shared entanglement. By using a QPD of the Bell state into product states $|\Psi\rangle\langle\Psi| = \sum_i a_i \rho_i \otimes \sigma_i$, we can write this protocol as a quasiprobabilistic mixture of entanglement breaking channels.

On the converse side, a Bell-pair shared between two parties can also be prepared using one instance of an identity channel from one side to the other. As such, any QPD of the identity channel into EBC* channels automatically induces a QPD of $|\Psi\rangle\langle\Psi|$ into separable states.

The above arguments for the n=1 case straightforwardly generalize to arbitrary $n \geq 1$ by using n parallel instances of gate teleportation, or identity channels respectively. Hence, $\gamma_{\text{EBC}^*}(\mathrm{id}^{\otimes n}) = \gamma_{\text{SEP}}(|\Psi\rangle\langle\Psi|^{\otimes n}) = 2^{n+1} - 1$. The equivalence between nonlocal identity channels and Bell pairs that we have found is strongly reminiscent of the equivalence of Clifford gates and their Choi states that we previously explored in Section 5.1.4.1.

We also note that the sub-multiplicativity of $\gamma_{\rm EBC^*}({\rm id})$ can be practically harnessed for time-like cuts that do not occur in the same time slice of the circuit. We refrain from an in-depth discussion of this approach, since it exactly mirrors the technique utilized for black box cuts of Clifford gates in Section 5.1.5. The main idea is to realize all space-like cuts using the state teleportation protocol, and to jointly simulate the preparation of all Bell pairs to harness the sub-multiplicative behavior of $\gamma_{\rm SEP}$. To avoid having to use a large number of ancilla qubits, an *entanglement factory* can be utilized, which re-uses ancillas for Bell pair simulation. In fact, the same entanglement factory can be shared between space-like and time-like cuts.

In a black box cut setting, it seems unlikely that one could entirely avoid such ancilla qubits. However, in the parallel cut setting, there exist constructions that do not require any ancilla qubits in their implementation, which allows for a significantly simpler practical realization. We refer the reader to Ref. [73], [122] for details on these constructions.

As a somewhat surprising conclusion, this chapter demonstrated that classical communication can significantly reduce the overhead of circuit knitting with timelike cuts. Indeed, to cut one single wire in a quantum circuit, the associated extent is $\gamma_{\text{MPC}^{\star}}(\text{id}) = 4$ and $\gamma_{\text{EBC}^{\star}}(\text{id}) = 3$ respectively. When we cut many wires in parallel, the gap widens even more, as the regularized gamma factors are $\overline{\gamma}_{\text{MPC}^{\star}}(\text{id}) = 4$ and

$$\overline{\gamma}_{EBC^*}(id) = \lim_{n \to \infty} (2^{n+1} - 1)^{1/n} = 2.$$
 (5.367)

This is especially remarkable, considering that for space-like cuts, we were unable to find an advantage provided by classical communication for a large variety of unitary gates.

5.2.4 Going beyond the identity channel

In this section, we briefly study the quasiprobability extents $\gamma_{\mathrm{MPC}^\star}(\mathcal{E})$ and $\gamma_{\mathrm{EBC}^\star}(\mathcal{E})$ for channels $\mathcal{E} \in \mathrm{CPTP}(A \to A')$ beyond the identity. A first insight is that the identity channel is in some sense the maximally difficult channel to cut.

Corollary 5.78: Let $\mathcal{E} \in \operatorname{CPTP}(A \to A')$ and either $S = \operatorname{MPC}^*$ or $S = \operatorname{EBC}^*$. Then

$$\gamma_{S(A \to A')}(\mathcal{E}) \le \min\{\gamma_{S(A,A)}(id_A), \gamma_{S(A',A')}(id_{A'})\}.$$
 (5.368)

This follows from a relatively simple insight, namely that a cut of $\mathcal E$ can be straightforwardly realized by first cutting the identity channel and then running $\mathcal E$ on the receiver's system, or analogously, running $\mathcal E$ on the sender's system and then cutting the identity. This insight is more generally captured by a stronger version of the chaining property.

Lemma 5.79: Let $\mathcal{E} \in \operatorname{CPTP}(A \to A')$, $\mathcal{F} \in \operatorname{CPTP}(A' \to A'')$ and either $S = \operatorname{MPC}^*$ or $S = \operatorname{EBC}^*$. Then

$$\gamma_{S(A \to A'')}(\mathcal{F} \circ \mathcal{E}) \le \min\{\gamma_{S(A \to A')}(\mathcal{E}), \gamma_{S(A', A'')}(\mathcal{F})\}. \tag{5.369}$$

Proof. It suffices to show that any QPD of \mathcal{E} w.r.t. $S(A \to A')$ and any QPD of \mathcal{F} w.r.t. $S(A' \to A'')$ both can induce a QPD of $\mathcal{F} \circ \mathcal{E}$ w.r.t. $S(A \to A'')$ with the same 1-norm of coefficients. This follows immediately from the fact that if you prepend or append a CPTP map \mathcal{G} to any map in $S(A \to A')$, one again obtains an element in $S(A \to A')$. For $S = EBC^*$ this is quite evident as

$$\mathcal{G}\left(\sum_{i}\beta_{i}\operatorname{tr}[M_{i}\rho]\sigma_{i}\right) = \sum_{i}\beta_{i}\operatorname{tr}[M_{i}\rho]\mathcal{G}(\sigma_{i})$$
(5.370)

and

$$\sum_{i} \beta_{i} \operatorname{tr}[M_{i} \mathcal{G}(\rho)] \sigma_{i} = \sum_{i} \beta_{i} \operatorname{tr}[\mathcal{G}^{\dagger}(M_{i})\rho] \sigma_{i}$$
 (5.371)

where \mathcal{G}^{\dagger} is the adjoint channel of \mathcal{G} and $\mathcal{G}^{\dagger}(M_i) \geq 0$, $\sum_i \mathcal{G}(M_i) = 1$ since \mathcal{G} is completely positive and unital (see e.g. [175, Proposition 2.18, Theorem 2.26]. An analogous argument applies for $S = \text{MPC}^{\star}$.

Another consequence of Lemma 5.79 is that unitaries are exactly as difficult to cut as the identity channel.

Corollary 5.80: Let $\mathcal{U} \in \operatorname{CPTP}(A)$ be a unitary channel and either $S = \operatorname{MPC}^*$ or $S = \operatorname{EBC}^*$. Then

$$\gamma_S(\mathcal{U}) = \gamma_S(\mathrm{id}).$$
 (5.372)

Proof. By Corollary 5.78, we know that $\gamma_S(\mathcal{U}) \leq \gamma_S(\mathrm{id})$. For the converse, we use Lemma 5.79

$$\gamma_S(\mathrm{id}) = \gamma_S(\mathcal{U}\mathcal{U}^{-1}) \le \gamma_S(\mathcal{U}).$$
 (5.373)

So if we want to find some interesting physics beyond the identity channel, we necessarily need to consider non-unitary channels. As for space-like cuts, non-unitarity significantly increases the complexity of evaluating the quasiprobability extent, so in the following we content ourselves with summarizing various techniques for finding upper and lower bounds.

As a first result, we note that the proof technique of Theorem 5.75 can be straightforwardly generalized to arbitrary channels, leading to following lower and upper bounds to γ_{MPC^*} .

Proposition 5.81

Let *A* and *A'* be *n*-qubit systems and $\mathcal{E} \in HP(A \to A')$. Then

$$\operatorname{tr}[M] \le \gamma_{\operatorname{MPC}^{*}(A \to A')}(\mathcal{E}) \le \sum_{Q \in P_{n}} \left\| \mathcal{E}^{\dagger}(Q) \right\|_{\infty}$$
 (5.374)

where M is the Pauli transfer matrix representation of \mathcal{E} , P_n denotes the set of n-qubit Pauli strings and \mathcal{E}^{\dagger} is the adjoint map of \mathcal{E} (in the sense that $\langle \mathcal{E}(X), Y \rangle = \langle X, \mathcal{E}^{\dagger}(Y) \rangle$).

Proof. The lower bound is achieved by the exact same steps as in the proof of Theorem 5.75. For the upper bound, we use

$$\mathcal{E}(\rho) = \sum_{Q \in P_n} \langle Q, \mathcal{E}(\rho) \rangle Q \tag{5.375}$$

$$= \sum_{Q \in P_n} \langle \mathcal{E}^{\dagger}(Q), \rho \rangle Q \tag{5.376}$$

$$= \sum_{Q \in P_n} \left\| \mathcal{E}^{\dagger}(Q) \right\|_{\infty} \operatorname{tr} \left[\frac{\mathcal{E}^{\dagger}(Q)}{\| \mathcal{E}^{\dagger}(Q) \|_{\infty}} \rho \right] \frac{Q}{2^n}$$
 (5.377)

which provides us with a QPD of \mathcal{E} w.r.t. the decomposition set $\underline{\mathsf{MPC}}(A \to A')$ with 1-norm $\sum_{Q \in P_n} \|\mathcal{E}^{\dagger}(Q)\|_{\infty}$.

Example 5.82. Consider the 1-qubit depolarizing channel

$$\mathcal{D}_{p}(\rho) = (1-p)\rho + \text{tr}[\rho]\frac{1}{2}\mathbb{1}.$$
 (5.378)

It is easy to verify that it is self-adjoint $\mathcal{D}_p^{\dagger} = \mathcal{D}_p$ and that its Pauli transfer matrix representation is M = diag(1, 1 - p, 1 - p, 1 - p) since

$$\mathcal{D}_{p}(1) = 1$$
, $\mathcal{D}_{p}(X) = (1 - p)X$, (5.379)

$$\mathcal{D}_p(Y) = (1-p)Y,$$
 $\mathcal{D}_p(Z) = (1-p)Z.$ (5.380)

As such, we get matching lower and upper bounds from Proposition 5.81 and

$$\gamma_{\text{MPC}^{\star}}(\mathcal{D}_p) = 4 - 3p. \tag{5.381}$$

Δ

Next, we focus purely on the quasiprobability extent w.r.t. EBC, which is the same as the one w.r.t. EBC* as shown in Proposition 5.74. Through the Choi characterization of Proposition 5.71, $\gamma_{\text{EBC}(A \to A')}(\mathcal{E})$ is almost the same quantity as $\gamma_{\text{SEP}(A;A')}(J_{\mathcal{E}})$, except that we impose an additional partial trace constraint on the decomposition set. As such, many techniques utilized for space-like cutting can be directly applied to bound γ_{EBC} . In fact, we can even consider EBC($A \to B'$) to be a special case of the decomposition set SEPC($A; B \to A'; B'$) where A' and B are chosen to be of dimension 1. Therefore, the techniques from Section 5.1.4 to estimate γ_{SEPC} are directly applicable here. We briefly summarize them below.

State conversion bounds For any bipartite state $\rho_{AB} \in D(AB)$ and any $\mathcal{E} \in CPTP(A \to A')$ one has

$$\gamma_{\text{EBC}^{\star}}(\mathcal{E}) \ge \gamma_{\text{SEP}}(\mathcal{E} \otimes \text{id}_{B}(\rho)).$$
(5.382)

This can be verified by checking that any QPD of \mathcal{E} induces one of $\mathcal{E} \otimes \mathrm{id}_B(\rho)$ with equal 1-norm. In the special case where ρ_{AB} is chosen as the maximally entangled state, we get

$$\gamma_{\text{EBC}^*}(\mathcal{E}) \ge \gamma_{\text{SEP}}(J_{\mathcal{E}}).$$
(5.383)

PPT relaxation bound By using a PPT relaxation on the separability constraint in the Choi characterization of EBC, we obtain a lower bound for the quasiprobability extent of $\mathcal{E} \in \text{CPTP}(A \to A')$

$$\gamma_{\text{EBC}^{\star}(A \to A')}(\mathcal{E}) \geq \begin{cases}
\min_{\Lambda^{\pm} \in P(AA')} & \text{tr}[\Lambda^{+}] + \text{tr}[\Lambda^{-}] \\
\text{s.t.} & J_{\mathcal{E}} = \Lambda^{+} - \Lambda^{-}, \\
& \text{tr}_{A'}[\Lambda^{\pm}] = \text{tr}[\Lambda^{\pm}] \frac{1}{\dim(A)} \mathbb{1}_{A}, \\
& (\text{id}_{A} \otimes T_{A'})(\Lambda^{\pm}) \geq 0
\end{cases} (5.384)$$

which is an SDP that can be efficiently evaluated.

SDP hierarchy As a refinement of the above PPT relaxation, we can lower bound $\gamma_{\text{EBC}(A \to A')}(\mathcal{E})$ through the SDP $\gamma_{\text{EBC}_k(A \to A')}(\mathcal{E})$ where

$$EBC_k(A \to A') := \{ \mathcal{E} \in CPTP(A \to A') | J_{\mathcal{E}} \in SEP_k(A; A') \}$$
 (5.385)

is a converging outer approximation $EBC(A \to A') = \bigcap_{k=1}^{\infty} EBC_k(A \to A')$.

Asymptotic characterization Let us define the set of entanglement breaking-preserving superchannels EBPSC($A, A' \rightarrow B, B'$) to be the superchannels which map elements from EBC($A \rightarrow A'$) to EBC($B \rightarrow B'$). The single-shot dilution cost $E_{C, \text{EBPSC}}^{(1)\varepsilon}(\rho)$ is characterized by the smooth log-robustness

$$\log(1 + R_{\text{EBC}}^{\epsilon}(\mathcal{E})) \le E_{\text{C,EBPSC}}^{(1),\epsilon}(\mathcal{E}) \le \log(1 + R_{\text{EBC}}^{\epsilon}(\mathcal{E})) + 2. \tag{5.386}$$

This was independently proven in Ref. [180, Theorem 1 (supplementary materials)], though it can also be seen as a direct consequence of Lemma 5.55. As such, the regularized and asymptotic quasiprobability extent w.r.t. EBC* can be expressed as

$$\log \overline{\gamma}_{EBC^*}(\mathcal{E}) = E_{C,EBPSC}^{\text{exact}}(\mathcal{E})$$
 (5.387)

$$\log \gamma_{\text{EBC}^*}^{\infty}(\mathcal{E}) = E_{C,\text{EBPSC}}(\mathcal{E}). \tag{5.388}$$

5.3 Further reading

The idea of extending the size of a quantum computer was first proposed by Bravyi et al. [25], specifically in the context of sparse quantum circuits and Pauli-based computation. The idea is to simulate an n+k-qubit quantum circuit with an n-qubit quantum computer by using a hybrid classical-quantum scheme that simulates k "virtual qubits". The overhead of the technique is exponential in k, which is generally less favorable compared to circuit knitting.

This shortcoming was addressed in the first circuit knitting scheme by Peng *et al.* [123]. Their approach uses the time-like QPD of the identity channel in Example 5.76 to decompose the total circuit into sub-circuits that fit on a small quantum device, though their approach is phrased in the language of tensor networks. The individual circuit clusters are either combined through QPS or via a tensor network contraction.

The first circuit knitting scheme with space-like cuts was introduced by Mitarai and Fujii in Ref. [111], [112]. The authors propose explicit QPDs of arbitrary two-qubit gates in terms of a decomposition set that contains single-qubit rotations and projective measurements (including classical side information derived from those). This decomposition set is a strict subset of LO*, and the authors explicitly define the quasiprobability extent of a channel w.r.t. that set (which they call *channel robustness of nonlocality*), however without being able to evaluate it.

In a later paper by Piveteau and Sutter [129], the decomposition sets LO* and LOCC* were introduced for the first time, though their definition of LO* did not encompass the full generality of classical side information (an inconsequential technicality which we discussed around Equation (5.6)). The QPD from [112] was also proven to be optimal w.r.t. LO* and LOCC* for two-qubit gates with at most two non-zero KAK parameters. In that work, the connection to the robustness of entanglement was first established. The authors also considered for the first time the possibility to utilize real-time classical communication between the two parties. They proposed teleportation-based circuit knitting in the black box setting and proved the optimal quasiprobability extent of Clifford gates w.r.t. LOCC*.

Two subsequent papers, one by Schmitt, Piveteau and Sutter [140] and the other by Harrow and Lowe [74], simultaneously and independently established the quasiprobability extent for unitaries with unitary operator Schmidt decompositions (and as such arbitrary two-qubit gates). The former paper additionally extended the black-box results to arbitrary two-qubit unitaries. The latter instead applied the techniques to clustered Hamiltonian simulation and also provided an improved wire cut-based circuit kniting scheme that can further reduce the sampling overhead given a certain sparsity assumption on the observable. A third paper by Ufrecht *et al.* [159] also appeared simultaneously which describes an optimal sub-multiplicative black box cutting method for the more restricted set of two-qubit rotation gates.

The PPTC-based lower bound for the quasiprobability extent γ_{LOCC} (without intermediate measurements) of channels was first proposed by Jing *et al.* [89]. They also established the lower bounds for the regularized quasiprobability extent $\overline{\gamma}_{PPTC}$ and $\overline{\gamma}_{SEPC}$ in terms of the entanglement cost for the first time.

We briefly note that the quasiprobability extent $\gamma_{SEPC}(\mathcal{U})$ of unitary channels \mathcal{U} with unitary operator Schmidt decompositions was introduced and characterized long before the advent of circuit knitting by Harrow and Nielsen in Ref. [76].

Focusing on wire cutting, the quasiprobability extent γ_{EBC} (without classical side information) of the identity channel was first characterized in Ref. [180]. This work focuses more on resource theoretic aspects instead of circuit knitting and the authors were seemingly not aware of the previous work by Peng *et al.* [123]. Similarly, many subsequent circuit knitting papers were not aware of [180] and did not refer to it.

Lowe *et al.* [106] proposed the first circuit knitting scheme based on time-like cuts that utilizes classical communication and exhibits strictly sub-multiplicative scaling. While their approach does achieve the optimal asymptotic scaling, its finite-size extent is strictly sub-optimal.

Later work by Brenner, Piveteau and Sutter [29] established the notions of EBC* and MPC and they provided and exact characterization of $\gamma_{EBC^*}(id)$ and $\gamma_{MPC}(id)$ with explicit QPDs achieving the two extents. They also introduced teleportation-based time-like cutting which allowed for its application in the black-box setting.

The optimal EBC* QPD in Ref. [29] requires the utilization of ancilla qubits for its practical realization. Two subsequent works by Pednault [122] and by Harada, Wada and Yamamoto[73] provide alternative optimal QPDs with explicit circuit implementations that do not require these ancilla qubits.

To date, multiple experimental demonstrations of QPS-based circuit knitting have been realized [36], [148], [179].

5.4 Contributions

This chapter almost exclusively contains novel contributions by the author. Some of them were previously presented in Ref. [29], [129], [140] while many others are new and previously unpublished. We briefly summarize the contributions section-by-section.

Section 5.1.1 The formal definition of the quasiprobability extent for circuit cutting

was introduced by the author in Ref. [129] for LO and LOCC. In this section, we establish connections with the novel formalism from Chapter 3 and consider the extension to the relaxations SEPC and PPTC.

- Section 5.1.2 This section only contains previously unpublished results.
- **Section 5.1.3** The characterization of γ_{SEP} for pure states is originally due to Vidal and Tarrach [166]. Though we note that the convex programming-based proof presented in Section 5.1.3.1 is new, arguably simpler and as a side result also provides a characterization of γ_{PPT} for pure states.
 - All results in Sections 5.1.3.2 and 5.1.3.3 are new and previously unpublished.
- **Section 5.1.4** The Sections 5.1.4.1 and 5.1.4.2 summarize the authors' results published in Ref. [129], [140]. During the preparation of the second publication, we became aware of efforts by Lowe and Harrow who had independently discovered a proof of Theorem 5.44 [74]. While their work was published later than ours, we acknowledge an equal claim to the discovery of the result.

The work in Sections 5.1.4.3 and 5.1.4.5 to 5.1.4.7 is new and previously unpublished.

We note that the idea of using the PPTC relaxation for lower bounds, as presented in Section 5.1.4.4, is due to Jing, Zhu and Wang [89]. Furthermore, the characterization of the regularized quasiprobability extent in Section 5.1.4.7 can be seen as a refinement of the work in Ref. [89], where the authors showed that $\overline{\gamma}_{LOCC}$ can be lower bounded by an entanglement cost.

- Section 5.1.5 This section summarizes results by the author from Ref. [129], [140].
- **Section 5.1.6** To our knowledge, the analysis of the quantum Fourier transform is new and previously unpublished. The results on clustered Hamiltonian simulation are due to Harrow and Lowe [74].
- **Section 5.2.1** The decomposition sets <u>MPC</u> and EBC* were initially introduced by the author in Ref. [29]. This section extends this work by formally introducing MPC* and determining the Choi characterization of the various decomposition sets.
- Section 5.2.2 This section only contains previously unpublished results.
- **Section 5.2.3** The two main results that characterize $\gamma_{\text{MPC}^*}(\text{id})$ and $\gamma_{\text{EBC}^*}(\text{id})$ (Theorems 5.75 and 5.77) were both introduced by the author in Ref. [29]. After publishing this paper, we were made aware of a previous work that had independently obtained a characterization of $\gamma_{\text{EBC}}(\text{id})$, though with a different proof technique [180, Lemma 1 (supplementary materials)]. Together with our insight from Proposition 5.74, this also leads to a characterization of $\gamma_{\text{EBC}^*}(\text{id})$.
- **Section 5.2.4** This section only contains previously unpublished results, except for the PPT lower bound, which was proposed in Ref. [180].

Chapter 6

Simulating magic computation

In this chapter, we discuss the simulation of non-Clifford circuits using a Clifford quantum computer. This application of QPS was already briefly covered as a motivational example in Section 3.1 and the discussion here provides a more in-depth treatment.

The famous Gottesman-Knill theorem states that any quantum circuit can be efficiently simulated classically if it only consists of Clifford gates as well as preparations and measurements of qubits in the computational basis [2], [63]. Clifford gates also play an important role in quantum error correction, as the naturally fault-tolerant (transversal) gates often form a subgroup of the Clifford group. The Clifford gates are not universal, and in order to reach universality, they need to be complemented by states or gates which are in some sense "non-classical". The most common such choice is the T gate, or equivalently the magic state $|H\rangle := \left(|0\rangle + e^{i\pi/4} \, |1\rangle\right)/\sqrt{2}$ which can realize a T gate using the injection circuit in Figure 3.2.

The resource theory of magic aims to quantify the amount of non-classicality (called "magic") that non-Clifford channels or non-stabilizer states contain. This chapter will formally introduce this resource theory and study the induced quasiprobability extent. We will restrict ourselves to qubits (as opposed to higher-dimensional qudits) since that represents the computationally most relevant setting. Intriguingly, the set of free states is finite (respectively convex with finitely many extremal points). At first, this might suggest that this resource theory could be simpler to work with compared to others like the QRT of entanglement. Alas, the opposite is the case: While it is indeed possible to numerically evaluate the quasiprobability extent of any state for small systems through the linear program in Section 3.6, it is very hard to find analytic closed-form expressions, and we know very little about the asymptotic behavior of the quasiprobability extent, even for the simplest cases. Very little is also known about the utility of classical side information through intermediate measurements in this theory - this topic has not previously been studied to our knowledge.

The obvious practical application of QPS w.r.t. the QRT of magic is classical simulation of non-Clifford circuits. However, it is important to note that there exist other approaches that can simulate near-Clifford circuits more efficiently. The state-of-theart approach is instead based on the stabilizer rank simulation algorithm proposed in Ref. [21], [22] and generally seems to exhibit roughly a square root speedup over

the QPS approach. The main difference of the stabilizer rank simulation approach is that it is based on quasiprobability decompositions of magic states and gates in the pure state quantum mechanics formalism instead of the mixed state formalism. That is, one decomposes non-Clifford unitaries (or non-stabilizer states) into Clifford unitaries (or stabilizer states) instead of working with channels (or density matrices). The runtime of the algorithm is then related to the number of elements in this decomposition, which interestingly can also be related to the 1-norm of the coefficients [21, Theorem 1]. However, due to the nature of the algorithm, stabilizer-rank simulators cannot incorporate mixed states and non-unitary channels the same way as QPS does, at least not without modifications.

Another simulation algorithm that outperforms QPS w.r.t. the QRT of magic is the *dyadic frame simulator* [142]. This technique can actually also be considered an instance of QPS itself, but with an extended decomposition set.

In summary, there exist better simulation algorithms for the task of simulating near-Clifford circuits with a classical computer. However, these improved techniques explicitly rely on using a *classical* computer that allows for strong simulation. We framed QPS as a method for simulating quantum computers using a *quantum* computer itself (a Clifford quantum computer in this case), so these improved techniques are not on the same footing. Still, the practical implications of this chapter are less clear compared to the other ones. For this reason, we will go into less detail and content ourselves with sketching some basic properties of the decomposition sets and the quasiprobability extent.

6.1 Decomposition sets

The core idea behind the Gottesman-Knill theorem is that any intermediate state of a Clifford circuit necessarily belongs to the class of *stabilizer states*, a finite subset of all n-qubit quantum states which can be represented with only poly(n) memory. Conversely, these states precisely represent the set of reachable states from a computational basis state using a Clifford unitary.

Definition 6.1: Let $S \subset P_n$ be an Abelian subgroup of the Pauli group which does not contain $-(\mathbb{1}^{\otimes n})$. The *n*-qubit *stabilizer code* C(S) associated to S is defined to be the simultaneous +1 eigenspace of the operators in S.

It can easily be shown that the dimension of a stabilizer code is precisely $\dim(\mathcal{C}(S)) = 2^k$ where n - k is the minimal number of generators of S. We say that $\mathcal{C}(S)$ is a [n,k] stabilizer code. A [n,0] stabilizer code only represents a single state (up to a complex scaling factor). States that can be represented as a [n,0] stabilizer code are called *stabilizer states*, and we denote the set of stabilizer states of an n-qubit system A by STAB(A).

Having defined the free states of our QRT, we next turn to defining the free channels. There are at least two reasonable choices that can be made here. The first choice is very operational in nature, as it precisely entails the set of operations that can be efficiently simulated through the Gottesman-Knill theorem.

Definition 6.2: Let A and A' be multi-qubit systems. We say that a quantum channel from A to A' is a *stabilizer operation* if it is a composition of

- preparations of stabilizer states,
- Clifford gates,
- Pauli measurements,
- discarding of qubits,

which may include classical control (based on previous measurement outcomes) that can involve randomness. We denote the set of such stabilizer operations by $SO(A \rightarrow A')$.

Note that we do not impose any sort of computational restrictions on the classical control logic - it could in principle entail very long and expensive classical calculations. This is a simplifying assumption akin to the one we made for circuit knitting with classical communication, where we allowed LOCC protocols with an unbounded number of rounds.

Another possible choice for the free operations is the set of completely resource non-generating channels.

Definition 6.3: Let A and A' be multi-qubit systems. The set of completely stabilizer-preserving channels $CSP(A \to A') \subset CPTP(A \to A')$ is defined to contain all channels \mathcal{E} such that for any ancillary system E

$$\forall \rho \in \operatorname{STAB}(AE) : (\mathcal{E} \otimes \operatorname{id}_E)(\rho) \in \operatorname{STAB}(A'E). \tag{6.1}$$

While $CSP(A \to A')$ is a more axiomatic choice of free operations, the operationally meaningful decomposition set for the purposes of QPS is clearly given by $SO(A \to A')$. The relation between $SO(A \to A')$ and $CSP(A \to A')$ was recently studied in detail by Heimendahl *et al.* [78]. Interestingly, they observed that $SO(A \to A')$ is a strict subset, i.e., there exist completely stabilizer-preserving operations which lie outside of $SO(A \to A')$. It can still be useful to consider $CSP(A \to A')$ as a well-behaved relaxation of $SO(A \to A')$, since it exhibits a rather simple characterization in terms of the Choi representation.

Proposition 6.4: [141, Theorem 3.1]

Let *A* and *A'* be multi-qubit systems and $\mathcal{E} \in T(A \to A')$. Then

$$\mathcal{E} \in \mathrm{CSP}(A \to A') \iff J_{\mathcal{E}} \in \mathrm{conv}(\mathrm{STAB}(AA')) \text{ and } \mathrm{tr}_{A'}[J_{\mathcal{E}}] = \frac{1}{\dim(A)} \mathbb{1}_A.$$
 (6.2)

Similarly, there exists a "normal form" of stabilizer operations, though it is not quite as simple.

Proposition 6.5: [78, Theorem 4]

Let *A* and *A'* be *n*-qubit and *m*-qubit systems. Every $\mathcal{E} \in SO(A \to A')$ can be written as a convex mixture of stabilizer operations of the type

$$\mathcal{E}(\rho) = \operatorname{tr}_{m+1,\dots,n+r} \left[\sum_{i} U_{i} \left(\Pi_{i} \rho \Pi_{i}^{\dagger} \otimes |0\rangle\langle 0|^{\otimes r} \right) U_{i}^{\dagger} \right]$$
(6.3)

for some $r \ge 0$, n + r-qubit Clifford unitaries U_i and a projective measurement $\{\Pi_i\}_i$ of mutually orthogonal stabilizer code projectors Π_i .

In fact, the proof of Proposition 6.5 in Ref. [78] implies even a slightly stronger statement. The free quantum instrument $\mathfrak{I}_{SO}(A \to A')$ from A to A' (recall Definition 2.13) are precisely given by coarse-grainings of convex mixtures of instruments $(\mathcal{E}_i)_i \in QI(A \to A')$ of the form

$$\mathcal{E}_{i}(\rho) = \operatorname{tr}_{m+1,\dots,n+r} \left[U_{i} \left(\Pi_{i} \rho \Pi_{i}^{\dagger} \otimes |0\rangle\langle 0|^{\otimes r} \right) U_{i}^{\dagger} \right]$$
(6.4)

for some $r \ge 0$, n + r-qubit Clifford unitaries U_i and a projective measurement $\{\Pi_i\}_i$ of mutually orthogonal stabilizer code projectors Π_i . Therefore, we can write the decomposition set $SO^*(A \to A') := \text{chan}^*[\mathfrak{I}_{SO}(A \to A')]$, which captures the notion of intermediate measurements, as the absolute convex hull of all superoperators of the form

$$\rho \mapsto \operatorname{tr}_{E} \left[\sum_{i} \beta_{i} U_{i} \left(\Pi_{i} \rho \Pi_{i}^{\dagger} \otimes |0\rangle\langle 0|^{\otimes r} \right) U_{i}^{\dagger} \right]$$

$$(6.5)$$

for some $\beta_i \in [-1, 1]$.

We note that SO and CSP are convex sets and SO* is even absolutely convex (recall Lemma 3.28). As such, many properties from Section 3.4 directly apply to the associated quasiprobability extents, such as the relation to a robustness measure and the reduction to QPDs with at most two (respectively one) elements. Note also that any CPTP map can be simulated using the decomposition sets SO and CSP.

Lemma 6.6: Let A and A' be multi-qubit systems. Then

$$\operatorname{span}_{\mathbb{R}}(\operatorname{SO}(A \to A')) = \operatorname{span}_{\mathbb{R}}(\operatorname{CSP}(A \to A')) = \mathbb{R}\operatorname{HPTP}(A \to A')\,. \tag{6.6}$$

Recall that $\mathbb{R}HPTP(A \to B) := \{r \cdot \mathcal{E} | r \in \mathbb{R}, \mathcal{E} \in HPTP(A \to B)\}.$

Proof. Since $SO(A \to A') \subset CSP(A \to A') \subset HPTP(A \to A')$ it suffices to show $\mathbb{R}HPTP(A \to A') \subset \operatorname{span}_{\mathbb{R}}(SO(A \to A'))$. For this purpose, consider an arbitrary basis $\{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_M\} \subset CPTP(A \to A')$ of $HPTP(A \to A')$ (recall from Lemma 4.1 that CPTP spans HPTP). Denote by $D \neq 0$ the determinant of the matrix obtained by stacking the vectorized representations of the \mathcal{E}_i . Because of the Stinespring dilation and Solovay-Kitaev theorems and the fact that the T gate can be decomposed in terms of Clifford gates (see Equation (3.5)), all the channels \mathcal{E}_i can be approximated

arbitrarily well by linear superpositions of elements in $SO(A \rightarrow A')$. By choosing the approximation error to be small enough, we can ensure that the change in determinant is smaller than |D| such that the approximated channels still remain linearly independent (since the determinant is continuous).

Similarly, one can verify that

$$\operatorname{span}_{\mathbb{R}}(\operatorname{SO}^{\star}(A)) = \operatorname{HP}(A) \tag{6.7}$$

which immediately follows from the fact that the basis of HP(A) presented in Example 3.24 lies in $SO^*(A)$.

6.2 Quasiprobability extent of states

Due to unfortunate historical circumstances, the quasiprobability extent γ_{STAB} is commonly called the "robustness of magic" in literature. This clashes with the more commonly used notion of robustness R_{STAB} (as introduced in Definition 3.16) which relates to the extent by $\gamma_{\text{STAB}}(\rho) = 1 + 2R_{\text{STAB}}(\rho)$ for $\rho \in D(A)$ by Lemma 3.17. In this manuscript, we will stick to the nomenclature used in other resource theories and call γ_{STAB} the "quasiprobability extent" and R_{STAB} the "robustness".

Since the decomposition set is finite, we can express the quasiprobability extent $\gamma_{\text{STAB}(A)}(\rho)$ in terms of a linear program (recall Section 3.6). As such, we can numerically evaluate it for small systems with a runtime of poly(|STAB(A)|). Unfortunately, in contrast to the resource theory of entanglement, we do not know an explicit analytical expression for the quasiprobability extent of product states $\gamma_{STAB(A)}(\rho^{\otimes m})$ for arbitrary m, even for simple states like the magic state $|H\rangle$. This makes the estimation of the regularized extent $\overline{\gamma}_{STAB}$ very difficult. To make matters worse, $|STAB(A)| = 2^n \prod_{k=1}^n (2^k + 1) \ge 2^{n(n+3)/2}$ grows super-exponentially in the number of qubits n [2], [147], making the cost of numerically evaluating γ_{STAB} grow out of control extremely quickly. By exploiting the symmetry structure of the stabilizer polytope [79] as well as algorithmic improvements in calculating stabilizer state overlaps [72], state-of-the-art techniques are able to compute the quasiprobability extent of arbitrary states up to 8-qubits. By considering states with certain symmetries or by allowing for approximations, even larger systems can be handled - see [72, Table 1] for a detailed overview. We especially note an interesting approach in Ref. [79] which proposes a hierarchy of inner approximations of the stabilizer polytope by restricting the decomposition set to stabilizer states which are at most k-partite entangled for $1 \le k \le n$. This is somewhat reminiscent of the SDP hierarchy we proposed in Section 5.1.3.2, and allows for efficiently computable upper bounds to the quasiprobability extent together with explicit QPDs that achieve these upper bounds.

A characteristic that γ_{STAB} does share with its entanglement counterpart γ_{SEP} is that it exhibits strict sub-multiplicativity.

Example 6.7. The magic state $|H\rangle := (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$ has a quasiprobability extent $\gamma_{\text{STAB}}(|H\rangle\langle H|) = \sqrt{2}$ which is achieved by the QPD

$$|H\rangle\langle H| = \frac{1}{2}|+\rangle\langle +|+\frac{\sqrt{2}}{2}|i+\rangle\langle i+|-\frac{\sqrt{2}-1}{2}|-\rangle\langle -|$$
(6.8)

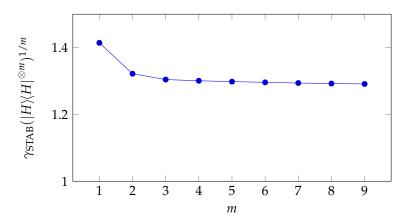


Figure 6.1: Normalized quasiprobability extent $\gamma_{\text{STAB}}(|H\rangle\langle H|^{\otimes m})^{1/m}$ of m magic states. Data taken from Ref. [79].

where $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$ and $|i\pm\rangle := (|0\rangle \pm i |1\rangle)/\sqrt{2}$. The normalized quasiprobability extent of m copies of the magic state $\gamma_{\text{STAB}}(|H\rangle\langle H|^{\otimes m})^{1/m}$ was evaluated in Ref. [79] and is depicted in Figure 6.1.

The sub-multiplicative behavior is very important as the asymptotic simulation overhead of a Clifford+T circuit with t T gates can be reduced from $\mathcal{O}(\gamma_{\text{STAB}}(|H\rangle\langle H|)^{2t})$ to $\mathcal{O}(\overline{\gamma}_{\text{STAB}}(|H\rangle\langle H|)^{2t})$ where the regularized quasiprobability extent is estimated to be around $\overline{\gamma}_{\text{STAB}}(|H\rangle\langle H|) \approx 1.283 \pm 0.002$ in Ref. [79].

6.3 Quasiprobability extent of channels

Generally, finding the quasiprobability extent $\gamma_{SO^*}(\mathcal{E})$ of a channel $\mathcal{E} \in CPTP(A \to A')$ is very difficult. In the case where there is a one-to-one correspondence between \mathcal{E} and a state, we can relate $\gamma_{SO^*}(\mathcal{E})$ to the extent of the state.

Example 6.8. A T gate and a magic state $|H\rangle$ are equivalent under stabilizer operations, as one can be transformed into the other (recall the magic state injection in Figure 3.2 and $|H\rangle = T|+\rangle$). As such, we have for all $m \in \mathbb{N}$

$$\gamma_{SO^*}(T^{\otimes m}) = \gamma_{SO}(T^{\otimes m}) = \gamma_{STAB}(|H\rangle\langle H|^{\otimes m})$$
(6.9)

by Lemmas 3.29 and 3.31
$$\triangle$$

Unfortunately, such a state-channel correspondence is generally not available. Due to the difficulty of estimating the quasiprobability extent w.r.t. SO^* , we will from now on focus on estimating $\gamma_{SO}(\mathcal{E})$. Below, we summarize a few strategies for finding lower and upper bounds, the latter through possibly non-optimal QPDs.

To find explicit QPDs for upper bounds, a general approach consists of finding inner approximations of SO and use those as decomposition sets. For instance, one could pick the finite set of all Clifford unitary channels (respectively its convex hull).

Interestingly those are sufficient for achieving an optimal QPD for the T gate as

$$Ad[T] = \frac{1}{2}id + \frac{1}{\sqrt{2}}Ad[S] - \frac{\sqrt{2}-1}{2}Ad[Z]$$
 (6.10)

achieves the 1-norm of $\sqrt{2}$. A more refined choice would be all stabilizer operators that require only a bounded number of ancillas r in Equation (6.3). This is also a finite decomposition set, and in the limit of $r \to \infty$ the quasiprobability extent will achieve γ_{SO} .

To find lower bounds, we instead turn to outer approximations of SO. A natural candidate in this regard is $CSP(A \rightarrow A')$, which is precisely the intersection of the polytope STAB(AA') with the hyperplane describing the partial trace constraint. It is itself a convex polytope and hence has finitely many extremal points. In fact, in some cases this partial trace constraint can even be dropped, which reduces the problem to finding a QPD of the Choi state w.r.t. STAB.

Proposition 6.9: [141, Theorem 3.2]

Let A, A' be multi-qubit systems. If \mathcal{E} is a unitary channel corresponding to the third level of the Clifford hierarchy, then $\gamma_{\text{CSP}}(\mathcal{E}) = \gamma_{\text{STAB}}(J_{\mathcal{E}})$.

It is a simple exercise to show that the Choi state of the T gate is equivalent to $|0\rangle \otimes |H\rangle$ up to Clifford gates. From Example 6.8, we can see that the CSP and Choi-based lower bounds are therefore tight for the T gate

$$\gamma_{\text{SO}}(T^{\otimes m}) \ge \gamma_{\text{CSP}}(T^{\otimes m}) = \gamma_{\text{STAB}}(J_T^{\otimes m}) = \gamma_{\text{STAB}}(|H\rangle\langle H|^{\otimes m}) = \gamma_{\text{SO}}(T^{\otimes m}).$$
(6.11)

6.4 Further reading

Historically, he resource theory of magic has was first developed for odd-dimensional quantum systems, since they exhibit well-behaved finite-dimensional analogs of the Wigner function [69]. For instance, Veitch *et al.* [162] introduced a magic monotone called *mana*, which has a striking resemblance to the quasiprobability extent, as it is defined to be the logarithm of the 1-norm of the Wigner function.

The first instance of QPS for magic simulation (or as a matter of fact of QPS in general) dates back to Pashayan *et al.* [120]. They frame their method in terms of quasiprobability representations w.r.t. some frame and its dual frame. As such, this formulation lends itself very naturally to odd-dimensional qudit systems. Later work by Howard *et al.* [86] extended QPS to qubit systems and studied QPDs of qubit magic states to be used in injection circuits. They introduced the quasiprobability extent of states and characterized some of its properties, like the strict sub-multiplicativity. Furthermore, they also used the quasiprobability extent to find bounds on optimal rates of synthesis for certain gates. It should be noted that their definition of robustness slightly differs from the canonical notion introduced by Vidal and Tarrach [166]. They denote by "robustness of magic" what we call the quasiprobability extent.

The first notion of QPS of magic channels was proposed shortly thereafter in Ref. [14]. The authors apply their technique to the simulation of noisy quantum circuits, which allows them to simulate quantum error correction systems with non-Clifford errors — a regime that is out of reach for the more conventional pure-state stabilizer rank simulators. For the decomposition set, they consider the set of stabilizer-preserving operations. The quasiprobability extent of channels was further developed in Ref. [141]. The authors prove the Choi characterization of the set of completely stabilizer-preserving operation as well as various properties of the induced quasiprobability extent.

The difference between the completely stabilizer-preserving operations and and the more operational stabilizer operations was investigated in Ref. [78]. Interestingly, the authors prove an explicit separation between these two sets, and they provide a normal form of general stabilizer operations.

Due to the strictly sub-multiplicative behavior and the lack of analytic description for γ_{STAB} , a significant amount of research has gone into improving the efficiency of its numerical evaluation. Notably, Heinrich *et al.* demonstrated how the complexity of computing the quasiprobability extent can be reduced super-polynomially by exploiting symmetries of the stabilizer polytope and the state in question [79]. As such, they can exactly estimate $\gamma_{STAB}(|H\rangle\langle H|^{\otimes m})$ for up to m=9. Using approximations, they can estimate almost tight upper bounds of the quasiprobability extent for even larger m. Further algorithmic improvements by Hamaguchi *et al.* [72] allow the evaluation of the quasiprobability extent of arbitrary states for up to 8 qubits. We refer to [72, Table 1] for a nice succinct overview of the different approaches.

6.5 Contributions

This chapter mostly surveys recent developments in QPS for magic simulation from other authors in literature. It is included it in this thesis for the sake of completeness. The author's contributions are mainly focused on contextualizing these results within the QPS framework developed in previous chapters. We also note that to our knowledge, classical side information from intermediate measurements has not been explored in the context of the QRT of magic before.

Chapter 7

Simulating noise-free computation

Experimental realizations of quantum computers suffer from a large amount of inherent noise, which makes their practical utilization difficult. In this chapter, we will study a method to remove the effect of noise by using QPS to simulate ideal noise-free quantum gates with a noisy quantum computer. More concretely, we consider a decomposition set *S* that contains a selection of noisy quantum operations that our computer can execute, such as noisy gates, noisy measurements and combinations thereof. This noise suppression is relatively straightforward to implement in practical experiments, at least compared to full-blown quantum error correction (QEC). It avoids the need for encoding the quantum information into a larger system as well as all the other difficulties that come with syndrome extraction and decoding. However, it also entails a sampling overhead that scales exponentially in the circuit size. This is perhaps not too surprising, given that the coherence time of the qubits is not physically prolonged.

As such, QPS of noise-free computation is an instance of quantum error mitigation (QEM). In fact, it is one of the prototypical QEM methods and it commonly goes by the name *probabilistic error cancellation*. For near-term devices, probabilistic error cancellation holds the promise to systematically remove the effect of noise in a quantum computation without the need for any additional quantum resources. It has been at the heart of some of the most impressive experimental quantum computing demonstrations that have come out in recent times [94]. Even as quantum hardware is poised to improve in the future, the overhead of error correction in early fault-tolerant computers is likely to make any practical utilization very difficult. As such, some recent research has focused on finding ways to seamlessly combine QEC and QEM to split up the burden of suppressing noise and therefore interpolate between the advantages and disadvantages of the two approaches. We will touch upon some of these ideas in this chapter.

Contrary to other applications of QPS that we explored in previous chapters, probabilistic error cancellation does unfortunately not exhibit a comparatively rich mathematical formalism, as the decomposition set is defined in a more ad-hoc fashion and a priori lacks any sort of structure. Many of the challenges in probabilistic error cancellation are of more practical nature. The goal of this chapter will be to briefly outline some of these difficulties as well as a powerful technique based on

randomized compiling, which makes it possible to address these challenges while effectively enforcing some strict structure on the decomposition set.

Remark 7.1: Since the decomposition set in probabilistic error cancellation is not derived from a QRT, it is a priori not clear that QPDs of smaller circuit elements can be combined into a QPD of the total circuit (recall the discussion in Remark 3.36). Indeed, if an imperfect quantum computer can implement the noisy operations $\mathcal E$ and $\mathcal F$, it might not be able to implement $\mathcal E\otimes \operatorname{id}$ or $\mathcal E\circ\mathcal F$ due to cross-talk and temporally correlated noise. This is an additional difficulty that practical QEM techniques need to take into account.

7.1 Basic setup of probabilistic error cancellation

In the following, we describe a simple procedure how to simulate a unitary quantum channel \mathcal{U} using probabilistic error cancellation.

- 1. Pick some finite set of (noisy) operations that the noisy quantum computer can natively execute.
- 2. Characterize these noisy operations, for instance with a quantum process tomography protocol. The resulting quantum channels constitute the decomposition set *S*.
- 3. Find the optimal QPD of \mathcal{U} into the decomposition set S using the linear program introduced in Section 3.6.
- 4. Perform QPS of \mathcal{U} w.r.t. the decomposition set S.

There are some potential difficulties with this type of approach. For instance, the choice of the operations for the decomposition set is generally unclear, as the hardware typically supports the execution of a continuum of gates and the choice of the decomposition set can strongly impact the resulting quasiprobability extent of \mathcal{U} . To achieve a small quasiprobability extent, S should typically include the noisy realization $\mathcal{U}_{\text{noisy}}$ of the gate \mathcal{U} . As such, we would need to pick a different decomposition set for every different type of gate. This is not only cumbersome from a theoretical perspective, but more importantly also impractical, as the number of noisy operations that need to be characterized is very large.

As a solution to this problem, many protocols instead tackle the problem from a noise inversion perspective¹. A noisy gate $\mathcal{U}_{\text{noisy}}$ can always be seen as a composition of the ideal gate with a noise channel $\mathcal{N} \circ \mathcal{U}$ where $\mathcal{N} := \mathcal{U}_{\text{noisy}} \circ \mathcal{U}^{-1}$. So instead of quasiprobabilistically simulating \mathcal{U} , one can instead physically execute $\mathcal{U}_{\text{noisy}}$ and then immediately thereafter quasiprobabilistically simulate $^2 \mathcal{N}^{-1}$. Since (at least for

¹This is essentially the same noise inversion procedure we encountered as an application of non-physical computation in Section 4.3, but with a noisy decomposition set.

²Note that we assume the noise channel to be invertible here. The inverse is not required to be a physical CPTP map itself.

weak noise) \mathcal{N}^{-1} is close to the identity no matter the choice of gate \mathcal{U} , it is much more reasonable to use the same decomposition set S for every type of gate. This approach is sometimes called the *inverse method* to distinguish it from the *compensation* method that we introduced initially.

Example 7.2. As a simple toy model, consider a quantum computer that can only execute noisy single-qubit quantum gates, which consist of the ideal gate followed by the depolarizing channel

$$\mathcal{N}_{\epsilon}(\rho) := (1 - \epsilon)\rho + \epsilon \operatorname{tr}[\rho] \frac{1}{2}$$
 (7.1)

for some noise rate ϵ . For our decomposition set, we simply pick the four noisy single-qubit Pauli channels

$$S = \{ \mathcal{N}_{\epsilon}, \mathcal{N}_{\epsilon} \circ \mathcal{X}, \mathcal{N}_{\epsilon} \circ \mathcal{Y}, \mathcal{N}_{\epsilon} \circ \mathcal{Z} \}$$
 (7.2)

where X, Y, Z are the unitary channels induced by X, Y, Z. Then the sampling overhead associated with inverting the noise of any quantum gate is given by

$$\gamma_S(\mathcal{N}_{\epsilon}^{-1}) = \frac{1 + \epsilon - \epsilon^2}{1 - 2\epsilon + 2\epsilon^2} \tag{7.3}$$

which is achieved by the following QPD³

$$\mathcal{N}_{\epsilon}^{-1} = \frac{4 - \delta}{4(1 - \delta)} \mathcal{N}_{\epsilon} - \frac{\delta}{4(1 - \delta)} \left(\mathcal{N}_{\epsilon} \circ \mathcal{X} + \mathcal{N}_{\epsilon} \circ \mathcal{Y} + \mathcal{N}_{\epsilon} \circ \mathcal{Z} \right) \tag{7.4}$$

where $\delta := 2\epsilon(1 - \epsilon)$.

This example nicely demonstrates an important property of probabilistic error cancellation: as the physical noise rate of the hardware improves, the sampling overhead decreases. Moreover, the quasiprobability extent converges to 1 as the physical error rate vanishes. This is nicely illustrated by the first-order Taylor expansion for small ϵ

$$\gamma_S(\mathcal{N}_{\epsilon}^{-1}) \approx 1 + 3\epsilon$$
 (7.5)

In order to achieve a crude theoretical handle on what sort of quasiprobability extents are achievable through the inverse method, we can look at the maximal possible extension of the decomposition set

$$\gamma_S(\mathcal{N}^{-1}) \ge \gamma_{\text{CPTP}^*}(\mathcal{N}^{-1}) = \left\| \mathcal{N}^{-1} \right\|_{\Diamond} = \frac{1 + \epsilon/2}{1 - \epsilon} \approx 1 + \frac{3}{2}\epsilon$$
(7.6)

where the inequality follows from Lemma 3.10 and the equalities were shown in Example 4.8. As such, the first-order factor in the overhead differs by a factor of two from the theoretically optimal value. \triangle

 $^{^3}$ One way to verify this is by multiplying both sides by $\mathcal{N}_{\varepsilon}$ from the left and then applying Example 4.8.

7.2 Tailoring noise with randomized compilation

There are a few technical difficulties which need to be addressed for an experimental realization of probabilistic error cancellation. The most pressing one is that one requires exact knowledge of the underlying noise of the hardware. This is in stark contrast to QEC, which is typically quite robust w.r.t. the underlying noise model. For a simple back-of-the-envelope calculation, assume that we erroneously characterize the noisy operations $\mathcal{E}_1, \ldots, \mathcal{E}_n$ and instead use the wrong decomposition set $S = \{\tilde{\mathcal{E}}_1, \ldots, \tilde{\mathcal{E}}_n\}$ with an error of about $\|\mathcal{E}_i - \tilde{\mathcal{E}}_i\| \approx \epsilon$. If we choose a QPD w.r.t. the erroneous basis $\mathcal{U} = \sum_i a_i \tilde{\mathcal{E}}_i$ and try to simulate \mathcal{U} , we effectively cause an error

$$\left\| \mathcal{U} - \sum_{i} a_{i} \mathcal{E}_{i} \right\| = \left\| \sum_{i} a_{i} (\tilde{\mathcal{E}}_{i} - \mathcal{E}_{i}) \right\| \lesssim \|a\|_{1} \epsilon. \tag{7.7}$$

In summary, the quality of the simulation is at most as good as the characterization of the involved noisy operations.

The standard method for characterizing quantum channels is through the quantum process tomography protocol. However, in practice it is inherently unreliable due to state preparation and measurement (SPAM) errors. There exist modifications to process tomography that allow for the elimination of systematic errors caused by SPAM noise (see e.g. *gate-set tomography* [68] and its utilization in QEM discussed by Endo *et al.* in Ref. [57]), but they are very expensive and typically impractical. To make matters worse, typical quantum devices suffer from cross-talk and correlated noise. Modelling noisy gates by a simple quantum channel acting *only* on the involved qubits is generally a bad approximation, and one needs to also account for the effect on surrounding qubits. This further exacerbates the difficulties of tomography.

In this section, we briefly present an elegant solution to these problems due to van den Berg *et al.* [15]. The foundational idea is based on the technique of *randomized compilation* [168] which enforces the noise in the circuit to effectively be Pauli channels. We only present the main principles and refer the reader to Ref. [15], [168] for more details.

The main idea is to separate the gates into two categories: "easy gates" and "hard gates", which can correspondingly be implemented with more or less noise. For the purposes of QEM, it is natural to choose the easy gates to be single-qubit gates and the hard gates to be CNOT gates. The circuit is modelled as alternating layers of easy and hard gates. Randomized compiling effectively twirls the noise to enforce it to be of Pauli type. This is achieved by pushing the Pauli twirl gates through the CNOT layer (which is Clifford) and integrating them into the previous and subsequent easy gate layers. This noise tailoring is exact as long as the noise of the easy gates can be assumed to be gate-independent [168, Theorem 1].

Having effectively tailored the noise to be purely of Pauli type, it remains to characterize the resulting Pauli channel. Luckily, Pauli channels are much easier to estimate than general channels [59] and expensive error-prone process tomography can be avoided. The main idea is to repeat the CNOT gate layer multiple times and measure the corresponding Pauli expectation values at every depth. The noise

parameters can be fully extracted from the exponential decay rate, making them robust against any type of SPAM errors⁴. Still, the number of parameters to learn is generally exponentially large in the number of qubits. This can in term be avoided by assuming some sparsity structure in the noise. Sparse Pauli-Lindblad models have proven very effective in the context of superconducting quantum processors [15].

Since the resulting noise channels \mathcal{N} are Pauli, \mathcal{N}^{-1} can itself be decomposed into a QPD of Pauli channels. In fact, this QPD is even optimal in the sense $\gamma_{\mathcal{P}}(\mathcal{N}^{-1}) = \gamma_{\text{CPTP}^*}(\mathcal{N}^{-1})$ where \mathcal{P} denotes the set of Pauli channels (recall Proposition 4.6).

QPS based on randomized compilation and sparse Pauli-Lindblad models has been at the heart of some of the most impressive experimental demonstrations of quantum computing in recent years [94]. Though it should be noted that in Ref. [94], QPS is not used to remove the effect of noise, but rather to extrapolate its strength in order to combine the method with another well-known QEM technique called *zero-noise noise extrapolation* [154].

7.3 Combining QEM and QEC

Early implementations of QEC will likely be unable to implement any useful quantum algorithms due the restricted number of logical qubits and limited ability to sufficiently suppress the logical error rates. Recent research is thus increasingly focused on finding meaningful ways to combine QEC and QEM to split the burden of suppressing the noise. This prospect is very enticing, as one could interpolate between the drawbacks of QEC and QEM. When the QEC capabilities of the hardware are more limited, QEM would do most of the heavy lifting, and once the hardware improves, QEC would become more dominant and thus reduce the sampling overhead.

The most straightforward way to achieve this goal is by simply performing error mitigation on the logical level, i.e., simply treat the logical qubits as if they were physical qubits and then run any QEM technique on top of that. This is unlikely to be the most resource efficient approach, and more refined techniques should more specifically take into account and address the specific aspects of QEC that are especially difficult to realize in early fault-tolerant systems. In this section, we will present precisely such a technique.

Logical Clifford gates are typically easier to realize fault-tolerantly than non-Clifford gates. In fact, for some CSS codes they can be implemented transversally, i.e., by only coupling equivalent qubits across code blocks. Transversal gates have a manageable overhead, because they do not significantly spread errors and are hence fault-tolerant by definition. The Eastin-Knill theorem proves that no error correcting code can transversally implement a universal set of gates [55]. To achieve universal computation, an additional fault-tolerant non-Clifford gate is thus needed. The T gate is usually considered for this purpose. A common technique to implement a T gate is through through a magic state injection, as previously discussed in Figure 3.2. This approach reduces the problem of implementing a fault-tolerant T gate

⁴This procedure is reminiscent of randomized benchmarking [56].

to the task of a fault-tolerant preparation of an encoded magic state. The latter task can be achieved by magic state distillation [24], where several noisy magic states are repeatedly transformed into fewer magic states of better fidelity. While magic state distillation is a very elegant approach for achieving universal fault-tolerant quantum computing, the distillation process leads to a considerable overhead in practice. It is meaningful to investigate the possibility of using error mitigation techniques to reduce this overhead.

In this section, we show two possible techniques how to perform probabilistic error cancellation on the logical level to simulate the execution of noise-free logical T gates using a noisy T gate, thus circumventing the need for expensive magic state distillation. In both cases, the quasiprobability extent of the logical T gate will be $\gamma_{\epsilon} = 1 + 2\kappa\epsilon + \mathcal{O}(\epsilon^2)$ where ϵ is the physical error rate and κ is some constant. The first method achieves a value of $\kappa \approx 2/5$, however it requires an additional logical ancilla qubit. In contrast, the second method is easier to implement but incurs a larger overhead $\kappa \approx 30$. When simulating a Clifford+T circuit with this technique, the sampling overhead scales as γ_{ϵ}^{2t} where t is the number of T gates. This overhead is much smaller than the best currently known classical simulation algorithm, which scales as 1.3161^t [133], even for modestly low physical error rates ϵ .

In the following, we present the two methods in detail.

Error-mitigated logical T gates via noisy magic states We consider an implementation of a T gate through magic state injection, but with a noisy magic state instead of a distilled noise-free one. This will lead to a faulty T gate. In the second step we use probabilistic error cancellation to suppress the error. Let ρ be the noisy encoded magic state. We define the logical error rate as

$$\bar{\epsilon} := 1 - \langle H | \rho | H \rangle . \tag{7.8}$$

Recall that $|H\rangle := \left(|0\rangle + e^{i\pi/4}\,|1\rangle\right)/\sqrt{2}$. For simplicity, assume that the chosen error correction code enables a noiseless implementation of logical gates S, X, CNOT, preparation of logical $|0\rangle$ and $|+\rangle$ states, and measurements of a logical qubit in the Z and X basis 5 . We can twirl the noisy magic state ρ with respect to the group $\{1,A\}$, where A is the Clifford gate defined as

$$A := e^{-i\pi/4}SX = |H\rangle\langle H| - |\omega\rangle\langle\omega| , \qquad (7.9)$$

an $|\omega\rangle := Z|H\rangle$ is a state orthogonal to $|H\rangle$. The twirled state thus becomes

$$\tau := \frac{1}{2}(\rho + A\rho A^{\dagger}) = (1 - \bar{\epsilon}) |H\rangle\langle H| + \bar{\epsilon} |\omega\rangle\langle\omega|, \qquad (7.10)$$

where the second equality follows from the identity $A |H\rangle\langle\omega| A^{\dagger} + |H\rangle\langle\omega| = 0$. Therefore, we can prepare the twirled state τ by first preparing ρ and then applying the logical A gate with probability 1/2. The twirled state τ can be considered as the ideal

⁵The noise can be exponentially suppressed by choosing a large enough distance. These gates are Clifford gates that feature a transversal and hence fault-tolerant implementation for self-dual CSS codes.

magic state $|H\rangle$ that suffers from a random Pauli Z error applied with probability $\bar{\epsilon}$, since $|\omega\rangle = Z|H\rangle$. Note that a Pauli Z error on the magic state $|H\rangle$ has no effect on the measurement in Figure 3.2. Using the twirled state τ instead of the ideal magic state $|H\rangle$ inside the T gate gadget therefore leads to the noisy T gate

$$\mathcal{T}_{\bar{\epsilon}} = (1 - \bar{\epsilon})\mathcal{T} + \bar{\epsilon}\,\mathcal{Z} \circ \mathcal{T}\,,\tag{7.11}$$

where \mathcal{T} and \mathcal{Z} denote the quantum channels implementing the ideal T and Z gate, respectively, i.e., $\mathcal{T}(X) = TXT^{\dagger}$ and $\mathcal{Z}(X) = ZXZ^{\dagger}$. To use error mitigation to simulate a perfect T gate using $\mathcal{T}_{\bar{e}}$, we need a good estimate for the logical error rate \bar{e} . This could be done via standard tomography techniques, however we present a more robust method further below.

It is crucial to understand how the logical error rate $\bar{\epsilon}$ defined in Equation (7.8) depends on the physical error rate ϵ . It has been shown that there exist noisy state preparation circuits for the surface code that achieve logical error rates that are asymptotically constant w.r.t. to the code distance [104]. In fact, by choosing the right protocol, $\bar{\epsilon}$ can be in the same order as the physical error rate ϵ

$$\bar{\epsilon} = \kappa \epsilon + \mathcal{O}(\epsilon^2) \,, \tag{7.12}$$

where κ is a conversion constant that depends on the details of the noise model. For example, in the case of perfect single-qubit gates and no initialization errors, one can achieve $\kappa = 2/5$ [101]. The second order term in Equation (7.12) has numerically been shown to be negligible for realistic error rates using the protocol from Ref. [101].

We now discuss how to experimentally determine the logical error rate $\bar{\epsilon}$. Denote the phase flip channel by $\mathcal{N}_{\bar{\epsilon}} := (1 - \bar{\epsilon}) \mathrm{id} + \bar{\epsilon} \mathcal{Z}$. Since \mathcal{T} and \mathcal{Z} commute, we have for any $p \in \mathbb{N}$

$$\mathcal{T}^{p}_{\bar{\epsilon}} = (\mathcal{N}_{\bar{\epsilon}} \circ \mathcal{T})^{p} = \mathcal{N}^{p}_{\bar{\epsilon}} \circ \mathcal{T}^{p}. \tag{7.13}$$

For $p \equiv 0 \pmod 8$ we have $T^p = 1$ and $T^p = id$. Hence, Equation (7.13) ensures that $\mathcal{T}^p_{\bar{\epsilon}} = \mathcal{N}^p_{\bar{\epsilon}}$. We can now learn $\bar{\epsilon}$ by preparing a logical state $|+\rangle$ fault-tolerantly, applying $\mathcal{T}^p_{\bar{\epsilon}}$ and measuring the output in the $\{|+\rangle, |-\rangle\}$ basis, which gives us an outcome 1 or 0. Hence we can estimate the expectation value of the outcome

$$f(p) := \langle + | \mathcal{T}_{\bar{\epsilon}}^p(|+\rangle\langle + |) | + \rangle = \frac{1}{2} \left(1 + (1 - 2\bar{\epsilon})^p \right), \tag{7.14}$$

with an approximation error that decreases in the number of circuit shots ⁶. By measuring f(p) for p = 8k with $k \in \mathbb{N}$, we can obtain a good estimate for $\bar{\epsilon}$ using exponential fitting.

The following QPD optimally decomposes the ideal T gate into its noisy variant with and without a subsequent Z gate

$$\mathcal{T} = \left(\frac{1 - \bar{\epsilon}}{1 - 2\bar{\epsilon}}\right) \mathcal{T}_{\bar{\epsilon}} - \left(\frac{\bar{\epsilon}}{1 - 2\bar{\epsilon}}\right) \mathcal{Z} \circ \mathcal{T}_{\bar{\epsilon}}, \tag{7.15}$$

with a 1-norm of $\gamma_{\epsilon} = 1/(1-2\bar{\epsilon})$. Expanding this term around $\bar{\epsilon} = 0$ gives

$$\gamma_{\epsilon} = 1 + 2\bar{\epsilon} + \mathcal{O}(\bar{\epsilon}^2) = 1 + 2\kappa\epsilon + \mathcal{O}(\epsilon^2).$$
 (7.16)

⁶The final step in Equation (7.14) can be checked with a simple recursive argument.

Error-mitigated logical T gates via code switching The implementation of a T gate via magic states has the drawback of using two encoded qubits. It also includes a logical CNOT gate which may require more logical qubits for certain implementations [85]. One may ask if a logical T gate can be implemented using only one logical qubit. In this section we show how to accomplish this task by combining probabilistic error cancellation with the code switching method pioneered by Paetznick and Reichardt [118].

Suppose S_1 and S_2 are CSS-type [32], [150] quantum codes with one logical qubit. Our goal is to implement a logical T gate on a qubit encoded by S_1 . We assume that \mathcal{S}_1 has a large distance for both X and Z errors and enables a fault-tolerant noise-free implementation of the S gate. We assume that S_2 is an asymmetric code with a large distance for X errors and a distance 1 for Z errors such that a logical-Z operator of S_2 can be chosen as a single-qubit Pauli Z on some physical qubit. We shall denote this local logical-Z operator as \overline{Z}_{loc} . This allows us to perform a logical T gate on a qubit encoded by \mathcal{S}_2 simply by applying a physical T gate on the qubit acted upon by \overline{Z}_{loc} . Indeed, since the T gate is a linear combination of the identity and the Pauli Z, the physical and the logical T gates become equivalent. The code switching method enables a conversion between the codes S_1 and S_2 by measuring stabilizers of S_2 on a logical state encoded by S_1 or vice versa. In certain cases this conversion can be performed fault-tolerantly [118]. This is achieved by identifying stabilizers present in both codes and using the measured syndromes of such stabilizers to diagnose and correct errors. The error suppression typically scales exponentially with the code distance. In our case the code switching protects the encoded qubit exponentially well only from logical X errors. This ensures that the effective noise channel acting on the logical qubit is dominated by Z-type (possibly coherent) errors. Let $\mathcal{T}_n = \mathcal{N} \circ \mathcal{T}$ be the noisy logical T gate. We can tailor the noise $\mathcal N$ to become stochastic Pauli-Z noise using by appropriate twirling

$$\frac{1}{2}(\mathcal{N} + \mathcal{X} \circ \mathcal{N} \circ \mathcal{X}) = (1 - \bar{\epsilon})id + \bar{\epsilon}\mathcal{Z} =: \mathcal{N}_{\bar{\epsilon}},$$
 (7.17)

for some logical error rate $\bar{\epsilon}$ where $\mathcal{X}(B) = XBX^{\dagger}$. Using the identity XT = TA we get

$$\frac{1}{2}(\mathcal{T}_n + \mathcal{X} \circ \mathcal{T}_n \circ \mathcal{A}) = \mathcal{N}_{\bar{e}} \circ \mathcal{T} = \mathcal{T}_{\bar{e}}, \qquad (7.18)$$

where $\mathcal{A}(B) = ABA^{\dagger}$ and $\mathcal{T}_{\bar{e}}$ is defined as in Equation (7.11). By assumption, the code \mathcal{S}_1 enables a fault-tolerant implementation of the Pauli X and the S gate on the encoded qubit. Since A = SX up to an overall phase, the twirling in Equation (7.18) can be implemented fault-tolerantly on a qubit encoded by \mathcal{S}_1 by applying either \mathcal{T}_n or $\mathcal{X} \circ \mathcal{T}_n \circ \mathcal{A}$ with the probability 1/2 each. Since at this point we have implemented a noisy T gate with Pauli-Z noise, we can follow the same procedure as for the first method in order to determine the parameter \bar{e} efficiently and perform probabilistic error cancellation to mitigate the noise.

In Appendix C we specialize the code switching method to Kitaev's surface code. It is shown that the logical error rate scales as $\bar{\epsilon} \approx 30\epsilon$ for realistic error models. The code switching procedure is depicted in Figure 7.1.

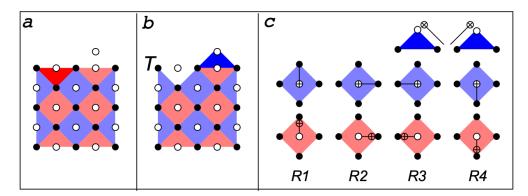


Figure 7.1: Logical T-gate by code switching. (a) Surface code S_1 with the distance d=3. Black and white circles indicate data and syndrome qubits. Red and blue faces indicate X and Z stabilizers. (b) Asymmetric surface code S_2 exhibiting a single-qubit logical T-gate at the north-west corner of the lattice. Code switching $S_1 \rightarrow S_2$ is performed by turning off X stabilizer F_1 (dark red triangle) and turning on Z stabilizer G_1 (dark blue triangle). (c) The syndrome extraction cycle consists of four rounds of CNOTs R1, R2, R3, R4. To avoid clutter we only show a local schedule of CNOTs for each stabilizer. Schedules for weight-3 stabilizers are properly truncated. The schedule extends to the full lattice in a translation invariant fashion.

7.4 Further reading

Probabilistic error cancellation was originally introduced by Temme, Bravyi and Gambetta [154] together with another important QEM technique, zero noise extrapolation. Their approach is formulated in terms of the compensation method, and they also formulated the linear program in Equation (3.101) expressing the optimal quasiprobability extent w.r.t. of a finite decomposition set. Shortly thereafter, Ref. [57] proposed several improvements to the technique, including the inverse method, an explicit decomposition set of single-qubit Clifford operations that spans the space of Hermitian-preserving superoperators as well as the usage of gate-set tomography to address the SPAM problem.

To minimize the sampling overhead of quantum error mitigation, it is of high interest to choose the best possible decomposition basis as to minimize the quasiprobability extent. This is generally very difficult, unless one uses the randomized compiling approach from Section 7.2. In Ref. [131], a heuristic method is proposed to iteratively construct the decomposition set in a greedy fashion. Importantly, no assumptions on the hardware noise is required, since it is considered in each iteration through an instance of gate tomography.

A different approach in Ref. [153] entirely avoids the choice of a finite decomposition set by modelling all gates to have identical noise \mathcal{N} . As such, the authors model their decomposition set as the union of all noisy gates $\{\mathcal{N} \circ \operatorname{Ad}[U] | U \in \operatorname{U}(A)\}$ together with (noise-free) state preparation channels. While this model is likely a strong oversimplification, it makes the computation of the quasiprobability extent

more tractable and allowed the authors to find bounds or even closed-form formulas in some instances.

PEC based on randomized compilation was first proposed in Ref. [15]. To characterize the twirled Pauli noise, the authors utilize a sparse Pauli-Lindblad model that can capture correlated noise across neighboring qubits on a superconducting quantum processor. With an extensive amount of theoretical work, the authors show how this characterization can be done efficiently and reliably. The methods are based on a broader line of research that shows that characterizing sparse Pauli noise can be done efficiently [59].

The idea of integrating QEM with QEC was first formulated in Ref. [152] and in Ref. [105], [130]. The former considers the possibility of using probabilistic error cancellation on the logical level to correct errors that either stem from erroneous decoding or from approximation errors in the Solovay-Kitaev decomposition. The latter independently developed similar ideas, which broadly correspond to the discussion in Section 7.3 on mitigating noisy non-transversal gates. Many subsequent works were dedicated to further study the interplay between QEC and QEM [54], [62], [156], [167].

Finally, we briefly mention a few other QEM techniques for the interested reader. Zero-noise extrapolation was introduced by Ref. [102], [154] and tries to remove the bias from expectation values by extrapolating the circuit result while varying the noise strength. A class of techniques called *symmetry verification* uses post-selection to verify that the quantum state fulfills certain symmetries that should be conserved in the absence of noise [17], [109]. Another approach called *subspace expansion* can mitigate noise in the task of eigenvalue estimation by classically solving the eigenvalue problem on a small subspace that is spanned by noisy states prepared by the hardware [110]. We refer to the review in Ref. [31] for an extensive overview of the state of quantum error mitigation.

7.5 Contributions

Most of this section serves as an overview of probabilistic error cancellation. The method presented in Section 7.3 is due to the author and was originally published in Ref. [130]. We briefly note that the author also developed a heuristic algorithm to find near-optimal decomposition sets for probabilistic error cancellation using the compensation method, which was published in Ref. [131]. We chose not to include this work in this thesis.

Chapter 8

Conclusion

In this thesis, we have explored the application of QPS to four different settings, each corresponding to a different choice of decomposition set. Looking back at these results, we revisit some of the important goals and questions that we originally stated in the introduction.

Utility of classical side information Since any linear combination of CPTP maps is itself proportional to an HPTP map, it is absolutely necessary to use classical side information for the simulation of any non-TP map. However, we saw only very few practical examples where such non-TP maps actually occur, as even non-physical maps of interest are usually HPTP. One exception is the imaginary time evolution discussed in Section 4.3.

For the simulation of CPTP maps, the utility of classical side information becomes a bit more murky and strongly depends on the underlying QRT. For space-like and time-like cuts without classical communication, we saw that classical side information is absolutely crucial, as nontrivial channels cannot even be represented as a QPD without it (see Lemmas 5.7 and 5.73). However, for many other decomposition sets we showed that side information is useless as the quasiprobability extent of a CPTP map $\mathcal E$ remains unchanged without it, i.e., $\gamma_{\text{CPTP}}(\mathcal E) = \gamma_{\text{CPTP}^*}(\mathcal E)$ (Proposition 4.2), $\gamma_{\text{SEPC}}(\mathcal E) = \gamma_{\text{SEPC}^*}(\mathcal E)$, $\gamma_{\text{PPTC}}(\mathcal E) = \gamma_{\text{PPTC}^*}(\mathcal E)$ (Proposition 5.10) and $\gamma_{\text{EBC}}(\mathcal E) = \gamma_{\text{EBC}^*}(\mathcal E)$ (Proposition 5.74). For space-like cuts with classical communication and for Clifford operations, it remains unknown whether $\gamma_{\text{LOCC}^*}(\mathcal E)$ and $\gamma_{\text{SO}^*}(\mathcal E)$ can be strictly smaller than $\gamma_{\text{LOCC}}(\mathcal E)$ and $\gamma_{\text{SO}}(\mathcal E)$. We highlight that currently known constructions for QPDs that achieve $\gamma_{\text{LOCC}^*}(\mathcal U)$ for non-Clifford unitary channels $\mathcal U$ heavily utilize classical side information, hinting at a possible separation $\gamma_{\text{LOCC}}(\mathcal U) \neq \gamma_{\text{LOCC}^*}(\mathcal U)$.

Remarkably, we do not know of any example where side information can reduce the quasiprobability extent, but is not absolutely required to find a valid QPD in the first place.

Submultiplicative behavior of quasiprobability extent For the resource theories of entanglement and magic, we saw that the associated quasiprobability extents exhibit a strictly sub-multiplicative behavior for many states and channels of interest.

Remarkably, in the QRT of entanglement we were able to evaluate $\overline{\gamma}_{\text{SEP}}$ and $\gamma_{\text{SEP}}^{\infty}$ for all pure states and show that they strictly differ in almost all instances. A similar feat seems out of reach for the QRT of magic, as even for the simplest magic state $|H\rangle$ there is no known closed-form expression of the regularized quasiprobability extent.

In stark contrast, the quasiprobability extent γ_{CPTP^*} is multiplicative and hence $\gamma_{\text{CPTP}^*} = \overline{\gamma}_{\text{CPTP}^*}$. This result also translates to some instances of probabilistic error cancellation, e.g., when using randomized compilation for noise tailoring as discussed in Section 7.2.

State of circuit knitting theory With the results presented in this thesis, we arguably have a mostly complete understanding of the practically relevant aspects of QPS-based circuit knitting. Concretely, we know the optimal overheads for cutting bipartite states, two-qubit unitaries and wire cuts, as well as explicit protocols that achieve these overheads. There is little room left to improve the sampling overhead in practical scenarios, unless a different circuit knitting paradigm emerges that operates in a different manner.

Practical utility of QPS Both circuit knitting and error mitigation have a clear practical appeal as they provide a potential avenue to overcome some of the stringent limitations of early quantum devices. Correspondingly, they have received a lot of attention in the last few years and have been experimentally demonstrated multiple times. However, we stress that it remains unclear if they can be used to demonstrate a useful quantum advantage on near-term devices, since the exponential sampling overhead is very steep and kicks in rather quickly. Especially for circuit knitting, it remains a crucial open problem to identify suitable quantum computations which are "nearly local" with only a small amount of nonlocality across the partitions. The clustered Hamiltonian simulation scheme by Harrow and Lowe [74], briefly discussed in Section 5.1.6, likely presents the most promising candidate in this regard. At least for QEM, there exists some experimental evidence that it enables current devices to start solving non-trivial problems at useful scales [94], though a convincing quantum advantage has not been demonstrated to date.

Due to the steep exponential sampling overhead, it seems plausible that error mitigation and circuit knitting could find practical utility in combination with more traditional approaches that scale better asymptotically. For example, we explored in Section 7.3 the possibility to combine error correction with QPS to simulate ideal logical T gates with noisy logical T gate. This avoids costly magic state distillation and comes with a sampling overhead $\mathcal{O}(\gamma_{\varepsilon}^{2t})$ exponential in the number of t gates where the quasiprobability extent γ_{ε} converges to 1 as the physical error rate vanishes. To further reduce the sampling overhead, one could perform a small number of rounds of magic state distillation (insufficient for the desired logical accuracy) and then use QEM to mitigate the remaining noise of the intermediate magic state 1 .

Similar approaches could also be envisaged for circuit knitting: Instead of simulating non-local computation purely with local operations, one could instead simulate

¹This combination of QEM with conventional magic state distillation only works with the noisy magic state approach, but not with the code switching method.

it with local operations and noisy nonlocal resources. As the fidelity of this nonlocal resource improves, the resulting sampling overhead would decrease. Some recent work has explored ideas in this direction [9].

The idea of using QPS to simulate nonphysical operations has only been sparsely explored to date. In Section 4.3, we briefly discussed some basic applications like noise inversion and entanglement detection, but further research is necessary. We highlight the simulation of imaginary time evolution as a particularly interesting idea in this direction.

Finally, we briefly re-iterate that QPS for magic simulation only has limited practical utility, as there exist significantly more efficient classical algorithms to simulate near-Clifford circuits. One advantage of QPS-based classical simulators is that they naturally allow for the inclusion of non-unitary channels, such as noise. This can be particularly useful for the simulation of quantum error correction protocols under weak non-Pauli noise.

Utility of classical communication for circuit knitting Classical communication seems to have only limited practical utility for space-like cuts. First, we saw that for states, γ_{LOCC^*} and γ_{LO^*} both reduced to γ_{SEP} (Lemma 5.5). Furthermore, we saw that for two-qubit gates, γ_{LOCC^*} and γ_{LO^*} were also identical (Theorem 5.44). This equality even remains when considering tensor products of multiple two-qubit unitaries and also in the black box setting (Theorem 5.66). It remains an open question if a gap between the two arises for general unitaries.

Quite surprisingly, this picture is completely different for time-like cuts. Cutting a single wire is already cheaper with classical communication as $\gamma_{EBC^*}(id) = 3$ (Theorem 5.77) and $\gamma_{MPC^*}(id) = 4$ (Theorem 5.75). Asymptotically, this gap widens even more as $\overline{\gamma}_{EBC^*}(id) = 2$, but $\overline{\gamma}_{MPC^*}(id) = 4$.

Similar techniques across different QRTs We briefly review various broad techniques that have found applications in multiple different instances of QPS.

Generally, expressing quasiprobability extents as convex programs proved to be very fruitful, both for numerical evaluation and also to analytically prove certain properties. In some cases, the desired quasiprobability extent can directly be expressed as a linear or semidefinite program (e.g. γ_{CPTP^*} for nonphysical simulation, γ_{STAB} for magic states and also in the setting of probabilistic cancellation). Whenever this is not possible, we instead reverted to finding outer approximations of the decomposition set, for which the inclusion does exhibit a semidefinite expression (e.g. PPT for SEP, PPTC for LOCC and CSP for SO).

Generally speaking, evaluating the quasiprobability extent tends to be easier for states than for channels. Interestingly, we observed that the completely resource non-generating channels of our QRTs were characterized as having free Choi states, e.g., SEPC channels have separable Choi states (Section 5.1.1) and CSP channels have stabilizer Choi states (Proposition 6.4). Hence, the relaxation to this larger set (e.g. SEPC instead of LOCC or CSP instead of SO) simplifies the problem of finding the quasiprobability extent of a channel as it is now almost the same task as finding the extent of a state.

Another useful insight is that for any channel $\mathcal E$ with a diagonal Pauli transfer matrix, the quasiprobability extent w.r.t. to the "largest" decomposition set $\gamma_{\text{CPTP}^{\star}}$ can also be achieved by restricting the decomposition set to the Pauli channels (recall Proposition 4.6). Since the Pauli channels are included in virtually every decomposition set we considered (they are local Clifford operations and typically also available in probabilistic error cancellation), the associated quasiprobability extent of $\mathcal E$ is hence also directly known.

8.1 Open questions

We summarize various open questions that we have left unanswered and that could be explored in future work.

General

- We explored QPS applied to the QRTs of entanglement, magic and the "maximal" QRT. What other resource theories can QPS be applied to in a practically meaningful manner? A natural candidate would for example be the resource theory of non-Gaussianity for continuous-variable quantum systems [176]. This QRT mirrors the resource theory of magic, as an analog of the Gottesman-Knill theorem exists for Gaussian information processing [8]. It would be particularly interesting to investigate the implications of switching from a finite-dimensional to an infinite-dimensional Hilbert space.
- It would be interesting to further investigate the properties of the smooth quasiprobability extent. Can any meaningful statements be made for a finite decomposition set (respectively a polytope)? How do the precise details of the smoothing matter?

Nonphysical simulation

- What interesting algorithmic applications (such as the imaginary time evolution in Section 4.3) can be achieved using QPS of nonphysical operations?
- How well can QPS emulate a universal quantum cloner? Does it provide an advantage over approximate CPTP universal quantum cloners?
- How does the overhead of entanglement detection with QPS compare to other approaches?

Nonlocal simulation

- What are suitable computational tasks for which circuit knitting can be used?
- Does there exist a bipartite CPTP map $\mathcal{E} \in \text{CPTP}(A; B \to A'; B')$ for which γ_{LOCC^*} is strictly smaller than γ_{LOCC} ?
- Is there a simple closed-form expression for the quasiprobability extent of arbitrary unitary channels $\mathcal{U} \in \operatorname{CPTP}(AB)$? Does there exist a bipartite unitary channel \mathcal{U} for which classical communication reduces the overhead of space-like

cutting, i.e., $\gamma_{LOCC^*}(\mathcal{U}) \lneq \gamma_{LO^*}(\mathcal{U})$? This question is difficult to address with current techniques, as essentially all of our proposed lower bounds are valid for both γ_{LO^*} and γ_{LOCC^*} , making them useless to demonstrate a separation between the two.

- Is there an adapted de Finetti theorem which can prove the convergence speed of the SDP hierarchy γ_{SEPC_k} introduced in Section 5.1.4.5?
- Our SDP hierarchies for γ_{SEP} and γ_{SEPC} discussed in Sections 5.1.3.2 and 5.1.4.5 are based on the original version of the Doherty-Parrilo-Spedalieri (DPS) hierarchy introduced in Ref. [51]. There exist improved versions of the DPS hierarchy that exhibit better convergence behavior, see for example Ref. [75]. It would be interesting to adapt our hierarchies for the quasiprobability extent with these improvements.
- Can a separation between the regularized and asymptotic quasiprobability extent be proven for arbitrary two-qubit gates, in analogy to the result for states in Corollaries 5.16 and 5.32?
- Can the exact PPT entanglement cost of a channel $E_{C,PPTSC}^{\text{exact}}(\mathcal{E})$ be efficiently evaluated analogously to the state counterpart?
- Can the finite-shot smoothed quasiprobability extent of a channel w.r.t. PPTC be characterized by the smoothed max-logarithmic negativity, to lift Lemma 5.58 to a statement analogous to Lemma 5.35? This would allow us to relate the asymptotic extent $\gamma_{\text{PPTC}}^{\infty}(\mathcal{E})$ with the PPTSC entanglement cost $E_{C,\text{PPTSC}}(\mathcal{E})$.
- Can we find bounds on the quasiprobability extent of the quantum Fourier transform that improve upon the proof of Proposition 5.67?
- Space-like cuts can be straightforwardly generalized to the *n*-partite settings for n > 2. Can the quasiprobability extent of pure states and unitary channels be evaluated in this more general setting?

Magic simulation

- Does approximate quasiprobability simulation with a vanishing error have an asymptotically lower overhead? More concretely, is $\gamma_{\text{STAB}}^{\infty}(|H\rangle\langle H|)$ strictly smaller than $\overline{\gamma}_{\text{STAB}}(|H\rangle\langle H|)$?
- Is there any utility in including classical side information in the QRT of magic, i.e., can γ_{SO^*} be strictly smaller than γ_{SO} for certain channels?

Noise-free simulation

- What is the most resource efficient way to combine QEC with QEM?
- Can classical side information be meaningfully used in some instance of probabilistic error cancellation?

Bibliography

- [1] S. Aaronson and A. Arkhipov, "The computational complexity of linear optics", in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, ser. STOC '11, San Jose, California, USA: Association for Computing Machinery, 2011, pp. 333–342, ISBN: 9781450306911. DOI: 10.1145/1993636. 1993682. Available: https://doi.org/10.1145/1993636.1993682.
- [2] S. Aaronson and D. Gottesman, *Improved simulation of stabilizer circuits, Phys. Rev. A*, vol. 70, p. 052328, 5 Nov. 2004. DOI: 10.1103/PhysRevA.70.052328. Available: https://link.aps.org/doi/10.1103/PhysRevA.70.052328.
- [3] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput., vol. 38, no. 4, pp. 1207–1282, Jul. 2008, ISSN: 0097-5397. DOI: 10.1137/S0097539799359385. Available: https://doi.org/10.1137/S0097539799359385.
- [4] T. Ando, Concavity of certain maps on positive definite matrices and applications to Hadamard products, Linear Algebra and its Applications, vol. 26, pp. 203–241, 1979. DOI: https://doi.org/10.1016/0024-3795(79)90179-4.
- [5] K. Audenaert, M. B. Plenio, and J. Eisert, Entanglement cost under positive-partial-transpose-preserving operations, Phys. Rev. Lett., vol. 90, p. 027901, 2 Jan. 2003. DOI: 10.1103/PhysRevLett.90.027901. Available: https://link.aps.org/doi/10.1103/PhysRevLett.90.027901.
- [6] J. Bae, Designing quantum information processing via structural physical approximation, Reports on Progress in Physics, vol. 80, no. 10, p. 104 001, Sep. 2017. DOI: 10.1088/1361-6633/aa7d45. Available: https://dx.doi.org/10.1088/1361-6633/aa7d45.
- [7] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, *Elementary gates for quantum computation*, *Phys. Rev. A*, vol. 52, pp. 3457–3467, 5 Nov. 1995. DOI: 10.1103/PhysRevA. 52.3457. Available: https://link.aps.org/doi/10.1103/PhysRevA.52.3457.
- [8] S. D. Bartlett, B. C. Sanders, S. L. Braunstein, and K. Nemoto, Efficient classical simulation of continuous variable quantum information processes, Phys. Rev. Lett., vol. 88, p. 097904, 9 Feb. 2002. DOI: 10.1103/PhysRevLett.88.097904. Available: https://link.aps.org/doi/10.1103/PhysRevLett.88.097904.

- [9] M. Bechtold, J. Barzen, F. Leymann, A. Mandl, and F. Truger, *Joint wire cutting with non-maximally entangled states*, 2024. arXiv: 2406.13315 [quant-ph]. Available: https://arxiv.org/abs/2406.13315.
- [10] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, *Concentrating partial entanglement by local operations*, *Phys. Rev. A*, vol. 53, pp. 2046–2052, 4 Apr. 1996. DOI: 10.1103/PhysRevA.53.2046. Available: https://link.aps.org/doi/10.1103/PhysRevA.53.2046.
- [11] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels, Phys. Rev. Lett., vol. 70, pp. 1895–1899, 13 Mar. 1993. DOI: 10.1103/PhysRevLett.70.1895. Available: https://link.aps.org/doi/10.1103/PhysRevLett.70.1895.
- [12] 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, Phys. Rev. A*, vol. 59, pp. 1070–1091, 2 Feb. 1999. DOI: 10.1103/PhysRevA.59.1070. Available: https://link.aps.org/doi/10.1103/PhysRevA.59.1070.
- [13] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal, *Unextendible product bases and bound entanglement, Physical Review Letters*, vol. 82, no. 26, pp. 5385–5388, Jun. 1999, ISSN: 1079-7114. DOI: 10.1103/physrevlett.82.5385. Available: http://dx.doi.org/10.1103/PhysRevLett.82.5385.
- [14] R. S. Bennink, E. M. Ferragut, T. S. Humble, J. A. Laska, J. J. Nutaro, M. G. Pleszkoch, and R. C. Pooser, *Unbiased simulation of near-clifford quantum circuits*, *Phys. Rev. A*, vol. 95, p. 062337, 6 Jun. 2017. DOI: 10.1103/PhysRevA.95.062337. Available: https://link.aps.org/doi/10.1103/PhysRevA.95.062337.
- [15] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme, *Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors, Nature Physics*, vol. 19, no. 8, pp. 1116–1121, May 2023, ISSN: 1745-2481. DOI: 10.1038/s41567-023-02042-2. Available: http://dx.doi.org/10.1038/s41567-023-02042-2.
- [16] M. Berta, F. Borderi, O. Fawzi, and V. B. Scholz, Semidefinite programming hierarchies for constrained bilinear optimization, Mathematical Programming, vol. 194, no. 1–2, pp. 781–829, Apr. 2021, ISSN: 1436-4646. DOI: 10.1007/s10107-021-01650-1. Available: http://dx.doi.org/10.1007/s10107-021-01650-1.
- [17] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, Low-cost error mitigation by symmetry verification, Phys. Rev. A, vol. 98, p. 062339, 6 Dec. 2018. DOI: 10.1103/PhysRevA.98.062339. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.062339.
- [18] F. G. S. L. Brandão and N. Datta, *One-shot rates for entanglement manipulation under non-entangling maps*, *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1754–1760, 2011. DOI: 10.1109/TIT.2011.2104531.

- [19] F. G. S. L. Brandão and G. Gour, Reversible framework for quantum resource theories, Phys. Rev. Lett., vol. 115, p. 070503, 7 Aug. 2015. DOI: 10.1103/PhysRevLett.115.070503. Available: https://link.aps.org/doi/10.1103/PhysRevLett.115.070503.
- [20] F. G. S. L. Brandão and M. B. Plenio, *A generalization of quantum stein's lemma, Communications in Mathematical Physics*, vol. 295, no. 3, pp. 791–828, Feb. 2010, ISSN: 1432-0916. DOI: 10.1007/s00220-010-1005-z. Available: http://dx.doi.org/10.1007/s00220-010-1005-z.
- [21] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, *Simulation of quantum circuits by low-rank stabilizer decompositions, Quantum*, vol. 3, p. 181, Sep. 2019, ISSN: 2521-327X. DOI: 10.22331/q-2019-09-02-181. Available: https://doi.org/10.22331/q-2019-09-02-181.
- [22] S. Bravyi and D. Gosset, Improved classical simulation of quantum circuits dominated by clifford gates, Phys. Rev. Lett., vol. 116, p. 250501, 25 Jun. 2016. DOI: 10.1103/PhysRevLett.116.250501. Available: https://link.aps.org/doi/10.1103/PhysRevLett.116.250501.
- [23] S. Bravyi, D. Gosset, and Y. Liu, How to simulate quantum measurement without computing marginals, Phys. Rev. Lett., vol. 128, p. 220503, 22 Jun. 2022. DOI: 10.1103/PhysRevLett.128.220503. Available: https://link.aps.org/doi/10.1103/PhysRevLett.128.220503.
- [24] S. Bravyi and A. Kitaev, *Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A*, vol. 71, p. 022316, 2 Feb. 2005. DOI: 10.1103/PhysRevA.71.022316. Available: https://link.aps.org/doi/10.1103/PhysRevA.71.022316.
- [25] S. Bravyi, G. Smith, and J. A. Smolin, *Trading classical and quantum computational resources*, *Phys. Rev. X*, vol. 6, p. 021043, 2 Jun. 2016. DOI: 10.1103/PhysRevX.6.021043. Available: https://link.aps.org/doi/10.1103/PhysRevX.6.021043.
- [26] S. Bravyi and A. Vargo, Simulation of rare events in quantum error correction, Phys. Rev. A, vol. 88, p. 062308, 6 Dec. 2013. DOI: 10.1103/PhysRevA.88.062308. Available: https://link.aps.org/doi/10.1103/PhysRevA.88.062308.
- [27] M. J. Bremner, R. Jozsa, and D. J. Shepherd, Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 467, no. 2126, pp. 459–472, 2011. DOI: 10.1098/rspa.2010.0301. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2010.0301. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2010.0301.
- [28] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Average-case complexity versus approximate simulation of commuting quantum computations, Phys. Rev. Lett., vol. 117, p. 080501, 8 Aug. 2016. DOI: 10.1103/PhysRevLett.117.080501. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.080501.

- [29] L. Brenner, C. Piveteau, and D. Sutter, *Optimal wire cutting with classical communication*, 2023. arXiv: 2302.03366 [quant-ph]. Available: https://arxiv.org/abs/2302.03366.
- [30] V. Bu žek, M. Hillery, and R. F. Werner, Optimal manipulations with qubits: Universal-not gate, Phys. Rev. A, vol. 60, R2626–R2629, 4 Oct. 1999. DOI: 10. 1103/PhysRevA.60.R2626. Available: https://link.aps.org/doi/10.1103/PhysRevA.60.R2626.
- [31] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, *Quantum error mitigation, Rev. Mod. Phys.*, vol. 95, p. 045005, 4 Dec. 2023. DOI: 10.1103/RevModPhys.95.045005. Available: https://link.aps.org/doi/10.1103/RevModPhys.95.045005.
- [32] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, Phys. Rev. A, vol. 54, pp. 1098–1105, 2 Aug. 1996. DOI: 10.1103/PhysRevA.54.1098. Available: https://link.aps.org/doi/10.1103/PhysRevA.54.1098.
- [33] E. T. Campbell, Optimal entangling capacity of dynamical processes, Phys. Rev. A, vol. 82, p. 042314, 4 Oct. 2010. DOI: 10.1103/PhysRevA.82.042314. Available: https://link.aps.org/doi/10.1103/PhysRevA.82.042314.
- [34] N. Cao, J. Lin, D. Kribs, Y.-T. Poon, B. Zeng, and R. Laflamme, *Nisq: Error correction, mitigation, and noise simulation*, 2022. arXiv: 2111.02345 [quant-ph]. Available: https://arxiv.org/abs/2111.02345.
- [35] E. Carlen, *Trace Inequalities and Quantum Entropy: An Introductory Course.* Contemporary Mathematics, 2009. DOI: 10.1090/conm/529.
- [36] A. Carrera Vazquez, C. Tornow, D. Ristè, S. Woerner, M. Takita, and D. Egger, *Combining quantum processors with real-time classical communication*, *Nature*, pp. 1–5, Nov. 2024. DOI: 10.1038/s41586-024-08178-2.
- [37] P. J. Cavalcanti, J. H. Selby, J. Sikora, and A. B. Sainz, Decomposing all multipartite non-signalling channels via quasiprobabilistic mixtures of local channels in generalised probabilistic theories, Journal of Physics A: Mathematical and Theoretical, vol. 55, no. 40, p. 404 001, Sep. 2022. DOI: 10.1088/1751-8121/ac8ea4. Available: https://dx.doi.org/10.1088/1751-8121/ac8ea4.
- [38] C. M. Caves, C. A. Fuchs, and R. Schack, *Unknown quantum states: The quantum de Finetti representation, Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4537–4559, Sep. 2002, ISSN: 0022-2488. DOI: 10.1063/1.1494475. eprint: https://pubs.aip.org/aip/jmp/article-pdf/43/9/4537/19182892/4537_1_online.pdf. Available: https://doi.org/10.1063/1.1494475.
- [39] J. Chen, E. Stoudenmire, and S. R. White, Quantum fourier transform has small entanglement, PRX Quantum, vol. 4, p. 040318, 4 Oct. 2023. DOI: 10.1103/PRXQuantum.4.040318. Available: https://link.aps.org/doi/10.1103/PRXQuantum.4.040318.

- [40] A. M. Childs, D. Leung, L. Mančinska, and M. Ozols, *A framework for bounding nonlocality of state discrimination, Communications in Mathematical Physics*, vol. 323, no. 3, pp. 1121–1153, Sep. 2013, ISSN: 1432-0916. DOI: 10.1007/s00220-013-1784-0. Available: http://dx.doi.org/10.1007/s00220-013-1784-0.
- [41] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Transforming quantum operations: Quantum supermaps, Europhysics Letters*, vol. 83, no. 3, p. 30004, Jul. 2008. DOI: 10.1209/0295-5075/83/30004. Available: https://dx.doi.org/10.1209/0295-5075/83/30004.
- [42] E. Chitambar, W. Cui, and H.-K. Lo, Increasing entanglement monotones by separable operations, Phys. Rev. Lett., vol. 108, p. 240504, 24 Jun. 2012. DOI: 10.1103/PhysRevLett.108.240504. Available: https://link.aps.org/doi/10.1103/PhysRevLett.108.240504.
- [43] E. Chitambar and G. Gour, *Quantum resource theories*, *Rev. Mod. Phys.*, vol. 91, p. 025 001, 2 Apr. 2019. DOI: 10.1103/RevModPhys.91.025001. Available: https://link.aps.org/doi/10.1103/RevModPhys.91.025001.
- [44] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, *Everything you always wanted to know about locc (but were afraid to ask)*, *Communications in Mathematical Physics*, vol. 328, no. 1, pp. 303–326, Mar. 2014, ISSN: 1432-0916. DOI: 10.1007/s00220-014-1953-9. Available: http://dx.doi.org/10.1007/s00220-014-1953-9.
- [45] E. Chitambar, J. I. de Vicente, M. W. Girard, and G. Gour, Entanglement manipulation beyond local operations and classical communication, Journal of Mathematical Physics, vol. 61, no. 4, p. 042201, Apr. 2020, ISSN: 0022-2488. DOI: 10.1063/1.5124109. eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.5124109/13959084/042201_1_online.pdf. Available: https://doi.org/10.1063/1.5124109.
- [46] M. Christandl, R. König, G. Mitchison, and R. Renner, *One-and-a-half quantum de finetti theorems, Communications in Mathematical Physics*, vol. 273, no. 2, pp. 473–498, Mar. 2007, ISSN: 1432-0916. DOI: 10.1007/s00220-007-0189-3. Available: http://dx.doi.org/10.1007/s00220-007-0189-3.
- [47] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein, Entangling operations and their implementation using a small amount of entanglement, Phys. Rev. Lett., vol. 86, pp. 544–547, 3 Jan. 2001. DOI: 10.1103/PhysRevLett.86.544. Available: https://link.aps.org/doi/10.1103/PhysRevLett.86.544.
- [48] W. Cui, E. Chitambar, and H.-K. Lo, Randomly distilling W-class states into general configurations of two-party entanglement, Phys. Rev. A, vol. 84, p. 052301, 5 Nov. 2011. DOI: 10.1103/PhysRevA.84.052301. Available: https://link.aps.org/doi/10.1103/PhysRevA.84.052301.
- [49] N. Datta, Max-relative entropy of entanglement, alias log robustness, International Journal of Quantum Information, vol. 07, no. 02, pp. 475–491, 2009. DOI: 10.1142/S0219749909005298. eprint: https://doi.org/10.1142/S0219749909005298. Available: https://doi.org/10.1142/S0219749909005298.

- [50] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, *Topological quantum memory, Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4452–4505, 2002. DOI: 10. 1063/1.1499754. eprint: https://doi.org/10.1063/1.1499754. Available: https://doi.org/10.1063/1.1499754.
- [51] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Complete family of separability criteria, Phys. Rev. A, vol. 69, p. 022308, 2 Feb. 2004. DOI: 10.1103/PhysRevA. 69.022308. Available: https://link.aps.org/doi/10.1103/PhysRevA.69.022308.
- [52] M. J. Donald, M. Horodecki, and O. Rudolph, The uniqueness theorem for entanglement measures, Journal of Mathematical Physics, vol. 43, no. 9, pp. 4252–4272, Sep. 2002, ISSN: 1089-7658. DOI: 10.1063/1.1495917. Available: http://dx.doi.org/10.1063/1.1495917.
- [53] T. Dubach, "Towards optimal circuit knitting", Master's thesis, ETH Zurich, 2023.
- [54] A. Dutkiewicz, S. Polla, M. Scheurer, C. Gogolin, W. J. Huggins, and T. E. O'Brien, *Error mitigation and circuit division for early fault-tolerant quantum phase estimation*, 2024. arXiv: 2410.05369 [quant-ph]. Available: https://arxiv.org/abs/2410.05369.
- [55] B. Eastin and E. Knill, Restrictions on transversal encoded quantum gate sets, Phys. Rev. Lett., vol. 102, p. 110502, 11 Mar. 2009. DOI: 10.1103/PhysRevLett.102. 110502. Available: https://link.aps.org/doi/10.1103/PhysRevLett.102. 110502.
- [56] J. Emerson, R. Alicki, and K. Życzkowski, *Scalable noise estimation with random unitary operators, Journal of Optics B: Quantum and Semiclassical Optics*, vol. 7, no. 10, S347, Sep. 2005. DOI: 10.1088/1464-4266/7/10/021. Available: https://dx.doi.org/10.1088/1464-4266/7/10/021.
- [57] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X, vol. 8, p. 031027, 3 Jul. 2018. DOI: 10.1103/PhysRevX.8.031027. Available: https://link.aps.org/doi/10.1103/PhysRevX.8.031027.
- [58] S. J. van Enk, Relations between cloning and the universal not derived from conservation laws, Phys. Rev. Lett., vol. 95, p. 010502, 1 Jun. 2005. DOI: 10.1103/PhysRevLett.95.010502. Available: https://link.aps.org/doi/10.1103/PhysRevLett.95.010502.
- [59] S. T. Flammia and J. J. Wallman, *Efficient estimation of pauli channels, ACM Transactions on Quantum Computing*, vol. 1, no. 1, Dec. 2020. DOI: 10.1145/3408039. Available: https://doi.org/10.1145/3408039.
- [60] A. G. Fowler, A. M. Stephens, and P. Groszkowski, *High-threshold universal quantum computation on the surface code*, *Phys. Rev. A*, vol. 80, p. 052312, 5 Nov. 2009. DOI: 10.1103/PhysRevA.80.052312. Available: https://link.aps.org/doi/10.1103/PhysRevA.80.052312.

- [61] S. Gharibian, Strong np-hardness of the quantum separability problem, Quantum Info. Comput., vol. 10, no. 3, pp. 343–360, Mar. 2010, ISSN: 1533-7146.
- [62] A. Gonzales, A. M. Babu, J. Liu, Z. Saleem, and M. Byrd, Fault tolerant quantum error mitigation, 2024. arXiv: 2308.05403 [quant-ph]. Available: https://arxiv.org/abs/2308.05403.
- [63] D. Gottesman, The heisenberg representation of quantum computers, 1998. DOI: 10.48550/arXiv.quant-ph/9807006. arXiv:quant-ph/9807006 [quant-ph].
- [64] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, Nature, vol. 402, no. 6760, pp. 390–393, Nov. 1999, ISSN: 1476-4687. DOI: 10.1038/46503. Available: http://dx.doi.org/10.1038/46503.
- [65] G. Gour, Resources of the quantum world, 2024. DOI: 10.48550/arXiv.2402. 05474. arXiv: 2402.05474 [quant-ph].
- [66] G. Gour and C. M. Scandolo, *Dynamical resources*, 2020. DOI: 10.48550/arXiv. 2101.01552. arXiv: 2101.01552 [quant-ph].
- [67] G. Gour and C. M. Scandolo, Entanglement of a bipartite channel, Phys. Rev. A, vol. 103, p. 062422, 6 Jun. 2021. DOI: 10.1103/PhysRevA.103.062422. Available: https://link.aps.org/doi/10.1103/PhysRevA.103.062422.
- [68] D. Greenbaum, *Introduction to quantum gate set tomography*, 2015. arXiv: 1509. 02921 [quant-ph]. Available: https://arxiv.org/abs/1509.02921.
- [69] D. Gross, Hudson's theorem for finite-dimensional quantum systems, Journal of Mathematical Physics, vol. 47, no. 12, p. 122107, Dec. 2006, ISSN: 0022-2488. DOI: 10.1063/1.2393152. eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.2393152/16702548/122107_1_online.pdf. Available: https://doi.org/10.1063/1.2393152.
- [70] L. Gurvits, "Classical deterministic complexity of edmonds' problem and quantum entanglement", in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, ser. STOC '03, San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 10–19, ISBN: 1581136749. DOI: 10.1145/780542.780545. Available: https://doi.org/10.1145/780542.780545.
- [71] G. Gutoski, *Properties of local quantum operations with shared entanglement*, *Quantum Info. Comput.*, vol. 9, no. 9, pp. 739–764, Sep. 2009, ISSN: 1533-7146.
- [72] H. Hamaguchi, K. Hamada, and N. Yoshioka, *Handbook for Quantifying Robustness of Magic*, *Quantum*, vol. 8, p. 1461, Sep. 2024, ISSN: 2521-327X. DOI: 10.22331/q-2024-09-05-1461. Available: https://doi.org/10.22331/q-2024-09-05-1461.
- [73] H. Harada, K. Wada, and N. Yamamoto, Doubly optimal parallel wire cutting without ancilla qubits, PRX Quantum, vol. 5, p. 040308, 4 Oct. 2024. DOI: 10. 1103/PRXQuantum.5.040308. Available: https://link.aps.org/doi/10.1103/PRXQuantum.5.040308.

- [74] A. W. Harrow and A. Lowe, *Optimal quantum circuit cuts with application to clustered hamiltonian simulation*, 2024. arXiv: 2403.01018 [quant-ph]. Available: https://arxiv.org/abs/2403.01018.
- [75] A. W. Harrow, A. Natarajan, and X. Wu, *An improved semidefinite programming hierarchy for testing entanglement, Communications in Mathematical Physics*, vol. 352, no. 3, pp. 881–904, Mar. 2017, ISSN: 1432-0916. DOI: 10.1007/s00220-017-2859-0. Available: http://dx.doi.org/10.1007/s00220-017-2859-0.
- [76] A. W. Harrow and M. A. Nielsen, Robustness of quantum gates in the presence of noise, Phys. Rev. A, vol. 68, p. 012308, 1 Jul. 2003. DOI: 10.1103/PhysRevA. 68.012308. Available: https://link.aps.org/doi/10.1103/PhysRevA.68.012308.
- [77] P. M. Hayden, M. Horodecki, and B. M. Terhal, *The asymptotic entanglement cost of preparing a quantum state, Journal of Physics A: Mathematical and General*, vol. 34, no. 35, pp. 6891–6898, Aug. 2001, ISSN: 1361-6447. DOI: 10.1088/0305-4470/34/35/314. Available: http://dx.doi.org/10.1088/0305-4470/34/35/314.
- [78] A. Heimendahl, M. Heinrich, and D. Gross, The axiomatic and the operational approaches to resource theories of magic do not coincide, Journal of Mathematical Physics, vol. 63, no. 11, p. 112 201, Nov. 2022, ISSN: 0022-2488. DOI: 10.1063/5.0085774. eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/5.0085774/16591566/112201_1_online.pdf. Available: https://doi.org/10.1063/5.0085774.
- [79] M. Heinrich and D. Gross, *Robustness of Magic and Symmetries of the Stabiliser Polytope*, *Quantum*, vol. 3, p. 132, Apr. 2019, ISSN: 2521-327X. DOI: 10.22331/q-2019-04-08-132. Available: https://doi.org/10.22331/q-2019-04-08-132.
- [80] M. Horodecki, P. W. Shor, and M. B. Ruskai, Entanglement breaking channels, Reviews in Mathematical Physics, vol. 15, no. 06, pp. 629–641, 2003. DOI: 10.1142/S0129055X03001709. eprint: https://doi.org/10.1142/S0129055X03001709. Available: https://doi.org/10.1142/S0129055X03001709.
- [81] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: Necessary and sufficient conditions, Physics Letters A, vol. 223, no. 1, pp. 1–8, 1996, ISSN: 0375-9601. DOI: https://doi.org/10.1016/S0375-9601(96)00706-2. Available: https://www.sciencedirect.com/science/article/pii/S0375960196007062.
- [82] M. Horodecki, P. Horodecki, and R. Horodecki, Limits for entanglement measures, Phys. Rev. Lett., vol. 84, pp. 2014–2017, 9 Feb. 2000. DOI: 10.1103/PhysRevLett.84.2014. Available: https://link.aps.org/doi/10.1103/PhysRevLett.84.2014.
- [83] P. Horodecki, From limits of quantum operations to multicopy entanglement witnesses and state-spectrum estimation, Phys. Rev. A, vol. 68, p. 052101, 5 Nov. 2003. DOI: 10.1103/PhysRevA.68.052101. Available: https://link.aps.org/doi/10.1103/PhysRevA.68.052101.

- [84] P. Horodecki and A. Ekert, Method for direct detection of quantum entanglement, Phys. Rev. Lett., vol. 89, p. 127902, 12 Aug. 2002. DOI: 10.1103/PhysRevLett. 89.127902. Available: https://link.aps.org/doi/10.1103/PhysRevLett. 89.127902.
- [85] C. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, Surface code quantum computing by lattice surgery, New Journal of Physics, vol. 14, no. 12, p. 123011, Dec. 2012. DOI: 10.1088/1367-2630/14/12/123011. Available: https://doi.org/10.1088/1367-2630/14/12/123011.
- [86] M. Howard and E. Campbell, Application of a resource theory for magic states to fault-tolerant quantum computing, Phys. Rev. Lett., vol. 118, p. 090501, 9 Mar. 2017. DOI: 10.1103/PhysRevLett.118.090501. Available: https://link.aps.org/doi/10.1103/PhysRevLett.118.090501.
- [87] A. Jenčová, Base norms and discrimination of generalized quantum channels, Journal of Mathematical Physics, vol. 55, no. 2, p. 022 201, Feb. 2014, ISSN: 0022-2488. DOI: 10.1063/1.4863715. eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.4863715/15640731/022201_1_online.pdf. Available: https://doi.org/10.1063/1.4863715.
- [88] J. Jiang, K. Wang, and X. Wang, *Physical Implementability of Linear Maps and Its Application in Error Mitigation*, *Quantum*, vol. 5, p. 600, Dec. 2021, ISSN: 2521-327X. DOI: 10.22331/q-2021-12-07-600. Available: https://doi.org/10.22331/q-2021-12-07-600.
- [89] M. Jing, C. Zhu, and X. Wang, Circuit knitting faces exponential sampling overhead scaling bounded by entanglement cost, 2024. arXiv: 2404.03619 [quant-ph]. Available: https://arxiv.org/abs/2404.03619.
- [90] R. Jozsa and A. Miyake, Matchgates and classical simulation of quantum circuits, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 464, no. 2100, pp. 3089–3106, 2008. DOI: 10.1098/rspa.2008.0189. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2008.0189. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2008.0189.
- [91] A. Kenfack and K. Życzkowski, Negativity of the wigner function as an indicator of non-classicality, Journal of Optics B: Quantum and Semiclassical Optics, vol. 6, no. 10, p. 396, Aug. 2004. DOI: 10.1088/1464-4266/6/10/003. Available: https://dx.doi.org/10.1088/1464-4266/6/10/003.
- [92] N. Khaneja and S. Glaser, Cartan decomposition of su(2ⁿ), constructive controllability of spin systems and universal quantum computing, 2000. arXiv: quant-ph/0010100 [quant-ph]. Available: https://arxiv.org/abs/quant-ph/0010100.
- [93] H.-J. Kim, S. Lee, L. Lami, and M. B. Plenio, *One-shot manipulation of entanglement for quantum channels, IEEE Transactions on Information Theory*, vol. 67, no. 8, pp. 5339–5351, 2021. DOI: 10.1109/TIT.2021.3079938.

- [94] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. P. Zaletel, K. Temme, and A. Kandala, *Evidence for the utility of quantum computing before fault tolerance, Nature*, vol. 618, pp. 500–505, 2023. DOI: https://doi.org/10.1038/s41586-023-06096-3.
- [95] A. Y. Kitaev, Quantum measurements and the abelian stabilizer problem, 1995. arXiv: quant-ph/9511026 [quant-ph]. Available: https://arxiv.org/abs/quant-ph/9511026.
- [96] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation: Error models and thresholds, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1969, pp. 365–384, Jan. 1998, ISSN: 1471-2946. DOI: 10.1098/rspa.1998.0166. Available: http://dx.doi.org/10.1098/rspa.1998.0166.
- [97] M. Koashi, F. Takenaga, T. Yamamoto, and N. Imoto, *Quantum nonlocality with-out entanglement in a pair of qubits*, 2007. arXiv: 0709.3196 [quant-ph]. Available: https://arxiv.org/abs/0709.3196.
- [98] J. K. Korbicz, M. L. Almeida, J. Bae, M. Lewenstein, and A. Acín, *Structural approximations to positive maps and entanglement-breaking channels, Phys. Rev. A*, vol. 78, p. 062 105, 6 Dec. 2008. DOI: 10.1103/PhysRevA.78.062105. Available: https://link.aps.org/doi/10.1103/PhysRevA.78.062105.
- [99] L. Lami, F. A. Mele, and B. Regula, *Computable entanglement cost*, 2024. arXiv: 2405.09613 [quant-ph]. Available: https://arxiv.org/abs/2405.09613.
- [100] L. Lami and B. Regula, Computable lower bounds on the entanglement cost of quantum channels, Journal of Physics A: Mathematical and Theoretical, vol. 56, no. 3, p. 035 302, Feb. 2023. DOI: 10.1088/1751-8121/aca731. Available: https://dx.doi.org/10.1088/1751-8121/aca731.
- [101] Y. Li, A magic state's fidelity can be superior to the operations that created it, New Journal of Physics, vol. 17, no. 2, p. 023 037, Feb. 2015. DOI: 10.1088/1367-2630/17/2/023037. Available: https://doi.org/10.1088%2F1367-2630%2F17%2F2%2F023037.
- [102] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X, vol. 7, p. 021 050, 2 Jun. 2017. DOI: 10. 1103/PhysRevX.7.021050. Available: https://link.aps.org/doi/10.1103/PhysRevX.7.021050.
- [103] Z.-W. Liu, K. Bu, and R. Takagi, One-shot operational quantum resource theory, Phys. Rev. Lett., vol. 123, p. 020401, 2 Jul. 2019. DOI: 10.1103/PhysRevLett. 123.020401. Available: https://link.aps.org/doi/10.1103/PhysRevLett. 123.020401.
- [104] J. Łodyga, P. Mazurek, A. Grudka, and M. Horodecki, Simple scheme for encoding and decoding a qubit in unknown state for various topological codes, Scientific Reports, vol. 5, no. 1, p. 8975, 2015. DOI: 10.1038/srep08975. Available: https://doi.org/10.1038/srep08975.

- [105] M. Lostaglio and A. Ciani, Error mitigation and quantum-assisted simulation in the error corrected regime, Phys. Rev. Lett., vol. 127, p. 200506, 20 Nov. 2021. DOI: 10.1103/PhysRevLett.127.200506. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.200506.
- [106] A. Lowe, M. Medvidović, A. Hayes, L. J. O'Riordan, T. R. Bromley, J. M. Arrazola, and N. Killoran, *Fast quantum circuit cutting with randomized measurements, Quantum,* vol. 7, p. 934, Mar. 2023, ISSN: 2521-327X. DOI: 10.22331/q-2023-03-02-934. Available: https://doi.org/10.22331/q-2023-03-02-934.
- [107] A. Mari and J. Eisert, Positive wigner functions render classical simulation of quantum computation efficient, Phys. Rev. Lett., vol. 109, p. 230503, 23 Dec. 2012. DOI: 10.1103/PhysRevLett.109.230503. Available: https://link.aps.org/doi/10.1103/PhysRevLett.109.230503.
- [108] J. Matoušek and J. Vondrák, *The probabilistic method lecture notes*, Mar. 2008. Available: https://www.cs.cmu.edu/~15850/handouts/matousek-vondrak-prob-ln.pdf.
- [109] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, Phys. Rev. Lett., vol. 122, p. 180501, 18 May 2019. DOI: 10.1103/PhysRevLett.122.180501. Available: https://link.aps.org/doi/10.1103/PhysRevLett.122.180501.
- [110] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, *Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states*, *Phys. Rev. A*, vol. 95, p. 042308, 4 Apr. 2017. DOI: 10.1103/PhysRevA. 95.042308. Available: https://link.aps.org/doi/10.1103/PhysRevA.95.042308.
- [111] K. Mitarai and K. Fujii, Constructing a virtual two-qubit gate by sampling single-qubit operations, New Journal of Physics, vol. 23, no. 2, p. 023 021, Feb. 2021. DOI: 10.1088/1367-2630/abd7bc. Available: https://dx.doi.org/10.1088/1367-2630/abd7bc.
- [112] K. Mitarai and K. Fujii, Overhead for simulating a non-local channel with local channels by quasiprobability sampling, Quantum, vol. 5, p. 388, Jan. 2021, ISSN: 2521-327X. DOI: 10.22331/q-2021-01-28-388. Available: https://doi.org/10.22331/q-2021-01-28-388.
- [113] M. Navascués, M. Owari, and M. B. Plenio, Complete criterion for separability detection, Phys. Rev. Lett., vol. 103, p. 160404, 16 Oct. 2009. DOI: 10.1103/PhysRevLett.103.160404. Available: https://link.aps.org/doi/10.1103/PhysRevLett.103.160404.
- [114] M. Navascués, M. Owari, and M. B. Plenio, Power of symmetric extensions for entanglement detection, Phys. Rev. A, vol. 80, p. 052306, 5 Nov. 2009. DOI: 10. 1103/PhysRevA.80.052306. Available: https://link.aps.org/doi/10.1103/PhysRevA.80.052306.

- [115] M. A. Nielsen, "Quantum information theory", PhD thesis, The University of New Mexico, Albuquerque, New Mexico, USA, Dec. 1998. DOI: 10.48550/arXiv.quant-ph/0011036. Available: https://arxiv.org/abs/quant-ph/0011036.
- [116] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press, 2010.
- [117] M. A. Nielsen, C. M. Dawson, J. L. Dodd, A. Gilchrist, D. Mortimer, T. J. Osborne, M. J. Bremner, A. W. Harrow, and A. Hines, *Quantum dynamics as a physical resource*, *Phys. Rev. A*, vol. 67, p. 052301, 5 May 2003. DOI: 10.1103/PhysRevA.67.052301. Available: https://link.aps.org/doi/10.1103/PhysRevA.67.052301.
- [118] A. Paetznick and B. W. Reichardt, *Universal fault-tolerant quantum computation with only transversal gates and error correction*, *Phys. Rev. Lett.*, vol. 111, p. 090505, 9 Aug. 2013. DOI: 10.1103/PhysRevLett.111.090505. Available: https://link.aps.org/doi/10.1103/PhysRevLett.111.090505.
- [119] H. Pashayan, S. D. Bartlett, and D. Gross, From estimation of quantum probabilities to simulation of quantum circuits, Quantum, vol. 4, p. 223, Jan. 2020, ISSN: 2521-327X. DOI: 10.22331/q-2020-01-13-223. Available: https://doi.org/10.22331/q-2020-01-13-223.
- [120] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating outcome probabilities of quantum circuits using quasiprobabilities, Phys. Rev. Lett., vol. 115, p. 070501, 7 Aug. 2015. DOI: 10.1103/PhysRevLett.115.070501. Available: https://link.aps.org/doi/10.1103/PhysRevLett.115.070501.
- [121] P. Pechukas, Reduced dynamics need not be completely positive, Phys. Rev. Lett., vol. 73, pp. 1060–1062, 8 Aug. 1994. DOI: 10.1103/PhysRevLett.73.1060. Available: https://link.aps.org/doi/10.1103/PhysRevLett.73.1060.
- [122] E. Pednault, An alternative approach to optimal wire cutting without ancilla qubits, 2023. arXiv: 2303.08287 [quant-ph]. Available: https://arxiv.org/abs/2303.08287.
- [123] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, Simulating large quantum circuits on a small quantum computer, Phys. Rev. Lett., vol. 125, p. 150504, 15 Oct. 2020. DOI: 10.1103/PhysRevLett.125.150504. Available: https://link.aps.org/doi/10.1103/PhysRevLett.125.150504.
- [124] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett., vol. 77, pp. 1413–1415, 8 Aug. 1996. DOI: 10.1103/PhysRevLett.77.1413. Available: https://link.aps.org/doi/10.1103/PhysRevLett.77.1413.
- [125] H. D. Pfister, C. Piveteau, J. M. Renes, and N. Rengaswamy, *Belief propagation for classical and quantum systems: Overview and recent results, IEEE BITS the Information Theory Magazine*, vol. 2, no. 3, pp. 20–32, 2022. doi: 10.1109/MBITS. 2023.3285848.

- [126] C. Piveteau, "Advanced methods for quasiprobabilistic quantum error mitigation", en, Master Thesis, ETH Zurich, Zurich, 2020-09. DOI: 10.3929/ethz-b-000504508.
- [127] C. Piveteau, C. T. Chubb, and J. M. Renes, Tensor-network decoding beyond 2d, PRX Quantum, vol. 5, p. 040303, 4 Oct. 2024. DOI: 10.1103/PRXQuantum.5.040303. Available: https://link.aps.org/doi/10.1103/PRXQuantum.5.040303.
- [128] C. Piveteau and J. M. Renes, *Quantum message-passing algorithm for optimal and efficient decoding*, *Quantum*, vol. 6, p. 784, Aug. 2022, ISSN: 2521-327X. DOI: 10. 22331/q-2022-08-23-784. Available: https://doi.org/10.22331/q-2022-08-23-784.
- [129] C. Piveteau and D. Sutter, *Circuit knitting with classical communication, IEEE Transactions on Information Theory*, vol. 70, no. 4, pp. 2734–2745, 2024. DOI: 10. 1109/TIT.2023.3310797.
- [130] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, Error mitigation for universal gates on encoded qubits, Phys. Rev. Lett., vol. 127, p. 200505, 20 Nov. 2021. DOI: 10.1103/PhysRevLett.127.200505. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.200505.
- [131] C. Piveteau, D. Sutter, and S. Woerner, *Quasiprobability decompositions with reduced sampling overhead, npj Quantum Information,* vol. 8, no. 1, Feb. 2022, ISSN: 2056-6387. DOI: 10.1038/s41534-022-00517-3. Available: http://dx.doi.org/10.1038/s41534-022-00517-3.
- [132] M. B. Plenio, Logarithmic negativity: A full entanglement monotone that is not convex, Phys. Rev. Lett., vol. 95, p. 090503, 9 Aug. 2005. DOI: 10.1103/PhysRevLett.95.090503. Available: https://link.aps.org/doi/10.1103/PhysRevLett.95.090503.
- [133] H. Qassim, H. Pashayan, and D. Gosset, *Improved upper bounds on the stabilizer rank of magic states*, *Quantum*, vol. 5, p. 606, Dec. 2021, ISSN: 2521-327X. DOI: 10.22331/q-2021-12-20-606. Available: https://doi.org/10.22331/q-2021-12-20-606.
- [134] A. Ray, E. Swaroop, N. Cao, M. Vasmer, and A. Chowdhury, *In preparation*, pivate communication, 2025.
- [135] B. Regula, Convex geometry of quantum resource quantification, Journal of Physics A: Mathematical and Theoretical, vol. 51, no. 4, p. 045303, Dec. 2017. DOI: 10. 1088/1751-8121/aa9100. Available: https://dx.doi.org/10.1088/1751-8121/aa9100.
- [136] B. Regula, R. Takagi, and M. Gu, Operational applications of the diamond norm and related measures in quantifying the non-physicality of quantum maps, Quantum, vol. 5, p. 522, Aug. 2021, ISSN: 2521-327X. DOI: 10.22331/q-2021-08-09-522. Available: https://doi.org/10.22331/q-2021-08-09-522.

- [137] D. Rosset, F. Buscemi, and Y.-C. Liang, Resource theory of quantum memories and their faithful verification with minimal assumptions, Phys. Rev. X, vol. 8, p. 021 033, 2 May 2018. DOI: 10.1103/PhysRevX.8.021033. Available: https://link.aps.org/doi/10.1103/PhysRevX.8.021033.
- [138] V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, *Quantum cloning, Rev. Mod. Phys.*, vol. 77, pp. 1225–1256, 4 Nov. 2005. DOI: 10.1103/RevModPhys.77.1225. Available: https://link.aps.org/doi/10.1103/RevModPhys.77.1225.
- [139] J. Schluck, G. Murta, H. Kampermann, D. Bruß, and N. Wyderka, *Continuity of robustness measures in quantum resource theories, Journal of Physics A: Mathematical and Theoretical*, vol. 56, no. 25, p. 255303, Jun. 2023. DOI: 10.1088/1751-8121/acd500. Available: https://dx.doi.org/10.1088/1751-8121/acd500.
- [140] L. Schmitt, C. Piveteau, and D. Sutter, *Cutting circuits with multiple two-qubit unitaries*, 2024. arXiv: 2312.11638 [quant-ph]. Available: https://arxiv.org/abs/2312.11638.
- [141] J. R. Seddon and E. T. Campbell, Quantifying magic for multi-qubit operations, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 475, no. 2227, p. 20190251, 2019. DOI: 10.1098/rspa.2019.0251. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2019.0251. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2019.0251.
- [142] J. R. Seddon, B. Regula, H. Pashayan, Y. Ouyang, and E. T. Campbell, Quantifying quantum speedups: Improved classical simulation from tighter magic monotones, PRX Quantum, vol. 2, p. 010345, 1 Mar. 2021. DOI: 10.1103/PRXQuantum.2.010345. Available: https://link.aps.org/doi/10.1103/PRXQuantum.2.010345.
- [143] A. Shaji and E. Sudarshan, Who's afraid of not completely positive maps?, Physics Letters A, vol. 341, no. 1, pp. 48–54, 2005, ISSN: 0375-9601. DOI: https://doi.org/10.1016/j.physleta.2005.04.029. Available: https://www.sciencedirect.com/science/article/pii/S0375960105005748.
- [144] P. Shor, "Fault-tolerant quantum computation", in *Proceedings of 37th Conference on Foundations of Computer Science*, 1996, pp. 56–65. DOI: 10.1109/SFCS. 1996.548464.
- [145] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A, vol. 52, R2493-R2496, 4 Oct. 1995. DOI: 10.1103/PhysRevA.52.R2493. Available: https://link.aps.org/doi/10.1103/PhysRevA.52.R2493.
- [146] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM Journal on Computing,* vol. 26, no. 5, pp. 1484–1509, 1997. DOI: 10.1137/S0097539795293172. eprint: https://doi.org/10.1137/S0097539795293172. Available: https://doi.org/10.1137/S0097539795293172.

- [147] T. Singal, C. Chiang, E. Hsu, E. Kim, H.-S. Goan, and M.-H. Hsieh, *Counting stabiliser codes for arbitrary dimension*, *Quantum*, vol. 7, p. 1048, Jul. 2023, ISSN: 2521-327X. DOI: 10.22331/q-2023-07-06-1048. Available: https://doi.org/10.22331/q-2023-07-06-1048.
- [148] A. P. Singh, K. Mitarai, Y. Suzuki, K. Heya, Y. Tabuchi, K. Fujii, and Y. Nakamura, Experimental demonstration of a high-fidelity virtual two-qubit gate, Phys. Rev. Res., vol. 6, p. 013235, 1 Mar. 2024. DOI: 10.1103/PhysRevResearch. 6.013235. Available: https://link.aps.org/doi/10.1103/PhysRevResearch. 6.013235.
- [149] A. M. Steane, Error correcting codes in quantum theory, Phys. Rev. Lett., vol. 77, pp. 793-797, 5 Jul. 1996. DOI: 10.1103/PhysRevLett.77.793. Available: https://link.aps.org/doi/10.1103/PhysRevLett.77.793.
- [150] A. Steane, Multiple-particle interference and quantum error correction, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 452, no. 1954, pp. 2551–2577, 1996. DOI: 10.1098/rspa.1996.0136. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1996.0136. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1996.0136.
- [151] M. Steiner, Generalized robustness of entanglement, Phys. Rev. A, vol. 67, p. 054305, 5 May 2003. DOI: 10.1103/PhysRevA.67.054305. Available: https://link.aps.org/doi/10.1103/PhysRevA.67.054305.
- [152] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error reduction technique: Applications from the nisq to the fault-tolerant quantum computing eras, PRX Quantum, vol. 3, p. 010345, 1 Mar. 2022. DOI: 10.1103/PRXQuantum.3.010345. Available: https://link.aps.org/doi/10.1103/PRXQuantum.3.010345.
- [153] R. Takagi, Optimal resource cost for error mitigation, Phys. Rev. Res., vol. 3, p. 033178, 3 Aug. 2021. DOI: 10.1103/PhysRevResearch.3.033178. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.3.033178.
- [154] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett., vol. 119, p. 180509, 18 Nov. 2017. DOI: 10.1103/PhysRevLett.119.180509. Available: https://link.aps.org/doi/10.1103/PhysRevLett.119.180509.
- [155] B. M. Terhal and D. P. DiVincenzo, Classical simulation of noninteracting-fermion quantum circuits, Phys. Rev. A, vol. 65, p. 032325, 3 Mar. 2002. DOI: 10.1103/PhysRevA.65.032325. Available: https://link.aps.org/doi/10.1103/PhysRevA.65.032325.
- [156] K. Tsubouchi, Y. Mitsuhashi, K. Sharma, and N. Yoshioka, *Symmetric clifford twirling for cost-optimal quantum error mitigation in early ftqc regime*, 2024. arXiv: 2405.07720 [quant-ph]. Available: https://arxiv.org/abs/2405.07720.
- [157] R. R. Tucci, An introduction to cartan's kak decomposition for qc programmers, 2005. arXiv: quant-ph/0507171 [quant-ph]. Available: https://arxiv.org/abs/quant-ph/0507171.

- [158] J. Tyson, Operator-schmidt decomposition of the quantum fourier transform on $\mathbb{C}^{N_1} \otimes \mathbb{C}^{N_2}$, Journal of Physics A: Mathematical and General, vol. 36, no. 24, p. 6813, Jun. 2003. DOI: 10.1088/0305-4470/36/24/317. Available: https://dx.doi.org/10.1088/0305-4470/36/24/317.
- [159] C. Ufrecht, L. S. Herzog, D. D. Scherer, M. Periyasamy, S. Rietsch, A. Plinge, and C. Mutschler, *Optimal joint cutting of two-qubit rotation gates, Phys. Rev. A*, vol. 109, p. 052440, 5 May 2024. DOI: 10.1103/PhysRevA.109.052440. Available: https://link.aps.org/doi/10.1103/PhysRevA.109.052440.
- [160] L. G. Valiant, "Quantum computers that can be simulated classically in polynomial time", in *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, ser. STOC '01, Hersonissos, Greece: Association for Computing Machinery, 2001, pp. 114–123, ISBN: 1581133499. DOI: 10.1145/380752. 380785. Available: https://doi.org/10.1145/380752.380785.
- [161] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, *Negative quasi-probability as a resource for quantum computation*, *New Journal of Physics*, vol. 14, no. 11, p. 113011, Nov. 2012. DOI: 10.1088/1367-2630/14/11/113011. Available: https://dx.doi.org/10.1088/1367-2630/14/11/113011.
- [162] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, *The resource theory of stabilizer quantum computation, New Journal of Physics*, vol. 16, no. 1, p. 013 009, Jan. 2014. DOI: 10.1088/1367-2630/16/1/013009. Available: https://dx.doi.org/10.1088/1367-2630/16/1/013009.
- [163] G. Vidal, W. Dür, and J. I. Cirac, Entanglement cost of bipartite mixed states, Phys. Rev. Lett., vol. 89, p. 027901, 2 Jun. 2002. DOI: 10.1103/PhysRevLett.89.027901. Available: https://link.aps.org/doi/10.1103/PhysRevLett.89.027901
- [164] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A, vol. 65, p. 032314, 3 Feb. 2002. DOI: 10.1103/PhysRevA.65.032314. Available: https://link.aps.org/doi/10.1103/PhysRevA.65.032314.
- [165] G. Vidal, Efficient classical simulation of slightly entangled quantum computations, Phys. Rev. Lett., vol. 91, p. 147902, 14 Oct. 2003. DOI: 10.1103/PhysRevLett. 91.147902. Available: https://link.aps.org/doi/10.1103/PhysRevLett. 91.147902.
- [166] G. Vidal and R. Tarrach, Robustness of entanglement, Phys. Rev. A, vol. 59, pp. 141–155, 1 Jan. 1999. DOI: 10.1103/PhysRevA.59.141. Available: https://link.aps.org/doi/10.1103/PhysRevA.59.141.
- [167] M. A. Wahl, A. Mari, N. Shammah, W. J. Zeng, and G. S. Ravi, "Zero noise extrapolation on logical qubits by scaling the error correction code distance", in 2023 IEEE International Conference on Quantum Computing and Engineering (QCE), vol. 01, 2023, pp. 888–897. DOI: 10.1109/QCE57702.2023.00103.
- [168] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A, vol. 94, p. 052325, 5 Nov. 2016. DOI: 10.1103/PhysRevA.94.052325. Available: https://link.aps.org/doi/10.1103/PhysRevA.94.052325.

- [169] X. Wang and M. Wilde, Errata for "cost of quantum entanglement simplified" and "exact entanglement cost of quantum states and channels under ppt-preserving operations", Apr. 2024. DOI: 10.5281/zenodo.11061607. Available: https://doi.org/10.5281/zenodo.11061607.
- [170] X. Wang and M. M. Wilde, α-logarithmic negativity, Phys. Rev. A, vol. 102, p. 032416, 3 Sep. 2020. DOI: 10.1103/PhysRevA.102.032416. Available: https://link.aps.org/doi/10.1103/PhysRevA.102.032416.
- [171] X. Wang and M. M. Wilde, Cost of quantum entanglement simplified, Phys. Rev. Lett., vol. 125, p. 040502, 4 Jul. 2020. DOI: 10.1103/PhysRevLett.125.040502. Available: https://link.aps.org/doi/10.1103/PhysRevLett.125.040502.
- [172] X. Wang and M. M. Wilde, Exact entanglement cost of quantum states and channels under positive-partial-transpose-preserving operations, Phys. Rev. A, vol. 107, p. 012429, 1 Jan. 2023. DOI: 10.1103/PhysRevA.107.012429. Available: https://link.aps.org/doi/10.1103/PhysRevA.107.012429.
- [173] J. Watrous, Semidefinite programs for completely bounded norms, Theory of Computing, vol. 5, no. 11, pp. 217–238, 2009. DOI: 10.4086/toc.2009.v005a011.

 Available: https://theoryofcomputing.org/articles/v005a011.
- [174] J. Watrous, Simpler semidefinite programs for completely bounded norms, Chicago Journal of Theoretical Computer Science, vol. 2013, no. 8, Jul. 2013. Available: http://cjtcs.cs.uchicago.edu/articles/2013/8/contents.html.
- [175] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: doi.org/10.1017/9781316848142.
- [176] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, *Gaussian quantum information*, *Rev. Mod. Phys.*, vol. 84, pp. 621–669, 2 May 2012. DOI: 10.1103/RevModPhys.84.621. Available: https://link.aps.org/doi/10.1103/RevModPhys.84.621.
- [177] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, Phys. Rev. Lett., vol. 80, pp. 2245-2248, 10 Mar. 1998. DOI: 10.1103/PhysRevLett. 80.2245. Available: https://link.aps.org/doi/10.1103/PhysRevLett.80. 2245.
- [178] W. K. Wootters, Entanglement of formation and concurrence, Quantum Info. Comput., vol. 1, no. 1, pp. 27–44, Jan. 2001, ISSN: 1533-7146.
- [179] T. Yamamoto and R. Ohira, Error suppression by a virtual two-qubit gate, Journal of Applied Physics, vol. 133, no. 17, p. 174401, May 2023, ISSN: 0021-8979. DOI: 10.1063/5.0151037. eprint: https://pubs.aip.org/aip/jap/article-pdf/doi/10.1063/5.0151037/17230548/174401_1_5.0151037.pdf. Available: https://doi.org/10.1063/5.0151037.
- [180] X. Yuan, Y. Liu, Q. Zhao, B. Regula, J. Thompson, and M. Gu, *Universal and operational benchmarking of quantum memories, npj Quantum Information*, vol. 7, no. 1, Jul. 2021, ISSN: 2056-6387. DOI: 10.1038/s41534-021-00444-9. Available: http://dx.doi.org/10.1038/s41534-021-00444-9.

- [181] L.-J. Zhao and L. Chen, Additivity of entanglement of formation via an entanglement-breaking space, Phys. Rev. A, vol. 99, p. 032310, 3 Mar. 2019. DOI: 10.1103/PhysRevA.99.032310. Available: https://link.aps.org/doi/10.1103/PhysRevA.99.032310.
- [182] X. Zhao, L. Zhang, B. Zhao, and X. Wang, *Power of quantum measurement in simulating unphysical operations*, 2023. arXiv: 2309.09963 [quant-ph]. Available: https://arxiv.org/abs/2309.09963.

Appendix A

Hoeffding's inequality

Theorem A.1: Hoeffding's inequality

Let $X_1, ..., X_n$ be i.i.d random variables with $|X_i| \le K$. Define $S_n := \sum_{i=1}^n X_i$. Then for any t > 0

$$\mathbb{P}\left[|S_n - \mathbb{E}[S_n]| \ge t\right] \le 2\exp\left(-\frac{t^2}{2nK^2}\right). \tag{A.1}$$

It will be informative for us to look at a simple example in which Hoeffding's bound is almost achieved, up to some constant factors.

Lemma A.2: Consider i.i.d. random variables X_1, \ldots, X_n that take the values $\pm K$ with probability 1/2 each and let $S_n := \sum_{i=1}^n X_i$. Then for any $t \in [0, nK/4]$

$$\mathbb{P}\left[|S_n| \ge t\right] \ge \frac{2}{15} \exp\left(-\frac{4t^2}{nK^2}\right). \tag{A.2}$$

Proof. We can write $X_i = 2K \cdot Y_i - K$ where Y_i are i.i.d. random variables taking values 0 and 1. Clearly, $\sum_i Y_i$ follows a Binomial distribution on which we can use standard tail bounds, such as [108, Proposition 7.3.2]

$$\mathbb{P}\left[\sum_{i} Y_{i} \ge \frac{n}{2} + t'\right] \ge \frac{1}{15} e^{-16t'^{2}/n}.$$
(A.3)

for $t' \in [0, n/8]$. The desired statement follows directly from

$$\mathbb{P}\left[\left|\sum_{i} X_{i}\right| \geq t\right] = \mathbb{P}\left[\sum_{i} X_{i} \geq t\right] + \mathbb{P}\left[\sum_{i} X_{i} < -t\right] \tag{A.4}$$

$$=2\mathbb{P}\left[\sum_{i}X_{i}\geq t\right] \tag{A.5}$$

$$=2\mathbb{P}\left[\sum_{i}Y_{i}-\frac{n}{2}\geq\frac{t}{2K}\right] \tag{A.6}$$

where we used the symmetry of the distribution $\sum_{i} X_{i}$.

Appendix B

Circuits for QPDs of unitaries

In this appendix, we provide an explicit circuit implementation for the QPD of a gate with unitary operator Schmidt decomposition that was constructed in Lemma 5.46. Let $U \in U(AB)$ be a bipartite unitary with operator Schmidt decomposition

$$U = \sum_{j} u_j L_j \otimes R_j \tag{B.1}$$

for some unitaries $L_j \in U(A)$ and $R_j \in U(B)$. Adopting the notation in the proof of Lemma 5.46, the QPS of U involves three types of operations.

- $\mathcal{L}_j \otimes \mathcal{R}_j$ with quasiprobability coefficient u_i^2 .
- $\mathcal{A}_{j,k}^0 \otimes \mathcal{B}_{j,k}^0$ with quasiprobability coefficient $2u_ju_k$.
- $\mathcal{A}_{j,k}^{\pi/2} \otimes \mathcal{B}_{j,k}^{\pi/2}$ with quasiprobability coefficient $-2u_ju_k$.

The circuit implementation of the first one is trivial. In Figure B.1, we depict a circuit implementation of $\mathcal{A}_{j,k}^{\phi}\otimes\mathcal{B}_{j,k}^{\phi}$ which covers both the second and third types by appropriately choosing ϕ . Most importantly, this circuit implementation only requires one additional ancilla on each bipartition.

The two ancilla qubits are each is prepared in the $(|0\rangle + e^{-i\phi}|1\rangle)/\sqrt{2}$ state. Since the two-circuit halves do not interact, it suffices to show that the upper half acts correctly — the argument for the lower half is analogous. Consider an input state $|\psi\rangle$ on the bipartition A. Controlled on the ancilla qubit, we either apply the gate L_j or L_k , thus resulting into the state

$$\frac{1}{\sqrt{2}}\left(|0\rangle\otimes L_{j}|\psi\rangle+e^{-i\phi}|1\rangle\otimes L_{k}|\psi\rangle\right). \tag{B.2}$$

Next, we apply a Hadamard gate on the ancilla qubit, yielding

$$\frac{1}{2}\left(\left|0\right\rangle\otimes\left(L_{j}+e^{-i\phi}L_{k}\right)\left|\psi\right\rangle+\left|1\right\rangle\otimes\left(L_{j}-e^{-i\phi}L_{k}\right)\left|\psi\right\rangle\right).\tag{B.3}$$

If we next measure the ancilla qubit in the computation basis, the total process is thus described by the quantum instrument $(\mathcal{A}_{j,k}^{+,\phi},\mathcal{A}_{j,k}^{-,\phi})$. In the QPS simulation procedure, we will thus need to weight the final measurement outcome by +1 or -1 depending on the outcome of the ancilla measurement.

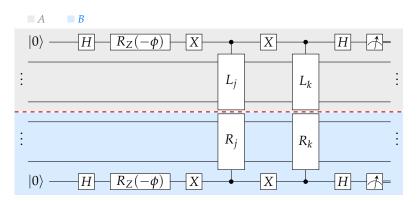


Figure B.1: Circuit implementation for the simulation the operation $\mathcal{A}_{j,k}^{\phi} \otimes \mathcal{B}_{j,k}^{\phi}$ used in the QPD constructed in Lemma 5.46.

Appendix C

Code switching on Kitaev's surface code

In Section 7.3 we described how to realize error-mitigated logical T-gates via code switching. In this appendix, we present a detailed construction for Kitaev's surface code depicted in Figure 7.1. The content of this appendix is taken and adapted from the author's previous work in Ref. [130].

A distance-d surface code is defined on a square lattice of linear size 2d-1. Code qubits live at sites indicated by black circles in Figure 7.1. Let S_1 be the stabilizer group of the surface code generated by X and Z stabilizers located on red (X) and blue (Z) faces of the lattice. We choose logical Pauli operators \overline{X} and \overline{Z} as products of single-qubit X and Z along the left (X) and the top (Z) boundary.

Next, let us define an asymmetric version of the surface code with a small distance for Z errors. Its stabilizer group \mathcal{S}_2 is obtained from \mathcal{S}_1 by introducing some new Z stabilizers and removing some X stabilizers. This modification only affects stabilizers located near the top boundary, as shown on Figure 7.1. To describe this formally, let d=2t+1 be the code distance. Label code qubits located at the top boundary by integers $1,2,\ldots,d$ in the order from the left to the right such that $\overline{Z}=Z_1Z_2\cdots Z_d$. Define Pauli operators

$$G_i := Z_{2i}Z_{2i+1}.$$

The group S_2 is obtained from S_1 by adding t new stabilizers G_1, \ldots, G_t and removing every second X stabilizer adjacent to the top boundary. An X stabilizer of S_1 is removed if it anti-commutes with some operator G_i . Let $q_{loc} := 1$ be the data qubit located at the north-west corner of the lattice. The code S_2 exhibits a local logical- S_2 operator acting only on the qubit S_2 operator acting only on the qubit S_2 exhibits a local logical- S_3

$$\overline{Z}_{loc} = \overline{Z} \prod_{i=1}^{t} G_i = Z_{q_{loc}}.$$
 (C.1)

Note that \overline{Z}_{loc} and \overline{Z} differ by stabilizers of S_2 and thus have the same action on any logical state of S_2 . Thus, the logical T-gate on a qubit encoded by S_2 can be

implemented by either of the following operators:

$$\overline{T} = e^{-i\frac{\pi}{8}\overline{Z}}$$
 and $\overline{T}_{loc} = e^{-i\frac{\pi}{8}\overline{Z}_{loc}}$. (C.2)

Here we ignore the overall phase. Note that \overline{T} also implements a logical T-gate for the code S_1 since both codes have the same logical Pauli operators. Let $\mathcal{G} = \langle G_1, \ldots, G_t \rangle$ be the group generated by G_1, \ldots, G_t . We claim that a logical T-gate on a qubit encoded by the surface code S_1 can be implemented as follows.

Algorithm 3 Logical T-gate by code switching

- 1: Initialize code qubits in a logical state of the surface code S_1
- 2: Measure the syndrome of the asymmetric surface code S_2
- 3: Let $\sigma_i \in \{-1,1\}$ be the measured syndrome of G_i
- 4: Compute $\sigma = \prod_{i=1}^t \sigma_i$
- 5: Apply a single-qubit operator $(\overline{T}_{loc})^{\sigma}$ to the post-measurement state
- 6: Measure the syndrome of the surface code S_1
- 7: Compute a Pauli operator $R \in \mathcal{G}$ consistent with the measured syndrome of \mathcal{S}_1
- 8: Apply *R* to the post-measurement state

Indeed, let $|\psi_j\rangle$ be a state obtained after executing the first j steps of this algorithm. We need to show that $|\psi_8\rangle = \overline{T}\,|\psi_1\rangle$. Let Π be a projector describing the measurement at step 2 such that $|\psi_2\rangle = \Pi\,|\psi_1\rangle$, ignoring the normalization. From $G_i\,|\psi_2\rangle = \sigma_i\,|\psi_2\rangle$ and (C.1),(C.2) one infers $(\overline{T}_{loc})^\sigma\,|\psi_2\rangle = \overline{T}\,|\psi_2\rangle$. Since \overline{Z} is a logical operator for both codes S_i , the logical gate \overline{T} commutes with the syndrome measurement of S_2 . Thus ,

$$|\psi_5\rangle = (\overline{T}_{\mathrm{loc}})^{\sigma} |\psi_2\rangle = \overline{T} |\psi_2\rangle = \Pi \overline{T} |\psi_1\rangle \ .$$

Let Λ be a projector describing the measurement at step 6. In the absence of errors the measured syndrome is non-trivial only for stabilizers of S_1 that anti-commute with some element of S_2 . One can check that there are exactly t such stabilizers which we denote F_1, \ldots, F_t . The stabilizer F_i acts on data qubits 2i - 1, 2i located at the top boundary as well as the data qubit located at the second row between 2i - 1 and 2i, see Figure 7.1. Let $\lambda_i \in \{1, -1\}$ be the syndrome of F_i measured at step 6. The operator R used at step 7 is defined as

$$R = \prod_{i:\lambda_i = -1} \prod_{a=i}^t G_a.$$

Using the commutation rules between the operators F_i and G_a one can easily check that R has the syndrome λ , that is, $RF_i = \lambda_i F_i R$ for all i = 1, ..., t. Write $\Lambda = R\Gamma R$, where Γ is a projector onto the logical subspace of the surface code S_1 . The above shows that

$$|\psi_8\rangle = R\Lambda |\psi_5\rangle = (R\Lambda R)R |\psi_5\rangle = \Gamma R |\psi_5\rangle = \Gamma R\Pi \overline{T} |\psi_1\rangle = \Gamma (R\Pi)\Gamma \overline{T} |\psi_1\rangle.$$

To obtain the last equality we noted that $\Gamma |\psi_1\rangle = |\psi_1\rangle$ and that \overline{T} commutes with Γ . The operator $R\Pi$ is a linear combination of Pauli operators from the group $\mathcal{S}_1 \cdot \mathcal{G}$. One can easily check that this group contains only stabilizers of the surface code and detectable errors. Thus, $\Gamma(R\Pi)\Gamma$ is proportional to Γ . The latter has trivial action on the logical state $\overline{T} |\psi_1\rangle$. This proves that $|\psi_8\rangle = \overline{T} |\psi_1\rangle$, as claimed.

We make Algorithm 3 partially fault-tolerant by repeating syndrome measurements at step 2 and step 6 sufficiently many times. The observed syndromes are used to compute an error-corrected version of the phase σ at step 4 and to perform the final error correction after step 8. Accordingly, the algorithm can fail in two distinct ways. First, the final state could be $(\overline{T})^{-1} |\psi_1\rangle$ instead of $\overline{T} |\psi_1\rangle$ because some syndrome σ_i has been flipped due to an error. Such error results in a wrong phase σ computed at step 4. We shall refer to such events as Pauli frame errors since they can be viewed as applying the T-gate in a wrong Pauli frame. Likewise, an undetected X error that occurs before step 5 on the qubit q_{loc} would result in a Pauli frame error, as can be seen from the identity $\overline{T}_{\text{loc}}X_1 = X_1(\overline{T}_{\text{loc}})^{-1}$. Secondly, the algorithm may fail if the final error correction performed after step 8 resulted in a logical error. We would like to achieve an exponential suppression for logical X errors and, at the same time, ensure that Pauli frame and logical Z errors occur with probability at most $\kappa\varepsilon$, where ε is the physical error rate and κ is a constant prefactor.

To measure the syndrome of the surface code S_1 we use a well-known quantum circuit proposed in Ref. [50], [60]. It requires one ancillary syndrome qubit per each stabilizer, see Figure 7.1. The circuit applies four rounds of CNOTs that compute the syndrome of each stabilizer into the respective syndrome qubit. A syndrome measurement cycle begins by resetting each syndrome qubit to $|0\rangle$ or $|+\rangle$ state for Z and X stabilizers respectively. The cycle ends by measuring each syndrome qubit in the Z- or X-basis. A circuit measuring the syndrome of S_2 requires only a few minor modifications, as described in the figure caption.

For numerical simulations we chose the depolarizing noise model [60]. It depends on a single error rate parameter ε such that each operation in the circuit (a gate, a measurement, or a qubit reset) becomes faulty with the probability ε . Faults on different operations are independent. The model can be described by stochastic Pauli errors enabling efficient simulation using the stabilizer formalism [2]. A faulty CNOT gate is modelled as the ideal CNOT followed by a Pauli error drawn uniformly at random from the two-qubit Pauli group. A faulty measurement is modelled as the ideal measurement whose outcome is flipped. A faulty qubit reset prepares a basis state orthogonal to the ideal one. A faulty idle qubit suffers from a Pauli error X, Y, or Z.

Let us first discuss how to correct Pauli frame errors. Suppose U is a quantum circuit describing the first two steps of Algorithm 3 with L syndrome measurement cycles at step 2. The circuit outputs a quantum state $|\psi_2\rangle$ and a classical syndrome history h specifying the syndrome of each stabilizer of S_2 measured in each of L syndrome cycles. In the absence of errors all stabilizers except for G_1, \ldots, G_t have trivial syndromes. Furthermore, the syndromes of G_i measured in different cycles are the same. Let $\mathcal{F}(U)$ be the set of all possible faulty implementations of U, as prescribed by the depolarizing noise model. Each circuit $\tilde{U} \in \mathcal{F}(U)$ differs from U

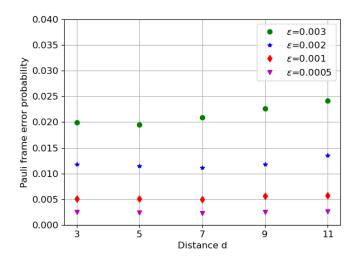


Figure C.1: Probability of a Pauli frame error P_F for the code switching $S_1 \to S_2$ with L=3 syndrome measurement cycles for the asymmetric surface code S_2 . The initial state is chosen as an ideal logical state of the surface code S_1 . A Pauli frame error in the code switching results in the implementation of the logical gate $(\overline{T})^{-1}$ instead of \overline{T} . Our data suggest a scaling $P_F \approx 6\varepsilon + O(d\varepsilon^2)$.

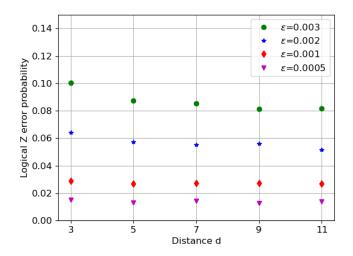


Figure C.2: Probability of a logical Z error P_Z for the code switching $S_1 \to S_2 \to S_1$ with L=3 syndrome cycles for the asymmetric surface code S_2 and d syndrome cycles for the regular surface code S_1 . To enable efficient simulation we skipped the T-gate application in Algorithm 3. Our data suggests a scaling $P_Z \approx 26\varepsilon$ for a large code distance. The large constant prefactor stems from leaving the logical qubit unprotected from Z errors for L+1 syndrome cycles when the full syndrome information for X stabilizers of the code S_1 is unavailable.

by inserting Pauli errors at some space-time locations (measurement and reset errors can be modeled by inserting suitable Pauli errors before the ideal measurement and after the ideal reset). We pick a faulty circuit $\tilde{U} \in \mathcal{F}(U)$ randomly, according to the depolarizing noise model. A simulation of the circuit \tilde{U} yields a noisy syndrome history \bar{h} . To perform error correction we used the maximum-weight matching (MWM) decoder proposed in Ref. [26] which is closely related to the one used in Ref. [50], [60]. The MWM decoder takes the syndrome history \tilde{h} as an input and outputs a candidate faulty circuit $\tilde{U}_{\text{dec}} \in \mathcal{F}(U)$ consistent with \tilde{h} . By propagating all Pauli errors contained in \tilde{U}_{dec} towards the final time step, we compute error-corrected syndromes σ_i and a Pauli correction C acting on the data qubits such that $CU |\psi_1\rangle$ is our guess of the noisy state $\tilde{U}|\psi_1\rangle$, based on the available syndrome information. Step 4 of Algorithm 3 computes an error-corrected phase $\sigma = (-1)^a \sigma_1 \cdots \sigma_t$, where a = 1 if Ccontains an X error on the qubit q_{loc} and a=0 otherwise. To avoid the actual application of the T-gate (which would require non-stabilizer simulators) we chose the initial logical state $|\psi_1\rangle$ as an eigenvector of \overline{Z} , that is, $|\psi_1\rangle = |\overline{b}\rangle$, where $b \in \{0,1\}$ is a random bit. Error correction succeeds if $\tilde{U}\ket{\psi_1}$ is an eigenvector of $(\overline{T}_{loc})^\sigma$ with the eigenvalue $(-1)^b$. Otherwise, a Pauli frame error is declared.

Our numerical results for the probability of a Pauli frame error P_F are shown in Figure C.1. Each data point represents an empirical estimate of P_F obtained by the Monte Carlo method (we used between 50,000 and 200,000 Monte Carlo trials per data point). We chose L=3 syndrome measurement cycles to ensure that any single fault in the measurement of σ_i is correctable. Recall that the MWM decoder detects errors by examining differences between syndromes measured in subsequent cycles. Thus, L=3 provides two parity checks to diagnose errors in the three measured values of σ_i , in a way analogous to the 3-bit repetition code. In contrast, choosing L=2 would only allow us to detect a faulty bit σ_i but not to correct it (if the two measured values of σ_i disagree, there is no way to select the correct value). Choosing L > 3 is undesirable as this leaves the logical qubit unprotected from Z errors for a longer time. Our data suggests a scaling $P_F \approx 6\varepsilon + O(d\varepsilon^2)$. We numerically observed that the dominant contribution to P_F stems from X errors occurring on the qubit $q_{\rm loc}$ in the last syndrome cycle as well as syndrome measurement errors on the Z stabilizer acting on the qubit q_{loc} in the last syndrome cycle. The term $O(d\varepsilon^2)$ can be understood as a result of a weight-two error affecting the measurement of some syndrome bit σ_i . Here $i=1,\ldots,t$ can be arbitrary. Such weight-two error cannot be corrected using only L = 3 syndrome cycles.

Finally, we estimated the probability of logical X and Z errors by simulating a simplified version of Algorithm 3 without steps 3,4,5. In other words, we skipped the application of the physical T-gate. This enables efficient simulation using the stabilizer formalism. The MWM decoder takes as input the combined syndrome history measured at steps 2,6 and calculates a Pauli correction to be applied after step 8. The standard implementation for MWM decoder [26] was properly modified to take into account that some stabilizers are turned off and on during the code switching. Namely, syndromes of the operators G_i measured in the first cycle of step 2 as well as syndromes of the operators F_i measured in the first cycle of step 6 are not used to diagnose errors. As commonly done in the literature, we chose the

number of syndrome cycles for the surface code S_1 equal to the code distance d and added a noiseless syndrome cycle at the end of the protocol.

Our numerical results for the probability of logical Z and X errors P_Z and P_X are shown on Figures C.2 and C.3, respectively. Our data suggests a scaling $P_Z \approx 26\varepsilon$ for a large code distance. The large constant prefactor can be seen as the price we pay for leaving the logical qubit unprotected from Z errors during L syndrome cycles at step 2. In fact, since the syndromes of X stabilizers F_i measured in the first cycle of step 6 are random, these syndrome provide no information about Z errors that occurred in this cycle. Thus, the logical qubit is unprotected from Z errors for L+1=4 cycles. Since each cycle is a depth-6 circuit (four CNOT rounds, reset round, and measurement round), the error correction for Z errors is effectively turned off for 24 time steps. This is in a good agreement with the scaling $P_Z\approx 26\varepsilon$. Finally, we observed that the probability of logical X errors is exponentially suppressed as one increases the code distance, see Figure C.3. Thus, the noise in the implemented logical T-gate is dominated by (coherent) Z errors, as expected.

We anticipate the scaling of P_Z can be improved by optimizing the code switching protocol. For example, one can reduce P_Z roughly by the factor 2/3 by choosing the number of syndrome measurement cycles L for the code S_2 adaptively such that L=2 if the syndromes σ_i measured in the first and the second cycles are the same for all $i=1,\ldots,t$ and L=3 otherwise. This would ensure that a single fault in the measurement of σ_i can be corrected, similar to the L=3 implementation described above. However, the logical qubit remains unprotected from logical Z errors for a shorter time (two cycles instead of three cycles, in the limit $\varepsilon \to 0$).

Under the assumption that we use a code with a sufficiently large distance, the logical X errors are negligible. Hence, the logical error rate $\bar{\epsilon}$ is determined by the logical Z and Pauli frame errors. The latter are described by the channel $(1-P_F)$ id + $P_F \mathcal{S}^{\dagger}$, where $\mathcal{S}^{\dagger}(B) = S^{\dagger}BS$ for any single qubit operator B. Twirling over the X gate gives a Z-type noise channel $(1-P_F/2)$ id + $(P_F/2)\mathcal{Z}$. Assuming that Pauli frame and logical Z errors are independent, the combined logical error rate is therefore $\bar{\epsilon} \approx P_Z + P_F/2 \approx 30\epsilon$ in the limit $\epsilon \ll 1$.

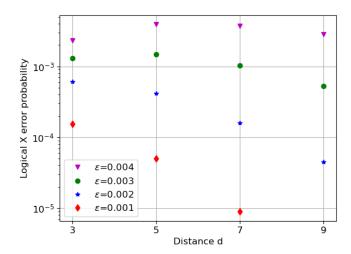


Figure C.3: Probability of a logical X **error** P_X for the code switching $S_1 \to S_2 \to S_1$ with L=3 syndrome cycles for the asymmetric surface code S_2 and d syndrome cycles for the regular surface code S_1 . To enable efficient simulation we skipped the T-gate application in Algorithm 3. Our data suggests that logical X errors are exponentially suppressed as one increases the code distance.