Applications of coherent classical communication and the Schur transform to quantum information theory

by

Aram Wettroth Harrow

B.S., Massachusetts Institute of Technology (2001)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

Author . . . .
Department of Physics
September 23, 2005

Certified by . . . .
Isaac L. Chuang
Associate Professor of Electrical Engineering and Computer Science, and Physics
Thesis Supervisor

Accepted by . . . .
Thomas J. Greytak
Professor of Physics

Applications of coherent classical communication and the Schur transform to quantum information theory

by

Aram Wettroth Harrow

Submitted to the Department of Physics
on September 23, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

Quantum mechanics has led not only to new physical theories, but also a new understanding of information and computation. Quantum information not only yields new methods for achieving classical tasks such as factoring and key distribution but also suggests a completely new set of quantum problems, such as sending quantum information over quantum channels or efficiently performing particular basis changes on a quantum computer. This thesis contributes two new, purely quantum, tools to quantum information theory—coherent classical communication in the first half and an efficient quantum circuit for the Schur transform in the second half.

The first part of this thesis (Chapters 1-4) is in fact built around two loosely overlapping themes. One is quantum Shannon theory, a broad class of coding theorems that includes Shannon and Schumacher data compression, channel coding, entanglement distillation and many others. The second, more specific, theme is the concept of using unitary quantum interactions to communicate between two parties. We begin by presenting new formalism: a general framework for Shannon theory that describes communication tasks in terms of fundamental information processing resources, such as entanglement and classical communication. Then we discuss communication with unitary gates and introduce the concept of *coherent classical communication*, in which classical messages are sent via some nearly unitary process. We find that coherent classical communication can be used to derive several new quantum protocols and unify them both conceptually and operationally with old ones. Finally, we use these new protocols to prove optimal trade-off curves for a wide variety of coding problems in which a noisy channel or state is consumed and two noiseless resources are either consumed or generated at some rate.

The second half of the thesis (Chapters 5-8) is based on the Schur transform, which maps between the computational basis of $(\mathbb{C}^d)^{\otimes n}$ and a basis (known as the *Schur basis*) which simultaneously diagonalizes the commuting actions of the symmetric group $\mathcal{S}_n$ and the unitary group $\mathcal{U}_d$. The Schur transform is used as a subroutine in many quantum communication protocols (which we review and further develop), but previously no polynomial-time quantum circuit for the Schur transform was known. We give such a polynomial-time quantum circuit based on the Clebsch-Gordan transform and then give algorithmic connections between the Schur transform and the quantum Fourier transform on $\mathcal{S}_n$.

Thesis Supervisor: Isaac L. Chuang
Title: Associate Professor of Electrical Engineering and Computer Science, and Physics

Acknowledgments

Very little of the work in this thesis, or indeed that I have done throughout my time in grad school, would have been possible without the support of countless colleagues, collaborators, advisors and friends.

I first want to thank Ike Chuang for guiding me in quantum information from even before I entered grad school and for doing such a good job of alternatingly pushing me, supporting me and turning me loose, always with an eye toward my growth as a scientist. I am also indebted to Eddie Farhi for his encouragement and for innumerable discussions, as well as for teaching me quantum mechanics in the first place. Thanks to Peter Shor for, among other things, sharing so many of his unpublished ideas with me, including a crucial improvement to remote state preparation in the summer of 2001 that led to my first research result in grad school.

I am deeply grateful to Charlie Bennett and Debbie Leung for being wonderful collaborators, mentors and friends throughout my time in grad school. In fact, most of my research direction came from my summers at IBM, for which I would also like to thank Nabil Amer, Guido Burkard, Igor Devetak, David DiVincenzo, Roberto Oliveira, Barbara Terhal and especially John Smolin.

I've been fortunate to have so many productive travel opportunities. Thanks to Michael Nielsen for inviting me to the U. of Queensland, where I enjoyed great discussions with Mick Bremner, Chris Dawson, Jen Dodd, Henry Haselgrove, Tobias Osborne and many other researchers. Thanks also to John Preskill at Caltech, Keiji Matsumoto at ERATO and Noah Linden at the Newton Institute for making these trips possible for me.

Most of my work in grad school has been collaborative and I am grateful to my many collaborators for teaching me, sharing their ideas with me and helping me improve my own ideas. In particular, the work in this thesis was done in collaboration with Dave Bacon, Charlie Bennett, Ike Chuang, Igor Devetak, Debbie Leung, John Smolin and Andreas Winter. I have also had valuable collaborations with Ken Brown, Patrick Hayden, Seth Lloyd, Hoi-Kwong Lo, Michael Nielsen (and many others at UQ), Roberto Oliveira, Ben Recht and Barbara Terhal.

Besides the colleagues I have mentioned so far, I want to thank Herb Bernstein, Carl Caves, Andrew Childs, Matthias Christandl, Andrew Cross, Chris Fuchs, Masahito Hayashi, Denes Petz and Krysta Svore for many valuable discussions. Thanks to Nolan Wallach for crucial discussions on the Schur transform and in particular for the idea behind Section 8.2.

Most of my work in grad school was funded by a QuaCGR grant (ARO contract DAAD19-01-1-06) for which I am grateful to Henry Everitt, Mark Heiligman, the NSA and ARDA.

This thesis is dedicated to my parents and to my teachers: in particular, to Mike Masterson, Mrs. Thomas, Mr. Steidle, Will Repko, Susan Schuur and Gian-Carlo Rota.

Contents

Chapter 0

Introduction

0.1 Motivation and context

Classical theories of information and computation: Though it may seem like a recent phenomenon, computation—the manipulation, storage and transmission of information—has long been one of the most central features of human civilization. Markets of buyers and sellers perform distributed computations to optimize the allocation of scarce resources, natural languages carefully balance the goals of reducing redundancy while correcting errors, and legal systems have long sought reliable algorithms of justice that are accurate and efficient even when implemented with unreliable components. Although these examples cannot be totally separated from human intelligence, they all rely on an impersonal notion of information that has two crucial attributes. First, information can be abstracted away from any particular physical realization; it can be photocopied, memorized, dictated, transcribed and broadcast, always in principle largely preserving the original meaning. Likewise an abstract algorithm for processing information can be performed equivalently using pencil and paper or with digital circuits, as long as it is purely mechanical and makes no use of human insight or creativity. Though the particular features and efficiency of each model of computation may differ, the class of problems they can solve is the same.*

These ideas of computation and information were expressed in their modern forms by Turing and Church in 1936[Tur36, Chu36] and Shannon in 1948[Sha48], respectively. Turing described a hypothetical machine meant to be able to perform any purely mechanical computation, and indeed every method of computation so far devised can be simulated by a Turing machine. Moreover, most practical algorithms used today correspond to the class of problems that a Turing machine can solve given a random number generator and running time bounded by a polynomial of the input size. While Turing showed the fungibility of computation, Shannon proved that information is fungible, so that determining whether any source can be reliably transmitted by any channel reduces, in the limit of long strings, to calculating only two numbers: the information content of the source and the information capacity of the channel.

The abstract theories of Turing and Shannon have been extraordinarily successful because they have happened to match the information-processing technology within our reach in the 20th century; Shannon capacities are nearly achievable by practical codes and most polynomial time algorithms are feasible on modern computers. However, our knowledge of quantum mechanics is now forcing us to rethink our ideas of information and computation, just as relativity revised our notions of space and time. The state of a quantum mechanical system has a number of properties which cannot be reduced to the former, classical, notion of information.

The challenge from quantum mechanics: The basic principles of quantum mechanics are simple to state mathematically, but hard to understand in terms we are familiar with from classical theories

*For two very different perspectives on these ideas, see *Cybernetics* (1948) by N. Weiner and *The Postmodern Condition* (1979) by J.-F. Lyotard.

of physics and information. A quantum system with d levels (e.g. an electron in the p orbital of an atom, which can be in the p_x, p_y or p_z states) has a state described by a unit vector $|\psi\rangle$ that belongs to a d-dimensional complex vector space. Thus, an electron could be in the p_x or p_y state, or in a linear combination of the two, known in chemistry as a hybrid orbital, or in quantum mechanics as a *superposition*. Systems combine via the tensor product, so the combined state space of n d-level systems is d^n-dimensional. A measurement with K outcomes is given by a collection of matrices $\{M_1, \ldots, M_K\}$ such that outcome k has probability $\langle\psi|M_k^\dagger M_k|\psi\rangle$ (here $\langle\psi|$ is the Hermitian conjugate of $|\psi\rangle$) and results in the normalized output state $M_k|\psi\rangle/\sqrt{\langle\psi|M_k^\dagger M_k|\psi\rangle}$; any measurement is possible (on a finite-dimensional system) as long as it satisfies the normalization condition $\sum_{k=1}^K M_k^\dagger M_k = \mathbb{1}$. The possible forms of time evolution are entirely described by the constraints of normalization and linearity; they correspond to maps from $|\psi\rangle$ to $U|\psi\rangle$, where U is a unitary operator ($U^\dagger U = \mathbb{1}$).

These principles bear a number of resemblances to classical wave mechanics, and at face value may not appear particularly striking. However, they have dramatic implications when quantum systems are used to store and manipulate information.

- *Exponentially long descriptions:* While n copies of a classical system require $O(n)$ bits to describe, n copies of a comparable quantum system cannot be accurately described with fewer than $\exp(O(n))$ bits. This is a direct consequence of the tensor product structure of composite quantum systems, in which n two-level systems are described by a unit vector in a 2^n-dimensional complex vector space. On the other hand, the largest classical message that can be reliably encoded in such a system is n bits long[Hol73]. This enormous gap cannot be explained by any classical model of information, even when probabilistic or analog models are considered.

- *Nonlocal state descriptions:* Another consequence of applying the tensor product to state spaces is that a composite system AB can be in an *entangled state* that cannot be separated into a state of system A and a state of system B. While correlated probability distributions have a similar property, an entangled quantum system differs in that the system as a whole can be in a definite state, while its parts still exhibit (correlated) randomness. Moreover, measuring entangled states yields correlations that cannot be obtained from any classical correlated random variable[Per93], though they nevertheless do not permit instantaneous communication between A and B.

- *Reversible unitary evolution:* Since time evolution is unitary, it is always reversible. (Measurement is also reversible once we include the measuring apparatus; see [Per93] or Section 1.1 of this thesis for details.) As an immediate consequence, quantum information can never be deleted, only rearranged, perhaps into a less accessible form. An only slightly more complicated argument can prove that it is impossible to copy an arbitrary quantum state[WZ82], unless we know that the state belongs to a finite set that is perfectly distinguishable by some measurement.

  This contrasts sharply with one of classical information's defining properties, its infinite reproducibility. The idea of *possessing* information takes on an entirely new meaning when referring to quantum information, one that we are only barely beginning to appreciate (e.g. see [Pre99, GC01]).

- *Complementary observables:* Another way to prove that quantum information cannot be cloned is via the *uncertainty principle*, which holds that complementary observables, such as position and momentum, cannot be simultaneously measured; observing one necessarily randomizes the other. The reason this implies no-cloning is that making a perfect copy of a particle would allow the position of one and the momentum of the other to be measured, thereby inferring both quantities about the original system.

  Even though the uncertainty principle describes limitations of quantum information, quantum cryptography turns this into a *strength* of quantum communication, by using uncertainty to hide information from an eavesdropper. The idea is to encode a random bit in one of two randomly chosen complementary observables, so that without knowing how the bit is encoded, it is impossible to measure it without risking disturbance. This can detect any eavesdropper, no matter how sophisticated, and even if the quantum information is sent through completely insecure channels. Combining this process with public classical communication can be used to send unconditionally secure messages[BB84].

- *Interference of amplitudes:* In the two-slit experiment, two beams of light from point sources (such as slits cut into a screen) overlap on a screen, but instead of simply adding, yield alternating bands of constructive and destructive interference. One insight of quantum mechanics is that particles are waves with complex amplitudes, so that interference is still found in the two-slit experiment with single photons, electrons, or even molecules. Measurement breaks the quantum coherence which makes this possible, so observing which slit an electron passes through, no matter how gently this is performed, completely eliminates the interference effect.

  The power of interference would be dramatically demonstrated by building a large-scale quantum computer and using it to solve classical problems. Such a computer could interfere different branches of a computation in much the same way that different paths of an electron can interfere.

These examples are significant not only because they expand the range of what is efficiently computable, but because they force us to revise the logical terms with which we understand the world around us. We can no longer say that an electron either went through one path or the other, or that a quantum computer took a particular computational path or that Schödinger's cat must be either alive or dead. At one point, this suggested that quantum theory needed to revised, but now a consensus is emerging that it is instead classical logic that needs to be rethought.

The operational approach to quantum information: Unfortunately, ever since quantum mechanics was first articulated seventy years ago, it has been difficult to give a clear philosophical interepretation of quantum information. In the last 10-20 years, though, a good deal of progress has been made by thinking about quantum information *operationally*, and studying how information-processing tasks can be accomplished using quantum systems. At the same time, we would like to study quantum information in its own right, preferably by abstracting it away from any particular physical realization.

This operational-yet-abstract approach to quantum information is best realized by the idea of quantum computation. While classical computers are based on bits, which can be either 0 or 1, quantum computers operate on quantum bits or *qubits*, which are 2-level quantum systems. Each state of a quantum memory register (a collection of n qubits, hence with 2^n states) has its own complex amplitude. Performing an elementary quantum gate corresponds to multiplying this (length 2^n) vector of amplitudes by a unitary matrix of size $2^n \times 2^n$. If we prepare an input with nonzero amplitude in many different states, we can run a computation in superposition on all of these input states and then interfere their output amplitudes, just as the amplitudes of differents paths of an electron can interfere. Certain problems appear to lend themselves well to this approach, and allow us to observe constructive interference in "correct" branches of the computation and destructive interference in "incorrect" branches; needless to say, this technique is completely impossible on classical probabilistic computers. For example, Shor's algorithm[Sho94] is able to use interference to factor integers on a quantum computer much faster than the best known classical algorithm can.

Other applications use the fact that amplitudes can add linearly, while probability (or intensity) is proportional to amplitude squared. This is used in Grover's algorithm[Gro96] to search a database of N items with time $O(\sqrt{N})$, or in the more colorful application of "interaction-free measurement," which can safely detect a bomb that will explode if it absorbs a single photon. Here the idea is to constructively interfere N photons, each of amplitude $1/N$, while randomizing the phase that the bomb sees, so that the bomb experiences a total intensity of $N \cdot (1/N)^2 = 1/N$, which can be made arbitrarily small (see [RG02] and references therein).

Purely quantum problems in quantum information: So far all of the examples of the power of quantum information describe goals that are defined entirely in terms of classical information (sharing secret bits, unstructured search, factoring integers) but are more efficiently achieved using quantum information processing resources; we might call these hybrid classical-quantum problems.

As our understanding of quantum information has improved, we have also begun to study information processing tasks which are purely quantum; for example, we might ask at what rate a noisy quantum channel can reliably transmit quantum messages. In fact, it is even possible to think of classical information entirely as a special case of quantum information, a philosophy known as the "Church of the Larger Hilbert Space"[*] which Section 1.1 will explain in detail. The two main contributions of this thesis involve such "purely quantum" tasks, in which both the problem and the solution are given in terms of quantum information. Before explaining them, we will discuss the fields of research that give them context.

Quantum information theory (or more specifically, *quantum Shannon theory*) seeks a quantitative understanding of how various quantum and classical communication resources, such as noisy channels or shared correlation, can be used to simulate other communication resources. The challenge comes both from the much richer structure of quantum channels and states, and from the larger number of communication resources that we can consider; for example, channels can be classical or quantum or can vary continuously between these possibilities. Moreover, (quantum) Shannon theory studies asymptotic capacities; we might ask that n uses of channel send $n(C - \delta_n)$ bits with error ϵ_n, where $\delta_n, \epsilon_n \to 0$ as $n \to \infty$. Since the state of n quantum systems generally requires $\exp(O(n))$ bits to describe, the set of possible communication strategies grows quite rapidly as the number of channel uses increases.

While early work (such as [Hol73, BB84]) focused on using quantum channels to transmit classical messages, the last ten years have seen a good deal of work on the task of sending quantum information for its own sake, or as part of a quantum computer. The main contribution of the first half of this thesis is to show that many tasks previously thought of in hybrid classical-quantum terms (such as using entanglement to help a noisy quantum channel send classical bits) are better thought of as purely quantum communication tasks. We will introduce a new tool, called *coherent classical communication*, to systematize this intuition. Coherent classical communication is actually a purely quantum communication resource; the name indicates that it is obtained by modifying protocols that use classical communication so that they preserve quantum coherence between different messages. We will find that coherent classical communication, together with a rigorous theory of quantum information resources, will give quick proofs of a wide array of optimal quantum communication protocols, including several that have not been seen before.

Quantum complexity theory asks how long it takes quantum computers to solve various problems. Since quantum algorithms include classical algorithms as a special case, the interesting question is when quantum algorithms can perform a task faster than the best possible or best known classical algorithm. The ultimate goal here is generally to solve classical problems (factoring, etc.) and the question is the amount of classical or quantum resources required to do so.

When considering instead "purely quantum" algorithms, with quantum inputs and quantum outputs, it is not immediately apparent what application these algorithms have. However, at the heart of Shor's factoring algorithm, and indeed almost all of the other known or suspected exponential speedups, is the quantum Fourier transform: a procedure that maps quantum input $\sum_x f(x)|x\rangle$ to quantum output $\sum_x \hat{f}(x)|x\rangle$, where $\hat{f}$ is the Fourier transform of the function f. Such a procedure, which Fourier transforms the amplitudes of a wavefunction rather than an array of floating point numbers, would not even make sense on a classical computer; complex probabilities do not exist and global properties of a probability distribution (such as periodicity) cannot be accessed by a single sample. Likewise, Grover's search algorithm can be thought of

[*] This term is due to John Smolin.

as an application of quantum walks[Sze04], a versatile quantum subroutine that is not only faster than classical random walks, but again performs a task that would not be well-defined in terms of classical probabilities. These quantum subroutines represent the core of quantum speedups, as well as the place where our classical intuition about algorithms as logical procedures breaks down. Thus, finding new nontrivial purely quantum algorithms is likely to be the key to understanding exactly how quantum computing is more powerful than the classical model.

The second half of this thesis is based on the Schur transform, a purely quantum algorithm which, like the quantum Fourier transform, changes from a local tensor power basis to a basis that reflects the global properties of the system. While the Fourier transform involves the cyclic group (which acts on an n-bit number by addition), the Schur transform is instead based on the symmetric and unitary groups, which act on n d-dimensional quantum systems by permuting them and by collectively rotating them. The primary contribution of this thesis will be an efficient quantum circuit implementing the Schur transform. As a purely quantum algorithm, the Schur transform does not directly solve any classical problem. However, it is a crucial subroutine for many tasks in quantum information theory, which can now be efficiently implemented on a quantum computer using our methods. More intriguingly, an efficient implementation of the Schur transform raises the hope of finding new types of quantum speedups.

This section has tried to give a flavor of why quantum information is an interesting subject, and of the sort of problems that this thesis contributes to. In the next section, we will set out the contributions of this thesis more precisely with a detailed technical summary.

0.2 Summary of results

This thesis is divided into two halves: Chapters 1-4 discuss information theory and Chapters 5-8 are on the Schur transform. The first chapter of each half is mostly background and the other chapters are mostly new work, though some exceptions to this rule will be indicated. A diagram of how the chapters depend on one another is given in Fig. 0-1.

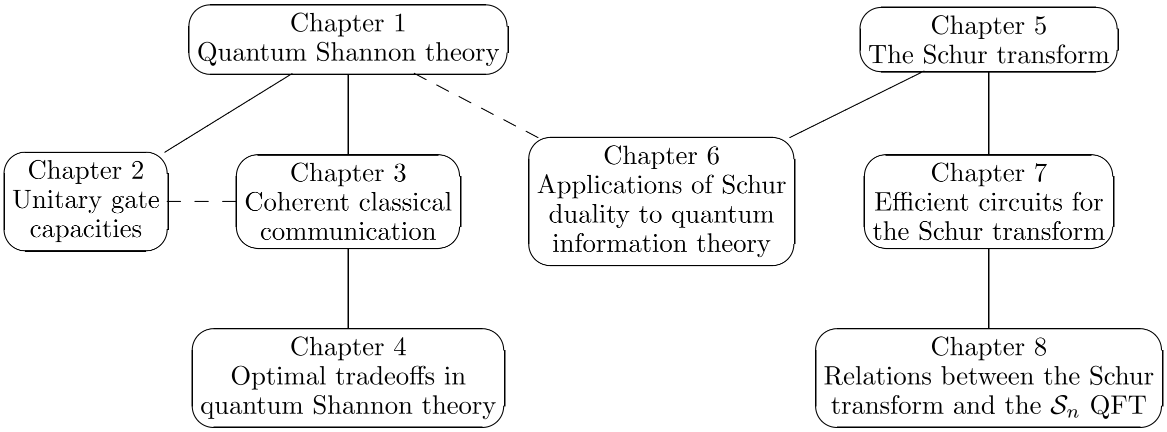

Chapter 1
Quantum Shannon theory
Chapter 5
The Schur transform
Chapter 2
Unitary gate capacities
Chapter 3
Coherent classical communication
Chapter 6
Applications of Schur duality to quantum information theory
Chapter 7
Efficient circuits for the Schur transform
Chapter 4
Optimal tradeoffs in quantum Shannon theory
Chapter 8
Relations between the Schur transform and the $\mathcal{S}_n$ QFT

Figure 0-1: Dependencies between different chapters of this thesis. The solid lines indicate that one chapter depends on another, while the dashed lines mean a partial dependence: Section 6.3 has references to some of the protocols in Section 1.4 and Chapter 3 is motivated by and extends the results of Chapter 2.

Chapter 1 introduces a rigorous framework for concisely stating coding theorems in quantum Shannon theory. The key idea, which has long been tacitly understood but not spelled out explicitly, is that communication protocols in quantum information theory can be thought of as *inequalities* between asymptotic information processing *resources*. Channel coding, for example, says that a noisy channel is at least as useful for communication as the use of a noiseless channel at a particular rate. This chapter rigorously defines and proves the sort of claims we would like to take for granted (e.g., that resources inequalities are transitive) in Section 1.2, goes on to prove some more advanced properties of resource inequalities in Section 1.3 and then summarizes many of the key results of quantum Shannon theory in terms of this new formalism in Section 1.4. Chapter 1 also lays out various definitions and notation used in the rest of the thesis, and in particular gives a detailed description of how the various purifications we use make up the Church of the Larger Hilbert Space (in Section 1.1). This chapter, as well as Chapter 4, is based on joint work with Igor Devetak and Andreas Winter, which is in the process of being turned into a paper[DHW05].

Chapter 2 applies this resource formalism to the problem of communication using a unitary gate that couples two parties. Unitary gates are in some ways more complicated than one-way quantum channels because they are intrinsically bidirectional, but in other ways they are simpler because they do not interact with the environment. The main results of this chapter are capacity formulae for entanglement creation and one-way classical communication using unlimited entanglement, as well as several relations among these and other capacities. We will see that most of these results are superseded by those in the next chapter; the capacity formulae will be simultaneously generalized while the relations between capacities will be explained in terms of a deeper principle. However this chapter helps provide motivation, as well as basic tools, for the results that follow. It is based on [BHLS03] (joint work with Charles Bennett, Debbie Leung and John Smolin), though the original manuscript has been rewritten in order to use the resource formalism of Chapter 1 (which has greatly simplified both definitions and proofs) and to add new material.

Chapter 3 introduces the concept of *coherent classical communication*, a new communication primitive that can be thought of either as classical communication sent through a unitary channel, or as classical communication in which the sender gets the part of the output that normally would go to the environment. This provides an efficient (and in fact, usually optimal) link from a wide variety of classical-quantum protocols (teleportation, super-dense coding, remote state preparation, HSW coding, classical capacities of unitary gates, and more in the next chapter) to purely quantum protocols that often would be much more difficult to prove by other means (super-dense coding of quantum states, quantum capacities of unitary gates, etc.).

This chapter describes some of the general properties of coherent communication, showing how it is equivalent to standard resources and proving conditions under which classical-quantum protocols can be made coherent. After describing how the examples in the last paragraph can all be fruitfully made coherent, we apply these results to find the tradeoff between the rates of classical communication and entanglement generation/consumption possible per use of a unitary gate.

Most of the material in this chapter is based on [Har04], with a few important exceptions. The careful proofs of the converse of Theorem 3.7 (which showed that unlimited back communication does not improve unitary gate capacities for forward communication or entanglement generation) and of coherent remote state preparation are new to the thesis. The full bidirectional version of Theorem 3.1 (showing that sending classical communication through unitary channels is as strong as coherent classical communication) and the discussion of bidirectional rate regions in Section 3.4.3 are both from [HL05], which was joint work with Debbie Leung. Finally, the formal rules for when classical communication can be made coherent were sketched in [DHW04] and will appear in the present form in [DHW05], both of which are joint work with Igor Devetak and Andreas Winter.

Chapter 4 uses coherent classical communication from Chapter 3, the resource formalism from Chapter 3 and a few other tools from quantum Shannon theory (mostly derandomization and measurement compression) to (1) derive three new communication protocols, (2) unify them with four old protocols into a family of related resource inequalities and (3) prove converses that yield six different optimal tradeoff curves for communication protocols that use a noisy channel or state to produce/consume two noiseless resources, such as classical communication, entanglement or quantum communication.

At the top of the family are two purely quantum protocols that can be related by exchanging states with channels: the "mother" protocol for obtaining pure entanglement from a noisy state assisted by a perfect quantum channel, and the "father" protocol for sending quantum information through a noisy channel assisted by entanglement. Combining the parent protocols with teleportation, super-dense coding and entanglement distribution immediately yields all of the other "child" protocols in the family. The parents can in turn be obtained from most of the children by simple application of coherent classical communication. It turns out that all of the protocols in the family are optimal, but since they involve finite amounts of two noiseless resources the converses take the form of two-dimensional capacity regions whose border is a tradeoff curve.

This chapter is based on joint work with Igor Devetak and Andreas Winter[DHW04, DHW05]. Most of the results first appeared in [DHW04], though proofs of the converses and more careful derivations of the parent protocols will be in [DHW05].

Chapter 5 begins the part of the thesis devoted to the Schur transform. Schur duality is a way of relating the representations that appear when the unitary group $\mathcal{U}_d$ and the symmetric group $\mathcal{S}_n$ act on $(\mathbb{C}^d)^{\otimes n}$. Schur duality implies the existence of a *Schur basis* which simultaneously diagonalizes these representations; and the unitary matrix relating the Schur basis to the computational basis is known as the *Schur transform.*

The chapter begins by describing general properties of group representations, such as how they combine in the Clebsch-Gordan transform and how the Fourier transform decomposes the regular representation, using the language of quantum information. Then we go on to describe the Schur transform, explain how it can be used to understand the irreps of $\mathcal{U}_d$ and $\mathcal{S}_n$, and give an idea of how Schur duality can generalized to other groups.

None of the material in this chapter is new (see [GW98] for a standard reference), but a presentation of this form has not appeared before in the quantum information literature. A small amount of the material has appeared in [BCH04] and most will later appear in [BCH05a, BCH05b], all of which are joint work with Dave Bacon and Isaac Chuang.

Chapter 6 describes how Schur duality can be applied to quantum information theory in a way analogous to the use of the method of types in classical information theory. It begins by reviewing the classical method of types in Section 6.1 (following standard texts[CT91, CK81]) and then collects a number of facts that justify the use of Schur duality as a quantum method of types in Section 6.2 (following [GW98, Hay02a, CM04]). Section 6.3 then surveys a wide variety of information theory results from the literature that are based on Schur duality. This section will appear in [BCH05a] and a preliminary version was in [BCH04] (both joint with Dave Bacon and Isaac Chuang).

The only new results of the chapter are in Section 6.4, which gives a way to decompose n uses of a memoryless quantum channel in the Schur basis, and shows how the components of the decomposition can be thought of as quantum analogues of joint types.

Chapter 7 turns to the question of computational efficiency and gives a poly$(n, d, \log 1/\epsilon)$ algorithm that approximates the Schur transform on $(\mathbb{C}^d)^{\otimes n}$ up to accuracy ϵ.

The main idea is a reduction from the Schur transform to the Clebsch-Gordan transform, which is described in Section 7.2. Then an efficient circuit for the Clebsch-Gordan transform is given in Section 7.3. Both of these algorithms are made possible by using *subgroup-adapted bases* which are discussed in Section 7.1.

Section 7.2 first appeared in [BCH04] and the rest of the chapter will soon appear in [BCH05a]. Again, all of this work was done together with Dave Bacon and Isaac Chuang.

Chapter 8 explores algorithmic connections between the Schur transform and the quantum Fourier transform (QFT) over $\mathcal{S}_n$. We begin by presenting *generalized phase estimation*, in which the QFT is used to measure a state in the Schur basis, and then discuss some generalizations and interpretations of the algorithm. Then we give a reduction in the other direction, and show how a variant of the standard $\mathcal{S}_n$ QFT can be derived from one application of the Schur transform.

Generalized phase estimation was introduced in the earlier versions of [BCH04], and will appear along with the other results in this chapter in [BCH05b] (joint with Dave Bacon and Isaac Chuang).

Recommended background: This thesis is meant to be understandable to anyone familiar with the basics of quantum computing and quantum information theory. The textbook by Nielsen and Chuang[NC00] is a good place to start; Chapter 2 (or knowledge of quantum mechanics) is essential for understanding this thesis, Chapters 9 and 11 (or knowledge of the HSW theorem and related concepts) are necessary for the first half of the thesis, and Sections 4.1-4.4, 5.1-5.2 and 12.1-12.5 are recommended. The first six chapters of Preskill's lecture notes[Pre98] are another option. Both [NC00] and [Pre98] should be accessible to anyone familiar with the basics of probability and linear algebra. Further pointers to the literature are contained in Chapters 1 and 5, which respectively introduce the information theory background used in the first half of the thesis and the representation theory background used in the second half.

Chapter 1

Quantum Shannon theory

Two communicating parties, a sender (henceforth called Alice) and a receiver (Bob), usually have, in a mathematical theory of communication, a predefined goal like the perfect transmission of a classical message, but at their disposal are only imperfect resources* like a noisy channel. This is Shannon's channel coding problem [Sha48]: allowing the parties arbitrary local operations (one could also say giving them local resources for free) they can perform encoding and decoding of the message to effectively reduce the noise of the given channel. Their performance is measured by two parameters: the error probability and the number of bits in the message, and quite naturally they want to minimize the former while maximizing the latter.

In Shannon theory, we are particularly interested in the case that the channel is actually a number of independent realizations of the same noisy channel and that the message is long: the efficiency of a code is then measured by the rate, i.e., the ratio of number of bits in a message by number of channel uses. And in particular again, we ask for the asymptotic regime of arbitrarily long messages and vanishing error probability.

Note that not only their given channel, but also the goal of the parties, noiseless communication, is a resource: the channel which transmits one bit perfectly (it is "noisy" in the extreme sense of zero noise), for which we reserve the special symbol $[c \to c]$ and call simply a *cbit*. Thus coding can be described more generally as the conversion of one resource into another, i.e., simulation of the target resource by using the given resource together with local processing. For a generic noisy channel, denoted $\{c \to c\}$, we express such an asymptotically faithful conversion of rate R as a *resource inequality*

$$\{c \to c\} \geq R[c \to c],$$

which we would like to think of as a sort of chemical reaction, and hence address the left hand side as *reactant resource(s)* and the right hand side as *product resource(s)* with R the conversion ratio between these two resoures. In the asymptotic setting, R can be any real number, and the maximum R is the (operational) capacity of the channel — to be precise: to transmit information in the absence of other resources.

Obviously, there exist other useful or desirable resources, such as perfect correlation in the form of a uniformly random bit (abbreviated *rbit*) known to both parties, denoted $[c\,c]$, or more generally some noisy correlation. In quantum information theory, we have further resources: noisy quantum channels and quantum correlations between the parties. Again of particular interest are the noiseless unit resources; $[q \to q]$ is an ideal quantum bit channel (*qubit* for short), and $[q\,q]$ is a unit of maximal entanglement, a two-qubit singlet state (*ebit*). The study of asymptotic conversion rates between the larger class of quantum information-theoretic resources is known as *quantum Shannon theory* and is the main focus of this half of the thesis.

To illustrate the goals of quantum Shannon theory, it is instructive to look at the conversions

[*] The term is used here in an everyday sense; later in this chapter we make it mathematically precise.

permitted by the unit resources $[c \to c]$, $[q \to q]$ and $[q\,q]$, where resource inequalities are finite and exact: the following inequalities always refer to a specific integral number of available resources of a given type, and the protocol introduces no error. We mark such inequalities by a $*$ above the $\geq$ sign. For example, it is always possible to use a qubit to send one classical bit, $[q \to q] \overset{*}{\geq} [c \to c]$, and to distribute one ebit, $[q \to q] \overset{*}{\geq} [q\,q]$; the latter is referred to as entanglement distribution (ED).

More inequalities are obtained by combining resources. Super-dense coding [BW92] is a coding protocol to send two classical bits using one qubit and one ebit:

$$[q \to q] + [q\,q] \overset{*}{\geq} 2[c \to c]. \tag{SD}$$

Teleportation [BBC$^+$93] is expressed as

$$2[c \to c] + [q\,q] \overset{*}{\geq} [q \to q]. \tag{TP}$$

In [BBC$^+$93] the following argument was used that the ratio of $1:2$ between $[q \to q]$ and $[c \to c]$ in these protocols is optimal, even with unlimited entanglement, and even asymptotically: assume, with $R > 1$, $[q \to q] + \infty[q\,q] \geq 2R[c \to c]$; then chaining this with (TP) gives $[q \to q] + \infty[qq] \geq R[q \to q]$. Hence by iteration $[q \to q] + \infty[q\,q] \geq R^k[q \to q] \geq R^k[c \to c]$ for arbitrary k, which can make R^k arbitrarily large, and this is easily disproved. Analogously, $2[c \to c] + \infty[q\,q] \geq R[q \to q]$, with $R > 1$, gives, when chained with (SD), $2[c \to c] + \infty[q\,q] \geq 2R[c \to c]$, which also easily leads to a contradiction. In a similar way, the optimality of the one ebit involved in both (SD) and (TP) can be seen.

While the above demonstration looks as if we did nothing but introduce a fancy notation for things understood perfectly well otherwise, in this chapter we want to make the case for a systematic theory of resource inequalities. We will present a framework general enough to include most two-player setups, specifically designed for the asymptotic memoryless regime. There are three main issues there: first, a suitably flexible definition of a *protocol*, i.e., a way of combining resources (and with it a mathematically precise notion of a resource inequality); second, a justification of the composition (chaining) of resource inequalities; and third, general tools to produce new protocols (and hence resource inequalities) from existing ones.

The benefit of such a theory should be clear then: while it does not mean that we get coding theorems "for free", we *do* get many protocols by canonical modifications from others, which saves effort and provides structural insights into the logical dependencies among coding theorems. As the above example shows, we also can relate (and sometimes actually prove) the converses, i.e. the statements of optimality, using the resource calculus.

The remainder of this chapter will systematically develop the resource formalism of quantum Shannon theory, and will show how it can concisely express and relate many familiar results in quantum information.

Section 1.1 covers the preliminaries and describes several complementary formalisms for quantum mechanics, which serve diverse purposes in the study of quantum information processing. Here also some basic facts are collected.

Section 1.2 sets up the basic communication scenario we will be interested in. It contains definitions and basic properties of so-called finite resources, and how they can be used in protocols. Building upon these we define asymptotic resources and inequalities between them, in such a way as to ensure natural composability properties.

Section 1.3 contains a number of general and useful resource inequalities.

Section 1.4 compiles most of the hitherto discovered coding theorems, rewritten as resource inequalities.

Section 1.5 concludes with a discussion of possible extensions to the resource formalism developed in the rest of the chapter.

The following three chapters will apply this formalism to develop new coding results.

Chapter 2 will examine the communication and entanglement-generating capacities of bipartite unitary gates. It is primarily based on [BHLS03] (joint work with Charles Bennett, Debbie Leung and John Smolin).

Chapter 3 develops the idea of *coherent classical communication* and applies it to a variety of topics in quantum Shannon theory, following the treatment of [Har04] and [HL05] (joint work with Debbie Leung).

Chapter 4 shows how coherent classical communicaton can be used to derive new quantum protocols and unify old ones into a family of resource inequalities. This chapter, as well as the present one, are based on [DHW04, DHW05] (joint work with Igor Devetak and Andreas Winter).

1.1 Preliminaries

This section is intended to introduce notation and ways of speaking about quantum mechanical information scenarios. We also state several key lemmas needed for the technical proofs. Most of the facts and the spirit of this section can be found in [Hol01]; a presentation slightly more on the algebraic side is [Win99b], appendix A.

1.1.1 Variations on formalism of quantum mechanics

We start by reviewing several equivalent formulations of quantum mechanics and discussing their relevance for the study of quantum information processing. As we shall be using several of them in different contexts, it is useful to present them in a systematic way. The main two observations are, first, that a classical random variable can be identified with a quantum systems equipped with a preferred basis, and second, that a quantum Hilbert space can always be extended to render all states pure (via a reference system) and all operations unitary (via an environment system) on the larger Hilbert space.

Both have been part of the quantum information processing folklore for at least a decade (the second of course goes back much farther: the GNS construction, Naimark's and Stinespring's theorems, see [Hol01]), and roughly correspond to the "Church of the Larger Hilbert Space" viewpoint.

Based on this hierarchy of embeddings C(lassical) $\Rightarrow$ Q(uantum) $\Rightarrow$ P(ure), in the above sense, we shall see how the basic "CQ" formalism of quantum mechanics gets modified to (embedded into) CP, QQ, QP, PQ and PP formalisms. (The second letter refers to the way quantum information is presented; the first, how knowledge about this information is presented.) We stress that from an operational perspective they are all equivalent — however, which formalism is the most useful depends on the context.

Throughout the thesis we shall use labels such as A (similarly, B, C, etc.) to denote not only a particular quantum system but also the corresponding Hilbert space (which is also denoted $\mathcal{H}_A$) and to some degree even the set of bounded linear operators on that Hilbert space (also denoted $\mathcal{L}(\mathcal{H}_A)$ or $\mathcal{L}(A)$). If $|\psi\rangle$ is a pure state, then we will sometimes use $|\psi\rangle$ to denote the density matrix $|\psi\rangle\langle\psi|$. When talking about tensor products of spaces, we will habitually omit the tensor sign, so $A \otimes B = AB$, etc. Labels such as X, Y, etc. will be used for classical random variables. For simplicity, all spaces and ranges of variables will be assumed to be finite.

The CQ formalism. This formalism is the most commonly used one in the literature, as it captures most of the operational features of a "Copenhagen" type quantum mechanics, which concerns itself more with the behavior of quantum systems than their meaning, reserving ontological statements about probabilities, measurement outcomes, etc. for classical systems. The postulates of quantum mechanics can be classified into static and dynamic ones. The static postulates define the static entities of the theory, while the dynamic postulates describe the physically allowed evolution of the static entities. In defining classes of static and dynamics entities, we will try to highlight their (quantum) information-theoretic significance.

The most general static entity is an *ensemble* of quantum states $(p_x, \rho_x)_{x \in \mathcal{X}}$. The probability distribution $(p_x)_{x \in \mathcal{X}}$ is defined on some set $\mathcal{X}$ and is associated with the random variable X. The ρ_x are density operators (positive Hermitian operators of unit trace) on the Hilbert space of a quantum system A. The state of the quantum system A is thus correlated with the classical index random variable X. We refer to XA as a hybrid classical-quantum system, and the ensemble $(p_x, \rho_x)_{x \in \mathcal{X}}$ is the "state" of XA. We will occasionally refer to a classical-quantum system as a "$\{c\,q\}$ entity". Special cases of $\{c\,q\}$ entities are $\{c\}$ entities ("classical systems", i.e. random variables) and $\{q\}$ entities (quantum systems).

The most general dynamic entity would be a map between two $\{c\,q\}$ entities (hence, and throughout the thesis, we describe dynamics in the Schrödinger picture). Let us highlight only a few special cases:

The most general map from a $\{c\}$ entity to a $\{q\}$ entity is a *state preparation map* or a "$\{c \to q\}$ entity". It is defined by a *quantum alphabet* $(\rho_x)_{x \in \mathcal{X}}$ and maps the classical index x to the quantum state ρ_x.

Next we have a $\{q \to c\}$ entity, a *quantum measurement*, defined by a positive operator-valued measure (POVM) $(M_x)_{x \in \mathcal{X}}$, where M_x are positive operators satisfying $\sum_x M_x = \mathbb{1}$, with the identity operator $\mathbb{1}$ on the underlying Hilbert space. The action of the POVM $(M_x)_{x \in \mathcal{X}}$ on some quantum system ρ results in the random variable defined by the probability distribution $(\operatorname{Tr} \rho M_x)_{x \in \mathcal{X}}$ on $\mathcal{X}$. POVMs will be denoted with roman capitals: L, M, N, P, etc.

A $\{q \to q\}$ entity is a *quantum operation*, a completely positive and trace preserving (CPTP) map $\mathcal{N} : A \to B$, described (non-uniquely) by its *Kraus representation*: a set of operators $\{N_x\}_{x \in \mathcal{X}}$, $\sum_x N_x^\dagger N_x = \mathbb{1}^B$, whose action is given by

$$\mathcal{N}(\rho) = \sum_x N_x \rho N_x^\dagger.$$

(When referring to operators, we use † for the adjoint, while $*$ is reserved for the complex conjugate. In Chapters 5-7, we will also apply $*$ to representation spaces to indicate the dual representation.) A CP map is defined as above, but with the weaker restriction $\sum_x A_x^\dagger A_x \leq \mathbb{1}^B$, and by itself is unphysical (or rather, it includes a postselection of the system). Throughout, we will denote CP and CPTP maps by calligraphic letters: $\mathcal{L}$, $\mathcal{M}$, $\mathcal{N}$, $\mathcal{P}$, etc. A special CPTP map is the identity on a system A, $\mathrm{id}^A : A \to A$, with $\mathrm{id}^A(\rho) = \rho$. More generally, for an isometry $U : A \to B$, we denote — for once deviating from the notation scheme outlined here — the corresponding CPTP map by the same letter: $U(\rho) = U \rho U^\dagger$.

A $\{q \to cq\}$ entity is an *instrument* $\mathbb{P}$, described by an ordered set of CP maps $(\mathcal{P}_x)_x$ that add up to a CPTP map. $\mathbb{P}$ maps a quantum state ρ to the ensemble $(p_x, \mathcal{P}_x(\rho)/p_x)_x$, with $p_x = \operatorname{Tr} \mathcal{P}_x(\rho)$. A special case of an instrument is one in which $\mathcal{P}_x = p_x \mathcal{N}_x$, and the $\mathcal{N}_x$ are CPTP; it is equivalent to an ensemble of CPTP maps, $(p_x, \mathcal{N}_x)_{x \in \mathcal{X}}$. Instruments will be denoted by blackboard style capitals: $\mathbb{L}$, $\mathbb{M}$, $\mathbb{N}$, $\mathbb{P}$, etc.

A $\{cq \to q\}$ entity is given by an ordered set of CPTP maps $(\mathcal{N}_x)_x$, and maps the ensemble $(p_x, \rho_x)_{x \in \mathcal{X}}$ to $\sum_x p_x \mathcal{N}_x(\rho_x)$. By contrast, a $\{c, q \to q\}$ map saves the classical label, mapping $(p_x, \rho_x)_{x \in \mathcal{X}}$ to $(p_x, \mathcal{N}_x(\rho_x))_{x \in \mathcal{X}}$.

In quantum information theory the CQ formalism is used for proving direct coding theorems of a part classical – part quantum nature, such as the HSW theorem [Hol98, SW97]. In addition, it is most suitable for computational purposes.

For two states, we write $\rho^{RA} \supseteq \sigma^A$ to mean that the state σ^A is a *restriction* of ρ^{RA}, namely $\sigma^A = \operatorname{Tr}_R \rho^{RA}$. The subsystem R is possibly null (which we write $R = \emptyset$), i.e., a 1-dimensional Hilbert space. Conversely, ρ^{RA} is called an *extension* of σ^A. Furthermore, if ρ^{RA} is pure it is called a *purification* of σ^R. The purification is unique up to a local isometry on R: this is an elementary consequence of the Schmidt decomposition (discussed in Section 2.1.2). These notions carry over to dynamic entities as well. For two quantum operations $\mathcal{A} : A \to BE$ and $\mathcal{B} : A \to B$ we write $\mathcal{A} \supseteq \mathcal{B}$ if $\mathcal{B} = \operatorname{Tr}_E \circ \mathcal{A}$. If $\mathcal{A}$ is an isometry, is is called an *isometric extension* of $\mathcal{B}$, and is unique up to an isometry on E — this and the existence of such a *dilation* are known as Stinespring's theorem [Sti55].

Observe that we can safely represent noiseless quantum evolution by isometries between systems (whereas quantum mechanics demands *unitarity*): this is because our systems are all finite, and we can embed the isometries into unitaries on larger systems. Thus we lose no generality but gain flexibility.

The CP formalism. In order to define the CP formalism, it is necessary to review an alternative representation of the CQ formalism that involves fewer primitives. For instance,

- $\{q\}$. A quantum state ρ^A is referred to by its purification $|\phi\rangle^{AR}$.
- $\{c\,q\}$, $\{c \to q\}$. The ensemble $(p_x, \rho^A_x)_x$ [resp. quantum alphabet $(\rho^A_x)_x$] is similarly seen as restrictions of a pure state ensemble $(p_x, |\phi_x\rangle^{AR})_x$ [resp. quantum alphabet $(|\phi_x\rangle^{AR})_x$].
- $\{q \to q\}$. A CPTP map $\mathcal{N} : A \to B$ is referred to by its isometric extension $U_{\mathcal{N}} : A \to BE$.
- $\{q \to c\}$. A POVM $(M_x)_x$ on the system A is equivalent to some isometry $U_M : A \to AE_X$, followed by a von Neumann measurement of the system E_X in basis $\{|x\rangle^{E_X}\}$, and discarding A.
- $\{q \to c\,q\}$. An instrument $\mathbb{P}$ is equivalent to some isometry $U_{\mathbb{P}} : A \to BEE_X$, followed by a von Neumann measurement of the system E_X in basis $\{|x\rangle^{E_X}\}$, and discarding E.
- $\{c, q \to q\}$ The ensemble of CPTP maps $(p_x, \mathcal{N}_x)_x$ is identified with the ensemble of isometric extensions $(p_x, U_{\mathcal{N}_x})_x$.

In this alternative representation of the CQ formalism all the quantum static entities are thus seen as restrictions of pure states and all quantum dynamic entities are combinations of performing isometries, von Neumann measurements, and discarding auxiliary subsystems. The *CP formalism* is characterized by never discarding (tracing out) the auxiliary subsystems (reference systems, environments, ancillas); they are kept in the description of our system. As for the auxiliary subsystems that get (von-Neumann-) measured, without loss of generality they may be discarded: the leftover state of such a subsystem may be set to a standard state $|0\rangle$ (and hence decoupled from the rest of the system) by a local unitary conditional upon the measurement outcome.

The CP formalism is mainly used in quantum information theory for proving direct coding theorems of a quantum nature, such as the quantum channel coding theorem (see e.g. [Dev05a]).

The QP formalism. The QP formalism differs from CP in that the classical random variables, i.e. classical systems, are embedded into quantum systems, thus enabling a unified treatment of the two.

- $\{c\}$. The classical random variable X is identified with a dummy quantum system X equipped with preferred basis $\{|x\rangle^X\}$, in the state $\sigma^X = \sum_x p_x |x\rangle\langle x|$. The main difference between random variables and quantum systems is that random variables exist without reference to a particular physical implementation, or a particular system "containing" it. In the QP formalism this is reflected in the fact that the state σ^X remains intact under the "copying" operation $\overline{\Delta} : X \to XX'$, with Kraus representation $\{|x\rangle^X |x\rangle^{X'} \langle x|^X\}$. In this way, instances of the same random variable may be contained in different physical systems.

- $\{c\,q\}$. An ensemble $(p_x, |\phi_x\rangle^{AR})_x$ is represented by a quantum state

$$\sigma^{XAR} = \sum_x p_x |x\rangle\langle x|^X \otimes \phi_x^{AR}.$$

- $\{c \to q\}$. A state preparation map $(|\phi_x\rangle^{AR})_x$ is given by the isometry $\sum_x |\phi_x\rangle^{AR}|x\rangle^X\langle x|^X$, followed by tracing out X.
- $\{cq \to q\}$. The ensemble of isometries (p_x, U_x) is represented by the controlled isometry

$$\sum_x |x\rangle\langle x|^X \otimes U_x.$$

- $\{q \to c\}, \{q \to c\,q\}$. POVMs and instruments are treated as in the CP picture, except that the final von Neumann measurement is replaced by a completely dephasing operation $\overline{\mathrm{id}} : E_X \to X$, defined by the Kraus representation $\{|x\rangle^X\langle x|^{E_X}\}_x$.

The QP formalism is mainly used in quantum information theory for proving converse theorems.

Other formalisms. The QQ formalism is obtained from the QP formalism by tracing out the auxiliary systems, and is also convenient for proving converse theorems. In this formalism the primitives are general quantum states (static) and quantum operations (dynamic).

The PP formalism involves further "purifying" the classical systems in the QP formalism; it is distinguished by its remarkably simple structure: all of quantum information processing is described in terms of isometries on pure states. There is also a PQ formalism, for which we don't see much use; one may also conceive of hybrid formalisms, such as QQ/QP, in which some but not all auxiliary systems are traced out. One should remain flexible. We will usually indicate, however, which formalism we are using as we go along.

1.1.2 Quantities, norms, inequalities, and miscellaneous notation

For a state ρ^{RA} and quantum operation $\mathcal{N} : A \to B$ we identify, somewhat sloppily,

$$\mathcal{N}(\rho) := (\mathrm{id}^R \otimes \mathcal{N})\rho^{RA}.$$

With each state ρ^B, one may associate a quantum operation that appends the state to the input, namely $\mathcal{A}^\rho : A \to AB$, defined by

$$\mathcal{A}^\rho(\sigma^A) = \sigma^A \otimes \rho^B.$$

The state ρ and the operation $\mathcal{A}^\rho$ are clearly equivalent in an operational sense.

Given some state, say ρ^{XAB}, one may define the usual entropic quantities with respect to it. Recall the definition of the von Neumann entropy $H(A) = H(A)_\rho = H(\rho^A) = -\operatorname{Tr}(\rho^A \log \rho^A)$, where $\rho^A = \operatorname{Tr}_{XB} \rho^{XAB}$. When we specialize to binary entropy this becomes $H_2(p) := -p \log p - (1-p) \log p$. Throughout this thesis exp and log are base 2. Further define the conditional entropy [CA97]

$$H(A|B) = H(A|B)_\rho = H(AB) - H(B),$$

the quantum mutual information [CA97]

$$I(A;B) = I(A;B)_\rho = H(A) + H(B) - H(AB),$$

the coherent information [Sch96, SN96]

$$I(A\,\rangle B) = -H(A|B) = H(B) - H(AB),$$

and the conditional mutual information

$$I(A;B|X) = H(A|X) + H(B|X) - H(AB|X).$$

Note that the conditional mutual information is always non-negative, thanks to strong subadditivity [LR73].

It should be noted that conditioning on classical variables (systems) amounts to averaging. For instance, for a state of the form

$$\sigma^{XA} = \sum_x p_x |x\rangle\langle x|^X \otimes \rho_x^A,$$

$$H(A|X)_\sigma = \sum_x p_x H(A)_{\rho_x}.$$

We shall freely make use of standard identities for these entropic quantities, which are formally identical to the classical case (see Ch. 2 of [CT91] or Ch. 1 of [CK81]). One such identity is the so-called chain rule for mutual information,

$$I(A;BC) = I(A;B|C) + I(A;C),$$

and using it we can derive an identity will later be useful:

$$I(X;AB) = H(A) + I(A\rangle BX) - I(A;B) + I(X;B). \tag{1.1}$$

We shall usually work in situations where the underlying state is unambiguous, but as shown above, we can emphasize the state by putting it in subscript.

We measure the distance between two quantum states ρ^A and σ^A by the trace norm,

$$\|\rho^A - \sigma^A\|_1,$$

where $\|\omega\|_1 = \operatorname{Tr}\sqrt{\omega^\dagger\omega}$; for Hermitian operators this is the sum of absolute values of the eigenvalues. If $\|\rho^A - \sigma^A\|_1 \leq \epsilon$, then we sometimes write that $\rho \overset{\epsilon}{\approx} \sigma$. An important property of the trace distance is its monotonicity under quantum operations $\mathcal{N}$:

$$\|\mathcal{N}(\rho^A) - \mathcal{N}(\sigma^A)\|_1 \leq \|\rho^A - \sigma^A\|_1.$$

In fact, the trace distance is operationally connected to the distinguishability of the states: if ρ and σ have uniform prior, Helstrom's theorem [Hel76] says that the maximum probability of correct identification of the state by a POVM is $\frac{1}{2} + \frac{1}{4}\|\rho - \sigma\|_1$.

The trace distance induces a metric on density matrices under which the von Neumann entropy is a continuous function. This fact is known as Fannes' inequality[Fan73, Nie00].

Lemma 1.1 (Fannes). *For any states ρ^A, σ^A defined on a system A of dimension d, if $\|\rho^A - \sigma^A\|_1 \leq \epsilon$ then*

$$|H(A)_\rho - H(A)_\sigma| \leq \epsilon \log d + \eta(\epsilon) \tag{1.2}$$

where $\eta(\epsilon)$ is defined (somewhat unconventionally) to be $-\epsilon \log \epsilon$ if $\epsilon \leq 1/e$ or $(\log e)/e$ otherwise.

Fannes' inequality leads to the following useful corollary:

Lemma 1.2. *For the quantity $I(A\rangle B)$ defined on a system AB of total dimension d, if $\|\rho^{AB} - \sigma^{AB}\|_1 \leq \epsilon$ then*

$$|I(A\rangle B)_\rho - I(A\rangle B)_\sigma| \leq \eta'(\epsilon) + K\epsilon \log d,$$

where $\lim_{\epsilon \to 0} \eta'(\epsilon) = 0$ and K is some constant. The same holds for $I(A;B)$ and other entropic quantities. □

Define a distance measure between two quantum operations $\mathcal{M}, \mathcal{N} : A_1 A_2 \to B$ with respect to some state ω^{A_1} by

$$\|\mathcal{M} - \mathcal{N}\|_{\omega^{A_1}} := \max_{\zeta^{RA_1A_2} \supseteq \omega^{A_1}} \left\| (\mathrm{id}^R \otimes \mathcal{M})\zeta^{RA_1A_2} - (\mathrm{id}^R \otimes \mathcal{N})\zeta^{RA_1A_2} \right\|_1. \tag{1.3}$$

The maximization may, w.l.o.g., be performed over pure states $\zeta^{RA_1A_2}$. This is due to the monotonicity of trace distance under the partial trace map. Important extremes are when A_1 or A_2 are null. The first case measures absolute closeness between the two operations (and in fact, $\| \cdot \|_{\emptyset}$ is the dual of the cb-norm[KW04]), while the second measures how similar they are relative to a particular input state. Eq. (1.3) is written more succinctly as

$$\|\mathcal{M} - \mathcal{N}\|_{\omega} := \max_{\zeta \supseteq \omega} \|(\mathcal{M} - \mathcal{N})\zeta\|_1.$$

We say that $\mathcal{M}$ and $\mathcal{N}$ are ϵ-close with respect to ω if

$$\|\mathcal{M} - \mathcal{N}\|_{\omega} \leq \epsilon.$$

Note that $\| \cdot \|_{\omega}$ is a norm only if ω has full rank; otherwise, different operations can be at distance 0. If ρ and σ are ϵ-close then so are $\mathcal{A}^{\rho}$ and $\mathcal{A}^{\sigma}$ (with respect to $\emptyset$, hence every state).

Recall the definition of the fidelity of two density operators with respect to each other:

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2 = \left(\mathrm{Tr} \sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}} \right)^2.$$

For two pure states $|\phi\rangle$, $|\psi\rangle$ this amounts to

$$F(|\phi\rangle\langle\phi|, |\psi\rangle\langle\psi|) = |\langle\phi|\psi\rangle|^2.$$

We shall need the following relation between fidelity and the trace distance [FvdG99]

$$1 - \sqrt{F(\rho, \sigma)} \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)}, \tag{1.4}$$

the second inequality becoming an equality for pure states. Uhlmann's theorem [Uhl76, Joz94] states that, for any fixed purification $|\phi\rangle\langle\phi|$ of σ,

$$F(\rho, \sigma) = \max_{|\psi\rangle\langle\psi| \supseteq \rho} F(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|).$$

As the fidelity is only defined between two states living on the same space, we are, of course, implicitly maximizing over extensions $|\psi\rangle\langle\psi|$ that live on the same space as $|\phi\rangle\langle\phi|$.

Lemma 1.3. *If $\|\rho - \sigma\|_1 \leq \epsilon$ and $\sigma' \supseteq \sigma$, then there exists some $\rho' \supseteq \rho$ for which $\|\rho' - \sigma'\|_1 \leq 2\sqrt{\epsilon}$.*

Proof. Fix a purification $|\phi\rangle\langle\phi|^{ABC} \supseteq \sigma'^{AB} \supseteq \sigma^A$. By Uhlmann's theorem, there exists some $|\psi\rangle\langle\psi|^{ABC} \supseteq \rho^A$ such that

$$F(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = F(\rho, \sigma) \geq 1 - 2\epsilon,$$

using also Eq. (1.4) Define $\rho'^{AB} = \mathrm{Tr}_C |\psi\rangle\langle\psi|^{ABC}$. By the monotonicity of trace distance under the partial trace map and Eq. (1.4), we have

$$\|\rho' - \sigma'\|_1 \leq \||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 \leq 2\sqrt{\epsilon},$$

as advertised. □

Lemma 1.4. *The following statements hold for density operators ω^A, ${\omega'}^{AA'}$, σ^A, $\rho^{A'}$, Ω^{A_1}, and quantum operations $\mathcal{M}'$,$\mathcal{N}'$: $AA'B \to C$, $\mathcal{M}$,$\mathcal{N}$: $AB \to C$, $\mathcal{K}$,$\mathcal{L}$: $A'B' \to C'$, and $\mathcal{M}_i$,$\mathcal{N}_i$: $A_iA_i^* \to A_{i+1}\hat{A}_{i+1}$.*

1. *If $\omega' \supseteq \omega$ then $\|\mathcal{M}' - \mathcal{N}'\|_{\omega'} \leq \|\mathcal{M}' - \mathcal{N}'\|_{\omega}$.*
2. *$\|\mathcal{M} - \mathcal{N}\|_{\omega} \leq \|\mathcal{M} - \mathcal{N}\|_{\sigma} + 2\sqrt{\|\omega - \sigma\|_1}$.*
3. *$\|\mathcal{M} \otimes \mathcal{K} - \mathcal{N} \otimes \mathcal{L}\|_{\omega\otimes\rho} \leq \|\mathcal{M} - \mathcal{N}\|_{\omega} + \|\mathcal{K} - \mathcal{L}\|_{\rho}$.*
4. *$\|\mathcal{M}_k \circ \cdots \circ \mathcal{M}_1 - \mathcal{N}_k \circ \cdots \circ \mathcal{N}_1\|_{\Omega} \leq \sum_i \|\mathcal{M}_i - \mathcal{N}_i\|_{(\mathcal{M}_{i-1}\circ\cdots\circ\mathcal{M}_1)(\Omega)}$.*

Proof. Straightforward. □

Finally, if we have systems A_1, A_2, ..., A_n, we use the shorthand $A^n = A_1 \ldots A_n$. Also, the set $\{1, \ldots, d\}$ is denoted $[d]$.

1.2 Information processing resources

In this section, the notion of a information processing resource will be rigorously introduced. Unless stated otherwise, we shall be using the QQ formalism (and occasionally the QP formalism) in order to treat classical and quantum entities in a unified way.

1.2.1 The distant labs paradigm

The communication scenarios we will be interested involve two or more separated parties. Each party is allowed to perform arbitrary local operations in his or her lab for free. On the other hand, non-local operations (a.k.a. *channels*) are valuable resources. In this thesis, we consider the following parties:

- *Alice* (A)
- *Bob* (B): Typically quantum Shannon theory considers only problems involving communication from Alice to Bob. This means working with channels from Alice to Bob (i.e. of the form $\mathcal{N} : A' \to B$) and arbitrary states ρ^{AB} shared by Alice and Bob. However, the next two chapters will also consider some bidirectional communication problems.
- *Eve* (E): In the CP and QP formalisms, we purify noisy channels and states by giving a share to the environment. Thus, we replace $\mathcal{N} : A' \to B$ with the isometry $U_{\mathcal{N}} : A' \to BE$ and replace ρ^{AB} with ψ^{ABE}.[*] We consider a series of operations equivalent when they differ only by a unitary rotation of the environment.
- *Reference* (R): Suppose Alice wants to send an ensemble of states $\{p_i, |\alpha_i\rangle^A\}$ to Bob with average density matrix $\rho^A = \sum_i p_i \alpha_i^A$. We would like to give a lower bound on the average fidelity of this transmission in terms only of ρ. Such a bound can be accomplished (in the CP/QP formalisms) by extending ρ^A to a pure state $|\phi\rangle^{AR} \supseteq \rho^A$ and finding the fidelity of the resulting state with the original state when A is sent through the channel and R is left untouched[BKN00]. Here the reference system R is introduced to guarantee that transmitting system A preserves its entanglement with an arbitrary external system. Like the environment, R is always inaccessible and its properties are not changed by local unitary rotations. Indeed the only freedom in choosing $|\phi\rangle^{AR}$ is given by a local unitary rotation on R.

[*] For our purposes, we can think of Eve as a passive environment, but other work, for example on private communication[Dev05a, BOM04], treats Eve as an active participant who is trying to maximize her information. In these settings, we introduce private environments for Alice and Bob E_A and E_B, so that they can perform noisy operations locally without leaking information to Eve.

- *Source* (S) In most coding problems Alice can choose how she encodes the message, but cannot choose the message that she wants to communicate to Bob; it can be thought of as externally given. Taking this a step further, we can identify the source of the message as another protagonist (S), who begins a communication protocol by telling Alice which message to send to Bob. Introducing S is useful in cases when the Source does more than simply send a state to Alice; for example in distributed compression, the Source distributes a bipartite state to Alice and Bob.

To each party corresponds a class of quantum or classical systems which they control or have access to at different times. The systems corresponding to Alice are labeled by A (for example, A', A_1, X_A, etc.), while Bob's systems are labeled by B. When two classical systems, such as X_A and X_B, have the same principal label it means that they are instances of the same random variable. In our example, X_A is Alice's copy and X_B is Bob's copy of the random variable X.

We turn to some important examples of quantum states and operations. Let A, B, A', X_A and X_B be d-dimensional systems with respective distinguished bases $\{|x\rangle^A\}, \{|x\rangle^B\}$, etc. The standard maximally entangled state on AB is given by

$$|\Phi_d\rangle^{AB} = \frac{1}{\sqrt{d}} \sum_{x=1}^{d} |x\rangle^A |x\rangle^B.$$

The decohered, "classical", version of this state is

$$\overline{\Phi}_d^{X_A X_B} = \frac{1}{d} \sum_{x=1}^{d} |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_B},$$

which may be viewed as two maximally correlated random variables taking values on the set $[d] = \{1, \ldots, d\}$. The local restrictions of either of these states is the maximally mixed state $\tau_d := \frac{1}{d}\mathbb{1}_d$. (We write τ to remind us that it is also known as the *tracial state*.) Define the identity quantum operation $\mathrm{id}_d : A' \to B$ by the isometry $\sum_x |x\rangle^B \langle x|^{A'}$ (Note that this requires fixed bases of A' and B!). It represents a perfect quantum channel between the systems A' and B. Its classical counterpart is the completely dephasing channel $\overline{\mathrm{id}}_d : X_{A'} \to X_B$, given in the Kraus representation by $\{|x\rangle^{X_B} \langle x|^{X_{A'}}\}_{x \in [d]}$. It corresponds to a perfect classical channel in the sense that it perfectly transmits random variables, as represented by density operators diagonal in the preferred basis. The channel $\overline{\Delta}_d : X_{A'} \to X_A X_B$ with Kraus representation $\{|x\rangle^{X_B} |x\rangle^{X_A} \langle x|^{X_{A'}}\}_{x \in [d]}$ is a variation on $\overline{\mathrm{id}}_d$ in which Alice first makes a (classical) copy of the data before sending it through the classical channel. The two channels are essentially interchangeable. In Chapter 3 we will discuss the so-called *coherent* channel $\Delta_d : A' \to AB$, given by the isometry $\sum_x |x\rangle^A |x\rangle^B \langle x|^{A'}$ which is a coherent version of the noiseless classical channel with feedback, $\overline{\Delta}_d$. Here and in the following, "coherent" is meant to say that the operation preserves coherent quantum superpositions.

The maximally entangled state $|\Phi_d\rangle^{AB}$ and perfect quantum channel $\mathrm{id}_d : A' \to B$ are locally basis covariant: $(U \otimes U^*)|\Phi_d\rangle^{AB} = |\Phi_d\rangle^{AB}$ and $U^\dagger \circ \mathrm{id}_d \circ U = \mathrm{id}_d$ for any unitary U. On the other hand, $\overline{\Phi}_d$, $\overline{\mathrm{id}}_d$, $\overline{\Delta}_d$ and Δ_d are all locally basis-dependent.

1.2.2 Finite resources

In this subsection we introduce "finite" or "non-asymptotic" resources. They can be either static or dynamic, but strictly speaking, thanks to the appending maps $\mathcal{A}^\rho$, we only need to consider dynamic ones.

Definition 1.5 (Finite resources). *A* finite static resource *is a quantum state ρ^{AB}. A* finite dynamic resource *is an ordered pair $(\mathcal{N} : \omega)$, where the $\mathcal{N} : A'B' \to AB$ is an operation, with Alice's* *and Bob's input systems decomposed as $A' = A^{\mathrm{abs}}A^{\mathrm{rel}}$, $B' = B^{\mathrm{abs}}B^{\mathrm{rel}}$, and $\omega^{A^{\mathrm{rel}}B^{\mathrm{rel}}}$ is a so-called* test state.

The idea of the resource character of states and channels (static and dynamic, resp.) ought be clear. The only thing we need to explain is why we assume that $\mathcal{N}$ comes with a test state (contained in the "relative" systems $A^{\mathrm{rel}}B^{\mathrm{rel}}$): for finite resources it serves only a syntactic purpose — the operation "expects" an extension of ω as input, which will play a role for the definition of (valid) protocols below. The test state may not comprise the entire input to $\mathcal{N}$, in which case the remainder of the input comes from the systems $A^{\mathrm{abs}}B^{\mathrm{abs}}$.

If $A^{\mathrm{rel}}B^{\mathrm{rel}} = \emptyset$, we identify $(\mathcal{N} : \omega)$ with the *proper* dynamic resource $\mathcal{N}$. Note that $\mathcal{A}^{\rho}$ is always a proper dynamic resource, as it has no inputs.

A resource $(\mathcal{N} : \omega)$ is called *pure* if $\mathcal{N}$ is an isometry. It is called *classical* if $\mathcal{N}$ is a $\{c \to c\}$ entity and ω is a $\{c\}$ entity (though they may be expressed in the QQ formalism).

We define a distance measure between two dynamic resources $(\mathcal{N} : \omega)$ and $(\mathcal{N}' : \omega)$ with the same test state as

$$\|(\mathcal{N}' : \omega) - (\mathcal{N} : \omega)\| := \|\mathcal{N}' - \mathcal{N}\|_{\omega}.$$

A central notion is that of comparison between resources: we take the operational view that one finite resource, $(\mathcal{N}_1 : \omega_1)$, is stronger than another, $(\mathcal{N}_2 : \omega_2)$, in symbols $(\mathcal{N}_1 : \omega_1) \overset{*}{\geq} (\mathcal{N}_2 : \omega_2)$, if it the former can be used to perfectly simulate the latter. We demand first that there exist local operations $\mathcal{E}_A : A_2' \to A_1'$ and $\mathcal{D}_A : A_1 \to A_2$ for Alice, and $\mathcal{E}_B : B_2' \to B_1'$ and $\mathcal{D}_B : B_1 \to B_2$ for Bob, such that

$$\mathcal{N}_2 = (\mathcal{D}_A \otimes \mathcal{D}_B)\mathcal{N}_1(\mathcal{E}_A \otimes \mathcal{E}_B); \tag{1.5}$$

and second that the simulation be *valid*, meaning that for every $\zeta_1 \supset \omega_1$,

$$\zeta_2 := (\mathcal{E}_A \otimes \mathcal{E}_B)\zeta_1 \supset \omega_2. \tag{1.6}$$

When this occurs, we also say that $(\mathcal{N}_2 : \omega_2)$ *reduces to* $(\mathcal{N}_1 : \omega_1)$.

Two important properties of this relation are that

1. It is *transitive*; i.e. if $(\mathcal{N}_1 : \omega_1) \overset{*}{\geq} (\mathcal{N}_2 : \omega_2)$ and $(\mathcal{N}_2 : \omega_2) \overset{*}{\geq} (\mathcal{N}_3 : \omega_3)$, then $(\mathcal{N}_1 : \omega_1) \overset{*}{\geq} (\mathcal{N}_3 : \omega_3)$.

2. It is continuous; i.e. if $(\mathcal{N}_1 : \omega_1) \overset{*}{\geq} (\mathcal{N}_2 : \omega_2)$, then for any channel $\mathcal{N}_1'$ there exists $\mathcal{N}_2'$ such that $(\mathcal{N}_1' : \omega_1) \overset{*}{\geq} (\mathcal{N}_2' : \omega_2)$ and $\|\mathcal{N}_2' - \mathcal{N}_2\|_{\omega_2} \leq \|\mathcal{N}_1' - \mathcal{N}_1\|_{\omega_1}$.

The tensor product of states naturally extends to dynamic resources:

$$(\mathcal{N}_1 : \omega_1) \otimes (\mathcal{N}_2 : \omega_2) := (\mathcal{N}_1 \otimes \mathcal{N}_2 : \omega_1 \otimes \omega_2).$$

However, contrary to what one might expect $(\mathcal{N}_1 \otimes \mathcal{N}_2 : \omega_1 \otimes \omega_2) \overset{*}{\geq} (\mathcal{N}_1 : \omega_1)$ holds if and only if $\omega_1^{A_1B_1}$ can be perfectly mapped with local operations to a state $\omega^{A_1B_1A_2B_2}$ such that $\omega^{A_1B_1} = \omega_1$ and $\omega^{A_2B_2} = \omega_2$. Thus, we will almost always consider resoures where the test state ω is a product state. Nevertheless, nontrivial examples exist when the tensor product is stronger than its component resources; for example, $(\mathcal{N} : \omega)^{\otimes 2}$ when ω is a classically correlated state.

A more severe limitation on these resource comparisons is that they do not allow for small errors or inefficiencies. Thus, most resources are incomparable and most interesting coding theorems do not yield useful exact resource inequalities. We will address these issues in the next section when we define asymptotic resources and asymptotic resource inequalities.

Resources as above are atomic primitives: "having" such a resource means (given an input state) the ability to invoke the operation (once). When formalizing the notion of "having" several resources, e.g., the choice from different channels, it would be too restrictive to model this by the tensor product, because it gives us just another resource, which the parties have to use in a sort of "block code". To allow for — finite — recursive depth (think, e.g., of feedback, where future channel uses depend on the past ones) in using the resources, we introduce the following:

Definition 1.6 (Depth-ℓ resources). *A finite depth-ℓ resource is an unordered collection of, w.l.o.g., dynamic resources*

$$(\mathcal{N}:\omega)^\ell := \big((\mathcal{N}_1:\omega_1),\ldots,(\mathcal{N}_\ell:\omega_\ell)\big).$$

Both static and dynamic resources are identified with depth-1 resources. To avoid notational confusion, for ℓ copies of the same dynamic resource, $\big((\mathcal{N}:\omega),\ldots,(\mathcal{N}:\omega)\big)$, we reserve the notation $(\mathcal{N}:\omega)^{\times\ell}$.

The definition of the distance measure naturally extends to the case of two depth-ℓ resources:

$$\|(\mathcal{N}':\omega)^\ell-(\mathcal{N}:\omega)^\ell\| := \min_{\pi\in\mathcal{S}_\ell,\omega_j=\omega_{\pi(j)}\forall j}\sum_{j\in[\ell]}\|(\mathcal{N}'_j:\omega_j)-(\mathcal{N}_{\pi(j)}:\omega_{\pi(j)})\|.$$

Here $\mathcal{S}_\ell$ is the set of permutations on ℓ objects; we need to minimize over it to reflect the fact that we're free to use depth-ℓ resources in an arbitrary order.

To *combine* resources there is no good definition of a tensor product (which operations should we take the products of?), but we can take tensor *powers* of a resource:

$$\big((\mathcal{N}:\omega)^\ell\big)^{\otimes k} := \big((\mathcal{N}_1:\omega_1)^{\otimes k},\ldots,(\mathcal{N}_\ell:\omega_\ell)^{\otimes k}\big).$$

The way we combine a depth-ℓ and a depth-ℓ' resource is by concatenation: let

$$(\mathcal{N}:\omega)^\ell+(\mathcal{N}':\omega')^{\ell'} := \big((\mathcal{N}_1:\omega_1),\ldots,(\mathcal{N}_\ell:\omega_\ell),(\mathcal{N}'_1:\omega'_1),\ldots,(\mathcal{N}'_{\ell'}:\omega'_{\ell'})\big).$$

We now have to extend the concept of one resource simulating another to depth-ℓ; at the same time we will introduce the notions of approximation that will become essential for the asymptotic resources below.

Definition 1.7 (Elementary protocols). *An* elementary protocol **P** *takes a depth-ℓ finite resource $(\mathcal{N}:\omega)^\ell$ to a depth-1 finite resource. Given $\mathcal{N}_i: A'_iB'_i\to A_iB_i$ and test states ${\omega_i}^{A_i^{\rm rel}B_i^{\rm rel}}$, $i=1\ldots\ell$, $\mathbf{P}[(\mathcal{N}:\omega)^\ell]$ is a finite depth-1 resource $(\mathcal{P}:\Omega^{A^{\rm rel}B^{\rm rel}})$, with a quantum operation $\mathcal{P}: A'B'\to AB$, which is constructed as follows:*[*]

1. *select a permutation π of the integers $\{1,\ldots,\ell\}$;*
2. *perform local operations $\mathcal{E}_0: A'\to A_0A_0^{\rm aux}$ and $\mathcal{E}'_0: B'\to B_0B_0^{\rm aux}$;*
3. *repeat, for $i=1,\ldots,\ell$,*
   - *(a)$_i$ perform local isometries $\mathcal{E}_i: A_{i-1}A_{i-1}^{\rm aux}\to A'_iA_i^{\rm aux}$ and $\mathcal{E}'_i: B_{i-1}B_{i-1}^{\rm aux}\to B'_iB_i^{\rm aux}$;*
   - *(b)$_i$ apply the operation $\mathcal{N}_{\pi(i)}$, mapping $A'_iB'_i$ to A_iB_i;*
4. *perform local operations $\mathcal{E}_{\ell+1}: A'_\ell A_\ell^{\rm aux}\to A$ and $\mathcal{E}'_{\ell+1}: B'_\ell B_\ell^{\rm aux}\to B$.*

[*] We use diverse notation to emphasize the role of the systems in question. The primed systems, such as A'_i, are channel inputs. The systems with no superscript, such as B_i, are channel outputs. Some systems are associated with Alice's sources (e.g. $A_i^{\rm rel}$) and Bob's possible side information about those sources (e.g. $B_i^{\rm rel}$). Furthermore, there are auxiliary systems, such as $A_i^{\rm aux}$.

We allow the arbitrary permutation of the resources π so that depth-ℓ resources do not have to be used in a fixed order. Denote by $\mathcal{P}_i$ the operation of performing the protocol up to, but not including, step 3.(b)$_i$. Define $\hat{\mathcal{P}}_i$ to be $\mathcal{P}_i$ followed by a restriction onto $A_i^{\mathrm{rel}}B_i^{\mathrm{rel}}$. The protocol **P** *is called* η-valid *on the input finite resource $(\mathcal{N}:\omega)^l$ if the conditions*

$$\|\hat{\mathcal{P}}_i(\Omega) - \omega_{\pi(i)}^{A_i^{\mathrm{rel}}B_i^{\mathrm{rel}}}\|_1 \leq \eta$$

are met for all i.

Definition 1.8 (Standard protocol). *Define the* standard protocol **S***, which is a 0-valid elementary protocol on a depth-ℓ finite resource $(\mathcal{N}:\omega)^\ell$, by*

$$\mathbf{S}[(\mathcal{N}:\omega)^l] = (\bigotimes_{i=1}^{l} \mathcal{N}_i : \bigotimes_{i=1}^{\ell} \omega_i).$$

That is, this protocol takes a list of resources, and flattens them into a depth-1 tensor product.

Whenever $(\mathcal{N}:\omega) \overset{*}{\geq} (\mathcal{N}':\omega')$, there is a natural protocol **R**, which is 0-valid on $(\mathcal{N}:\omega)$, *implementing* the reduction:

$$\mathbf{R}[(\mathcal{N}:\omega)] = (\mathcal{N}':\omega'),$$

which we write as

$$\mathbf{R} : (\mathcal{N}:\omega) \overset{*}{\geq} (\mathcal{N}':\omega').$$

For resources with depth > 1, $(\mathcal{N}:\omega)^\ell = ((\mathcal{N}_1:\omega_1),\ldots,(\mathcal{N}_\ell:\omega_\ell))$ and $(\mathcal{N}':\omega')^{\ell'} = ((\mathcal{N}_1':\omega_1'),\ldots,(\mathcal{N}_{\ell'}':\omega_{\ell'}'))$, we say that $(\mathcal{N}:\omega) \overset{*}{\geq} (\mathcal{N}':\omega')$ if there exists an injective function $f:[\ell'] \to [\ell]$ such that for all $i \in [\ell']$, $(\mathcal{N}_{f(i)}:\omega_{f(i)}) \overset{*}{\geq} (\mathcal{N}_i':\omega_i')$. In other words, for each $(\mathcal{N}_i':\omega_i')$ there is a unique $(\mathcal{N}_j:\omega_j)$ that reduces to $(\mathcal{N}_i':\omega_i')$. Note that this implies $\ell \geq \ell'$. Again there is a natural 0-valid protocol **R** implementing the reduction.

The next two lemmas help justify aspects of our definition of a protocol—η-validity and the fact that outputs are depth-1—that will later be crucial in showing how protocols may be composed.

First we show why η-validity is important. In general we want our distance measures for states to satisfy the triangle inequality, and to be nonincreasing under quantum operations. These properties guarantee that the error of a sequence of quantum operations is no more than the sum of errors of each individual operation (cf. part 4 of Lemma 1.4 as well as [BV93]). However, this assumes that we are using the same distance measure throughout the protocol; when working with relative resources, a small error with respect to one input state may be much larger for a different input state. Thus, for a protocol to map approximately correct inputs to approximately correct outputs, we need the additional assumption that the protocol is η-valid.

Lemma 1.9 (Continuity). *If some elementary protocol* **P** *is η-valid on $[(\mathcal{N}:\omega)^\ell]$ and*

$$\|(\mathcal{N}:\omega)^\ell - (\mathcal{A}:\omega)^\ell\| \leq \epsilon,$$

then

$$\|\mathbf{P}[(\mathcal{N}:\omega)^\ell] - \mathbf{P}[(\mathcal{A}:\omega)^\ell]\| \leq l(\epsilon + 2\sqrt{\eta})$$

and $\mathbf{P}[(\mathcal{A}:\omega)^\ell]$ is $(\eta + \ell(\epsilon + 2\sqrt{\eta}))$-valid.

Proof. Let $(\mathcal{P}:\Omega) = \mathbf{P}[(\mathcal{N}:\omega)^\ell]$ and $(\mathcal{P}':\Omega) = \mathbf{P}[(\mathcal{A}:\omega)^\ell]$. By definition 1.7, $\mathcal{P}$ is of the form

$$\mathcal{P} = \mathcal{E}_{\ell+1} \circ \mathcal{N}_\ell \circ \mathcal{E}_\ell \circ \cdots \circ \mathcal{N}_1 \circ \mathcal{E}_1$$

and similarly for $\mathcal{P}'$. The η-validity condition reads, for all i,

$$\|\hat{\mathcal{P}}_i(\Omega) - \omega_i\|_1 \leq \eta.$$

By part 3 of Lemma 1.4,

$$\|\mathcal{P} - \mathcal{P}'\|_\Omega \leq \sum_i \|\mathcal{A}_i - \mathcal{N}_i\|_{\mathcal{P}_i(\Omega)}.$$

By part 1 of Lemma 1.4,

$$\|\mathcal{A}_i - \mathcal{N}_i\|_{\mathcal{P}_i(\Omega)} \leq \|\mathcal{A}_i - \mathcal{N}_i\|_{\hat{\mathcal{P}}_i(\Omega)}.$$

By part 2 of Lemma 1.4 and η-validity

$$\|\mathcal{A}_i - \mathcal{N}_i\|_{\hat{\mathcal{P}}_i(\Omega)} \leq \|\mathcal{A}_i - \mathcal{N}_i\|_{\omega_i} + 2\sqrt{\eta}$$

Hence

$$\|\mathcal{P} - \mathcal{P}'\|_\Omega \leq \ell(\epsilon + 2\sqrt{\eta}),$$

which is one of the statements of the lemma. To estimate the validity of $\mathbf{P}$ on $[(\mathcal{A} : \omega)^\ell]$, note that one obtains in the same way as above, for all i,

$$\|\hat{\mathcal{P}}_i - \hat{\mathcal{P}}'_i\|_\Omega \leq \ell(\epsilon + 2\sqrt{\eta}).$$

Combining this with the η-validity condition via the triangle inequality finally gives

$$\|\hat{\mathcal{P}}'_i(\Omega) - \omega_i\|_1 \leq \eta + \ell(\epsilon + 2\sqrt{\eta}),$$

concluding the proof. □

We note that we do not have a concept of what it means to turn a depth-ℓ resource into a depth-ℓ' resource; instead, our basic concept of simulation produces a depth-1 resource. I.e., we can formulate what it means that a depth-ℓ resource simulates the standard protocol of a depth-ℓ' resource.

The following lemma states that the standard protocol is basically sufficient to generate any other, under some i.i.d.-like assumptions.

Lemma 1.10 (Sliding). *If for some depth-ℓ finite resource $(\mathcal{N} : \omega)^\ell = ((\mathcal{N}_1 : \omega_1), \ldots, (\mathcal{N}_\ell : \omega_\ell))$ and quantum operation $\mathcal{C}$,*

$$\|(\mathcal{C} : \bigotimes_i \omega_i) - \mathbf{S}[(\mathcal{N} : \omega)^\ell]\| \leq \epsilon, \tag{1.7}$$

then for any integer $m \geq 1$ and for any η- valid protocol $\mathbf{P}$ on $(\mathcal{N} : \omega)^\ell$, there exists a $((m + \ell - 1)(\epsilon + 2\sqrt{\eta}) + \eta)$-valid protocol $\mathbf{P}'$ on $(\mathcal{C} : \bigotimes_i \omega_i)^{\times(m+\ell-1)}$, such that

$$\|\mathbf{P}'[(\mathcal{C} : \bigotimes_i \omega_i)^{\times(m+\ell-1)}] - (\mathbf{P}[(\mathcal{N} : \omega)^\ell])^{\otimes m}\| \leq (m + \ell - 1)(\epsilon + 2\sqrt{\eta}).$$

Proof. Denoting by $\mathbf{P}'$ the sliding protocol (see Fig. 1-1) it is clear that

$$\mathbf{P}'[(\mathbf{S}[(\mathcal{N} : \omega)^\ell])^{\times(m+\ell-1)}] = (\mathbf{P}[(\mathcal{N} : \omega)^\ell])^{\otimes m}.$$

The result follows from Lemma 1.9. □

The sliding protocol shows how working with depth-1 resources is not overly restrictive. Another difficulty with resources is that relative resources are only guaranteed to work properly when given the right sort of input state. Here we show that using shared randomness, some of the standard relative resources can be "absolutized," removing the restriction to a particular input state.

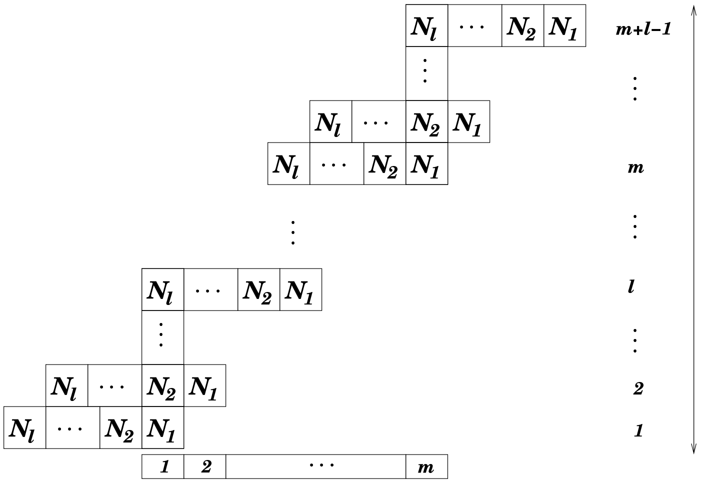


Figure 1-1: The sliding protocol. We would like to simulate $\mathbf{P}$, which uses $\mathcal{N}_1, \ldots, \mathcal{N}_\ell$ consecutively, but we are only given $\mathcal{N}_1 \otimes \ldots \mathcal{N}_\ell$. The horizontal blocks represent uses of $\mathcal{N}_1 \otimes \ldots \otimes \mathcal{N}_\ell$ and stacking them vertically indicates how we perform them consecutively with the output of one block becoming the input of the block above it (i.e. time flows from the bottom to the top). Thus $m+l-1$ consecutive uses of $\mathcal{N}_1 \otimes \ldots \otimes \mathcal{N}_\ell$ can simulate m copies of $\mathbf{P}$.

Lemma 1.11. *For a operation $\mathcal{N} : A' \to AB$ which is either the perfect quantum channel id_d, the coherent channel Δ_d or the perfect classical channel $\overline{\mathrm{id}}_d$, there exists a 0-valid protocol $\mathbf{P}$ such that*

$$\mathbf{P}[\overline{\Phi}^{X_A X_B}, (\mathcal{N} : \tau^{A'})] = \mathcal{N} \otimes \mathcal{A}^{\overline{\Phi}^{X_A X_B}},$$

where $\dim X_A = (\dim A')^2$, and $\tau^{A'}$ is the maximally mixed state on A'.

Proof. Consider first the case of $\mathcal{N}$ being either id_d or the coherent channel Δ_d. The main observation is that there exist a set of unitary operations $\{U_x\}_{x \in [d^2]}$ (the generalized Pauli, or discrete Weyl, operators) such that, for any state ρ living on a d-dimensional Hilbert space,

$$d^{-2} \sum_x U_x \rho U_x^\dagger = \tau_d, \tag{1.8}$$

with τ_d being the maximally mixed state on that space.

Let Alice and Bob share the common randomness state

$$\overline{\Phi}^{X_A X_B} = d^{-2} \sum_{x=1}^{d^2} |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_B},$$

where $d := \dim A'$. Consider an arbitrary input state $|\phi\rangle^{RA'}$, possibly entangled between Alice and a reference system R. Alice performs the conditional unitary $\sum_x |x\rangle\langle x|^{X_A} \otimes U_x^{A'}$, yielding a state whose restriction to A' is maximally mixed. She then applies the operation $\mathcal{N}$ (this is 0-valid!), which gives the state

$$d^{-2} \sum_{x=1}^{d^2} |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_B} \otimes (\mathcal{N} \circ U_x^{A'}) \phi^{RA'}.$$

In the case of the id_d channel, Bob simply applies the conditional unitary $\sum_x |x\rangle\langle x|^{X_B} \otimes (U_x^{-1})^B$. In the case of the Δ_d channel Alice must also perform

$$\sum_x |x\rangle\langle x|^{X_A} \otimes (U_x^{-1})^A.$$

Either way, the final state is

$$\overline{\Phi}^{X_A X_B} \otimes \mathcal{N}(\phi^{RA'}),$$

as advertised.

The case of the perfect classical channel $\overline{\mathrm{id}}_d$ is a classical analogue of the above. The observation here is that there exists a set of d unitaries $\{U_x\}_{x\in[d]}$ (all the cyclic permutations of the basis vectors), each member of which commutes with Δ, such that (1.8) holds for any state ρ diagonal in the preferred basis. Now Alice first applies a local Δ (diagonalizing the input), before proceeding as above. This concludes the proof. □

Observe that in the above lemma, the final output of $\mathcal{N}$ is uncorrelated with the shared randomness that is used. In the QQ formalism, this is immediately apparent from the tensor product between $\mathcal{N}$ and $\mathcal{A}^{\overline{\Phi}^{X_A X_B}}$. Thus we say that the shared randomness is (incoherently) *decoupled* from the rest of the protocol.

Now consider the case when $\mathcal{N} = \overline{\mathrm{id}}_d$ and use the QP formalism, so $\mathcal{N}$ is a map from A to BE. If we condition on a particular message sent by Alice, then the randomness is no longer decoupled from the composite BE system. This is the problem of reusing the key in a one-time pad: if the message is not uniformly random, then information about the key leaks to Eve.

On the other hand, if $\mathcal{N}$ is Δ_d or id_d then the shared randomness is decoupled even from the environment. This stronger form of decoupling is called *coherent decoupling*. Below we give formal definitions of these notions of decoupling.[*]

Definition 1.12 (Incoherent decoupling). *Consider a protocol* $\mathbf{P}$ *on* $((\overline{\mathcal{N}} : \overline{\omega})^\ell, (\mathcal{N} : \omega)^{\ell'})$, *where* $(\overline{\mathcal{N}} : \overline{\omega})^\ell$ *is classical. Recall that in the QQ formalism classical systems are unchanged under the copying operation* $\overline{\Delta}$. *This means we can consider an equivalent protocol in which the systems associated with the classical resource* $(\overline{\mathcal{N}} : \overline{\omega})^\ell$ *are copied into a composite classical system* Z, *which includes all the copies of all the random variables involved. Let* $\mathbf{P}'$ *be the modified version of* $\mathbf{P}$ *which retains* Z *in the final state. Now* $\mathcal{P}' := \mathbf{P}'[((\overline{\mathcal{N}} : \overline{\omega})^\ell, (\mathcal{N} : \omega)^{\ell'})] \supseteq \mathcal{P}$ *takes a particular extension* $\Upsilon^{RA'A^*B^*} \supseteq \Omega^{A^*B^*}$ *to some state* $\sigma^{ZRABA^*B^*}$.

We say that the classical resource $(\overline{\mathcal{N}} : \overline{\omega})^\ell$ *is* $\epsilon-$incoherently decoupled *(or just* $\epsilon-$decoupled*) with respect to the protocol* $\mathbf{P}$ *on* $((\overline{\mathcal{N}} : \overline{\omega})^\ell, (\mathcal{N} : \omega)^{\ell'})$ *if for any* $\Upsilon^{RA'A^*B^*}$ *the state* $\sigma^{ZRABA^*B^*}$ *satisfies*

$$\|\sigma^{ZRABA^*B^*} - \sigma^Z \otimes \sigma^{RABA^*B^*}\|_1 \le \epsilon. \tag{1.9}$$

We describe separately how classical resources used in the input and the output of a protocol may be coherently decoupled.

Definition 1.13 (Coherent decoupling of input resources). *Again, consider a protocol* $\mathbf{P}$ *on* $((\overline{\mathcal{N}} : \overline{\omega})^\ell, (\mathcal{N} : \omega)^{\ell'})$, *where* $(\overline{\mathcal{N}} : \overline{\omega})^\ell$ *is classical. Now we adopt a QP view in which all non-classical states are purified and all channels are isometrically extended. Again, we define a classical system* Z *which contains copies of all the classical variables associated with the resource* $(\mathcal{N} : \omega)^{\ell'}$. *The final state of the protocol is then some* $\sigma^{ZRABA^*B^*E}$. *We say that the classical resource* $(\overline{\mathcal{N}} : \overline{\omega})^\ell$ *is* $\epsilon-$coherently decoupled *with respect to the protocol* $\mathbf{P}$ *on* $((\overline{\mathcal{N}} : \overline{\omega})^\ell, (\mathcal{N} : \omega)^{\ell'})$ *if for any* $\Upsilon^{RA'A^*B^*}$ *the final state* $\sigma^{ZRABA^*B^*E}$ *satisfies*

$$\|\sigma^{ZRABA^*B^*E} - \sigma^Z \otimes \sigma^{RABA^*B^*E}\|_1 \le \epsilon.$$

[*] The notion of an "oblivious" protocol for remotely preparing quantum states is similar to coherent decoupling, but applies instead to quantum messages[LS03].

Definition 1.14 (Coherent decoupling of output resources). *Remaining within the QP formalism, let* $\mathbf{P}$ *be a protocol mapping* $(\mathcal{N} : \omega)^\ell$ *to* $(\overline{\mathcal{P}}_1 \otimes \mathcal{P}_2 : \overline{\Omega}_1^{A_1} \otimes \Omega_2^{A_2B_2})$*; i.e. the tensor product of a classical resource* $(\overline{\mathcal{P}}_1 : \overline{\Omega}_1^{A_1})$ *and a quantum resource* $(\mathcal{P}_2 : \Omega_2^{A_2B_2})$*. Define* Z *to consist of copies of* A_1B_1 *together with all the other classical resources associated with* $\overline{\mathcal{P}}_1$*, such as outputs (if different from* A_1*) and inputs other than* A_1 *(if any).*

We now say that the classical resource $(\mathcal{P}_1 : \Omega_1)$ *is* $\epsilon-$coherently decoupled *with respect to the protocol* $\mathbf{P}$ *on* $(\mathcal{N} : \omega)^\ell$ *if*

$$\|\sigma^{ZQ} - \sigma^Z \otimes \sigma^Q\|_1 \leq \epsilon,$$

where now Q *comprises all the quantum systems involved (including environments and reference systems).*

We will give some applications of decoupling in Section 1.3, but its primary utility will be seen in Chapters 3 and 4.

One simple example of decoupling is when a protocol involves several pure resources (i.e. isometries) and one noiseless classical resource. In this case, decoupling the classical resource is rather easy, since pure resources don't involve the environment. However, it is possible that the classical communication is correlated with the ancilla system Q that Alice and Bob are left with. If Q is merely discarded, then the cbits will be incoherently decoupled. To prove that coherent decoupling is in fact possible, we will need to carefully account for the ancillas produced by the classical communication. This will be accomplished in Section 3.5, where we prove that classical messages sent through isometric channels can always be coherently decoupled.

1.2.3 Asymptotic resources

Definition 1.15 (Asymptotic resources). *An* asymptotic resource α *is defined by a sequence of finite depth-*ℓ *resources* $(\alpha_n)_{n=1}^\infty$*, where* α_n *is w.l.o.g. of the form* $\alpha_n = (\mathcal{N}_n : \omega_n)^\ell := ((\mathcal{N}_{n,1} : \omega_{n,1}), (\mathcal{N}_{n,2} : \omega_{n,2}), \ldots, (\mathcal{N}_{n,\ell} : \omega_{n,\ell}))$*, such that*

- $$\alpha_n \overset{*}{\geq} \alpha_{n-1} \quad \textit{for all} n; \tag{1.10}$$
- *for any* $\delta > 0$*, any integer* k *and all sufficiently large* n*,*
$$\alpha_{\lfloor n(1+\delta) \rfloor} \overset{*}{\geq} (\alpha_{\lfloor n/k \rfloor})^{\otimes k} \overset{*}{\geq} \alpha_{\lfloor n(1-\delta) \rfloor}. \tag{1.11}$$

We sometimes refer to this as the requirement that a resource be "quasi-i.i.d."

Denote the set of asymptotic resources by $\mathcal{R}$.

Given two resources $\alpha = (\alpha_n)_{n=1}^\infty$ and $\beta = (\beta_n)_{n=1}^\infty$, if $\alpha_n \overset{*}{\geq} \beta_n$ for all sufficiently large n, then we write $\alpha \overset{*}{\geq} \beta$. We shall use the following convention: if $\beta = (\mathcal{N}_n)_n$, where all $\mathcal{N}_n$ are proper dynamic resources and $\gamma = (\omega_n)_n$, where all ω_n are proper static resources, then $(\beta : \gamma) := (\mathcal{N}_n : \omega_n)_n$. Note that typically ω_n is product state, so the resource γ reduces to the null resource $\emptyset$; however this is no problem as long as we are interested in γ only as a test state for β.

Our next goal is to define what it means to simulate one (asymptotic) resource by another.

Definition 1.16 (Asymptotic resource inequalities). *A resource* inequality $\alpha \geq \beta$ *holds between two resources* $\alpha = (\alpha_n)_n$ *and* $\beta = (\beta_n)_n$ *if for any* $\delta > 0$ *there exists an integer* k *such that for any* $\epsilon > 0$ *there exists* N *such that for all* $n \geq N$ *there exists an* ϵ*-valid protocol* $\mathbf{P}^{(n)}$ *on* $(\alpha_{\lfloor n/k \rfloor})^{\times k}$ *(i.e.* k *sequential uses of* $\alpha_{\lfloor n/k \rfloor}$*) for which*

$$\|\mathbf{P}^{(n)}[(\alpha_{\lfloor n/k \rfloor})^{\times k}] - \mathbf{S}[\beta_{\lfloor (1-\delta)n \rfloor}]\| \leq \epsilon.$$

α is called the input resource, β is called the output resource, δ is the inefficiency (or sometimes the fractional inefficiency) and ϵ (which bounds both the validity and the error) is called the accuracy (or sometimes just the error).

At first glance it may seem that we are demanding rather little from asymptotic resource inequalities: we allow the depth of the input resource to grow arbitrarily, while requiring only a depth-1 output. However, later in this section we will use tools like the sliding lemma to show that this definition is nevertheless strong enough to allow the sort of protocol manipulations we would like.

Also, for resources that consist entirely of states one-way channels, it is never necessary to use protocols with depth > 1. Thus, we state here a "flattening" lemma that will later be useful in proving converses; i.e. statements about when certain resource inequalities are impossible.

Lemma 1.17 (Flattening). *Suppose $\alpha \geq \beta$ and α is a "one-way" resource, meaning that it consists entirely of static resources ($\mathcal{A}^\rho$) and dynamic resources which leave nothing on Alice's side (e.g. $\mathcal{N}^{A' \to BE}$). Then for any $\epsilon, \delta > 0$ for sufficiently large n there is an ϵ-valid protocol $\mathbf{P}^{(n)}$ on α_n such that*

$$\|\mathbf{P}^{(n)}[\alpha_n] - \mathbf{S}[\beta_{\lfloor (1-\delta)n \rfloor}]\| \leq \epsilon.$$

Proof. To prove the lemma, it will suffice to convert a protocol on $(\alpha_{\lfloor n/k \rfloor})^{\times k}$ to a protocol on $(\alpha_{\lfloor n/k \rfloor})^{\otimes k}$. Then we can use the fact that $\alpha_{\lfloor n(1+\delta) \rfloor} \overset{*}{\geq} (\alpha_{\lfloor n/k \rfloor})^{\otimes k}$ and the lemma follows from a suitable redefinition of n and δ.

Since α is a one-way resource, any protocol that uses it can be assumed to be of the following form: first Alice applies all of the appending maps, then she does all of her local operations, then she applies all of the dynamic resources, and finally Bob does his decoding operations. The one-way nature of the protocol means that Bob can wait until all of Alice's operations are finished before he starts decoding. It also means that Alice can apply the dynamic resources last, since they have no outputs on her side, so none of her other operations can depend on them. Finally, the appending maps can be pushed to the beginning because they have no inputs. Thus $(\alpha_{\lfloor n/k \rfloor})^{\times k}$ can be simulated using $(\alpha_{\lfloor n/k \rfloor})^{\otimes k}$, completing the proof. □

Definition 1.18 (i.i.d. resources). *A resource α is called* independent and identically distributed (i.i.d.) *if $\alpha_n = (\mathcal{N}^{\otimes n} : \omega^{\otimes n})$ for some state ω and operation $\mathcal{N}$. We use shorthand notation $\alpha = \langle \mathcal{N} : \omega \rangle$.*

We shall use the following notation for *unit* asymptotic resources:

- ebit $[q\,q] := \langle \Phi_2 \rangle$
- rbit $[c\,c] := \langle \overline{\Phi}_2 \rangle$
- qubit $[q \to q] := \langle \mathrm{id}_2 \rangle$
- cbit $[c \to c] := \langle \overline{\mathrm{id}}_2 \rangle$
- cobit $[\![c \to c]\!] := \langle \Delta_2 \rangle$ (cobits will be explained in Chapter 3)

In this thesis, we tend to use symbols for asymptotic resource inequalities (e.g. "$\langle \mathcal{N} \rangle \geq C[c \to c]$") and words for finite protocols (e.g. "$\mathcal{N}^{\otimes n}$ can be used to send $\geq n(C - \delta_n)$ cbits with error $\leq \epsilon_n$"). However, there is no formal reason that they cannot be used interchangeably.

We also can define versions of the dynamic resources with respect to the standard "reference" state $\tau_2^{A'} = \mathbb{1}_2^{A'}/2$: a qubit in the maximally mixed state. These are denoted as follows:

- $[q \to q : \tau] := \langle \mathrm{id}_2 : \tau_2 \rangle$
- $[c \to c : \tau] := \langle \overline{\mathrm{id}}_2 : \tau_2 \rangle$

- $[\![c \to c : \tau]\!] := \langle \Delta_2 : \tau_2 \rangle$

Definition 1.19 (Addition). *The addition operation* $+ : \mathcal{R} \times \mathcal{R} \to \mathcal{R}$ *is defined for* $\alpha = (\alpha_n)_n$, $\alpha_n = ((\mathcal{N}_{n,1} : \omega_{n,1}), \ldots, (\mathcal{N}_{n,l} : \omega_{n,l}))$, *and* $\beta = (\beta_n)_n$, $\beta_n = ((\mathcal{N}'_{n,1} : \omega'_{n,1}), \ldots, (\mathcal{N}'_{n,l'} : \omega'_{n,l'}))$, *as* $\alpha + \beta = (\gamma_n)_n$ *with*

$$\gamma_n = (\alpha_n, \beta_n) := ((\mathcal{N}_{n,1} : \omega_{n,1}), \ldots, (\mathcal{N}_{n,l} : \omega_{n,l}), (\mathcal{N}'_{n,1} : \omega'_{n,1}), \ldots, (\mathcal{N}'_{n,l'} : \omega'_{n,l'})).$$

Closure is trivially verified. It is also easy to see that the operation $+$ is associative and commutative. Namely,

1. $\alpha + \beta = \beta + \alpha$
2. $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$

Definition 1.20 (Multiplication). *The multiplication operation* $\cdot : \mathcal{R} \times \mathbb{R}_+ \to \mathcal{R}$ *is defined for any positive real number* z *and resource* $\alpha = (\alpha_n)_n$ *by* $z\alpha = (\alpha_{\lfloor zn \rfloor})_n$.

Of course, we need to verify that $\mathcal{R}$ is indeed closed under multiplication. Define $\beta := z\alpha$, so that $\beta_n = \alpha_{\lfloor zn \rfloor}$. We know, for all sufficiently large n, that

$$\alpha_{\lfloor \lfloor zn \rfloor (1+\delta) \rfloor} \overset{*}{\geq} (\alpha_{\lfloor \lfloor zn \rfloor / k \rfloor})^{\otimes k} \overset{*}{\geq} \alpha_{\lfloor \lfloor zn \rfloor (1-\delta) \rfloor}.$$

We need to prove

$$\alpha_{\lfloor z \lfloor n(1+\delta') \rfloor \rfloor} \overset{*}{\geq} (\alpha_{\lfloor z \lfloor n/k \rfloor \rfloor})^{\otimes k} \overset{*}{\geq} \alpha_{\lfloor z \lfloor n(1-\delta') \rfloor \rfloor},$$

which is true for the right δ'.

Definition 1.21 (Asymptotic decoupling). *Consider a resource inequality of the form* $\alpha + \gamma \geq \beta$, *or* $\alpha \geq \gamma$, *where* γ *is a classical resource, and* α *and* β *are quantum resources. In either case, if in the definition above, for each sufficiently large* n *we also have that* γ_n *is* $\epsilon-$*(coherently) decoupled with respect to* $\mathbf{P}^{(n)}$, *then we say that* γ *is (coherently) decoupled in the resource inequality.*

The central purpose of our resource formalism is contained in the following "composability" theorem, which states that resource inequalities can be combined via concatenation and addition. In other words, the source of a resource (like cbits) doesn't matter; whether they were obtained via a quantum channel or a carrier pigeon, they can be used equally well in any protocol that takes cbits as an input. A well-known example of composability in classical information theory is Shannon's joint source-channel coding theorem which states that a channel with capacity $\geq C$ can transmit any source with entropy rate $\leq C$; the coding theorem is proved trivially by composing noiseless source coding and noisy channel coding.

Theorem 1.22 (Composability). *For resources in* $\mathcal{R}$*:*

1. *if* $\alpha \geq \beta$ *and* $\beta \geq \gamma$ *then* $\alpha \geq \gamma$
2. *if* $\alpha \geq \beta$ *and* $\gamma \geq \varepsilon$ *then* $\alpha + \gamma \geq \beta + \varepsilon$
3. *if* $\alpha \geq \beta$ *then* $z\alpha \geq z\beta$

Proof. 1. Fix $\delta > 0$. Then there exist k, k', such that for any ϵ and sufficiently large n

$$\|\mathbf{P}_1[(\alpha_{\lfloor n(1-\delta)/(mkk') \rfloor})^{\times k}] - \mathbf{S}[\beta_{\lfloor n(1-2\delta)/(mk') \rfloor}]\| \leq \epsilon, \tag{1.12}$$

$$\|\mathbf{P}_2[(\beta_{\lfloor n(1-2\delta)/(mk') \rfloor})^{\times k'}] - \mathbf{S}[\gamma_{\lfloor n(1-3\delta)/m \rfloor}]\| \leq \epsilon, \tag{1.13}$$

$$\gamma^{\otimes m}_{\lfloor n(1-3\delta)/m \rfloor} \overset{*}{\geq} \gamma_{\lfloor n(1-4\delta) \rfloor}, \tag{1.14}$$

with $m \geq k'l/\delta$, where l is the depth of β, and where $\mathbf{P}_1$ and $\mathbf{P}_2$ are both ϵ-protocols. Equation (1.14) implies the existence of a reduction protocol

$$\mathbf{R}_1 : \mathbf{S}[\gamma_{\lfloor n(1-3\delta)/m\rfloor}]^{\otimes m} \overset{*}{\geq} \mathbf{S}[\gamma_{\lfloor n(1-4\delta)\rfloor}].$$

By Eq. (1.13)

$$\|\mathbf{P}_2[(\beta_{\lfloor n(1-2\delta)/(mk')\rfloor})^{\times k'}]^{\otimes m} - \mathbf{S}[\gamma_{\lfloor n(1-3\delta)/m\rfloor}]^{\otimes m}\| \leq m\epsilon. \tag{1.15}$$

Define $\iota = \mathbf{P}_1[(\alpha_{\lfloor n(1-\delta)/(mkk')\rfloor})^{\times k}]^{\otimes k'}$, which, by Eq. (1.12), satisfies

$$\|\iota - \mathbf{S}[(\beta_{\lfloor n(1-2\delta)/(mk')\rfloor})^{\times k'}]\| \leq k'\epsilon. \tag{1.16}$$

Let $\epsilon' = (m + k'l - 1)(k'\epsilon + 2\sqrt{\epsilon}) + \epsilon$. We shall exhibit an ϵ'-valid protocol $\mathbf{P}_3$ such that

$$\|\mathbf{P}_3[\alpha_n] - \mathbf{S}[\gamma_{\lfloor n(1-4\delta)\rfloor}]\| \leq \epsilon' + m\epsilon. \tag{1.17}$$

By Eq. (1.11), there is a reduction $\mathbf{R}'$ from the initial finite resource α_n to $(\alpha_{\lfloor n(1-\delta)/(mkk')\rfloor})^{\times\lfloor mkk'(1+\delta)\rfloor}$, which in turn suffices to implement $\iota^{\times m+k'l-1}$. By the Sliding Lemma (1.10) and Eq. (1.16), there exists some ϵ'-valid protocol $\mathbf{P}'$ such that

$$\|\mathbf{P}'[\iota^{\times m+k'l-1}] - \mathbf{P}_2[(\beta_{\lfloor n(1-\delta)/(mk')\rfloor})^{\times k'}]^{\otimes m}\| \leq \epsilon'.$$

Now we claim that the protocol $\mathbf{P}_3 := \mathbf{R} \circ \mathbf{P}' \circ \mathbf{P}_1^{\otimes k} \otimes \mathbf{R}'$ satisfies Eq. (1.17). Indeed $\mathbf{P}_3[\alpha_n] = \mathbf{R} \circ \mathbf{P}'[\iota^{\times m+k'l-1}]$ maps α to γ with inefficiency $\delta' \leq 4\delta + 1/m \leq 5\delta$, depth $\leq mkk'(1+\delta) \leq k(k')^2 l(1+1/\delta)$ (where k, k' depend only on δ) and error $\epsilon'' \leq \epsilon' + m\epsilon$. Since $\delta' \to 0$ as depth increases and $\epsilon'' \to 0$ as $n \to \infty$, this satisfies our definition of an asymptotic protocol.

2. We begin with the standard quantifiers from our definition of a resource inequality: $\forall\delta > 0, \exists k, k', \forall\epsilon > 0, \exists N, \forall n \geq N$

$$\|\mathbf{P}_1[(\alpha_{\lfloor n/(kk')\rfloor})^{\times k}] - \mathbf{S}[\beta_{\lfloor n(1-\delta)/k'\rfloor}]\| \leq \epsilon, \tag{1.18}$$

$$\|\mathbf{P}_2[(\gamma_{\lfloor n/(kk')\rfloor})^{\times k'}] - \mathbf{S}[\varepsilon_{\lfloor n(1-\delta)/k\rfloor}]\| \leq \epsilon, \tag{1.19}$$

$$\mathbf{R}_1 : (\beta_{\lfloor n(1-\delta)/k'\rfloor})^{\otimes k'} \overset{*}{\geq} \beta_{\lfloor n(1-2\delta)\rfloor}, \tag{1.20}$$

$$\mathbf{R}_2 : (\varepsilon_{\lfloor n(1-\delta)/k\rfloor})^{\otimes k} \overset{*}{\geq} \varepsilon_{\lfloor n(1-2\delta)\rfloor}, \tag{1.21}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are both ϵ-protocols. Hence the depth-$(k+k')$ $(k+k')\epsilon$-protocol $\mathbf{P}_3$ given by

$$\mathbf{R}_1 \circ \mathbf{P}_1[(\alpha_{\lfloor n/(kk')\rfloor})^{\times k}]^{\otimes k'} \otimes \mathbf{R}_2 \circ \mathbf{P}_2[(\gamma_{\lfloor n/(kk')\rfloor})^{\times k'}]^{\otimes k},$$

satisfies

$$\|\mathbf{P}_3[((\alpha+\gamma)_{\lfloor n/(kk')\rfloor})^{\times kk'}] - \mathbf{S}[(\beta+\varepsilon)_{\lfloor n(1-2\delta)\rfloor}]\| \leq (k+k')\epsilon. \tag{1.22}$$

3. The proof is trivial.

□

It is worth noting that our definitions of resources and resource inequalities were carefully chosen with the above theorem in mind; as a result the proof exposes most of the important features of our definitions. (It is a useful exercise to try changing aspects of our definitions to see where the above proof breaks down.) By contrast, the remainder of this section will establish a number of details about the resource formalism that mostly depend only on Eqns. (1.10) and (1.11) and not so much on the details of how we construct protocols and resource inequalities.

Definition 1.23 (Equivalent resources). *Define an equivalence between resources $\alpha \equiv \beta$ iff $\alpha \geq \beta$ and $\beta \geq \alpha$.*

Example 1.24. *It is easy to see that $R[q\,q] \equiv (\Phi_{D'_n})_n$ with $D'_n = \lfloor 2^{nR} \rfloor$.*

Lemma 1.25. *For resources in $\mathcal{R}$:*

1. $(zw)\alpha \equiv z(w\alpha)$
2. $z(\alpha + \beta) = z\alpha + z\beta$
3. $(z + w)\alpha \equiv z\alpha + w\alpha$

Proof.

1. The $\geq$ is trivial, since $\lfloor zwn \rfloor \geq \lfloor z \lfloor wn \rfloor \rfloor$. The $\leq$ follows from $\lfloor zwn \rfloor \leq zwn \leq \lfloor z \lfloor wn \rfloor \rfloor + z + 1$.
2. Immediate from the definitions.
3. Let $k = \lfloor zm \rfloor$ and $k' = \lfloor wm \rfloor$, where m is a parameter we will choose later.

   For any δ and sufficiently large n (depending on δ and m),

$$\alpha_{\lfloor zn(1+2\delta) \rfloor} \overset{*}{\geq} (\alpha_{\lfloor \lfloor zn(1+\delta) \rfloor / k \rfloor})^{\otimes k},$$
$$\alpha_{\lfloor wn(1+2\delta) \rfloor} \overset{*}{\geq} (\alpha_{\lfloor \lfloor wn(1+\delta) \rfloor / k' \rfloor})^{\otimes k'},$$
$$\alpha_{\lfloor (z+w)n(1+2\delta) \rfloor} \overset{*}{\geq} (\alpha_{\lfloor \lfloor (z+w)n(1+\delta) \rfloor / (k+k') \rfloor})^{\otimes (k+k')},$$
$$(\alpha_{\lfloor \lfloor zn(1-\delta) \rfloor / k \rfloor})^{\otimes k} \overset{*}{\geq} \alpha_{\lfloor zn(1-2\delta) \rfloor},$$
$$(\alpha_{\lfloor \lfloor wn(1-\delta) \rfloor / k' \rfloor})^{\otimes k'} \overset{*}{\geq} \alpha_{\lfloor wn(1-2\delta) \rfloor},$$
$$(\alpha_{\lfloor \lfloor (z+w)n(1-\delta) \rfloor / (k+k') \rfloor})^{\otimes (k+k')} \overset{*}{\geq} \alpha_{\lfloor (z+w)n(1-2\delta) \rfloor}.$$

   Observe:

$$\begin{array}{rcl} |zn - kn/m| & \leq & n/m \\ |\lfloor zn \rfloor - \lfloor kn/m \rfloor| & \leq & n/m + 1 \\ |\lfloor zn \rfloor - k \lfloor n/m \rfloor| & \leq & n/m + k + 2 \\ |\lfloor \lfloor zn \rfloor / k \rfloor - \lfloor n/m \rfloor| & \leq & n/(km) + 2 + 2/k. \end{array}$$

   Thus, for sufficiently large n and an appropriate choice of m,

$$\lfloor \lfloor zn(1+\delta) \rfloor / k \rfloor \geq \lfloor n/m \rfloor \geq \lfloor \lfloor zn(1-\delta) \rfloor / k \rfloor.$$

   Analogously,

$$\lfloor \lfloor wn(1+\delta) \rfloor / k \rfloor \geq \lfloor n/m \rfloor \geq \lfloor \lfloor wn(1-\delta) \rfloor / k \rfloor$$

   and

$$\lfloor \lfloor (w+z)n(1+\delta) \rfloor / (k+k') \rfloor \geq \lfloor n/m \rfloor \geq \lfloor \lfloor (w+z)n(1-\delta) \rfloor / (k+k') \rfloor.$$

Let us start with the $\leq$ direction.

$$\begin{aligned}
\alpha_{\lfloor zn(1+2\delta)\rfloor} \otimes \alpha_{\lfloor wn(1+2\delta)\rfloor} &\overset{*}{\geq} \left(\alpha_{\lfloor \lfloor zn(1+\delta)\rfloor /k\rfloor}\right)^{\otimes k} \otimes \left(\alpha_{\lfloor \lfloor wn(1+\delta)\rfloor /k'\rfloor}\right)^{\otimes k'} \\
&\overset{*}{\geq} \left(\alpha_{\lfloor n/m\rfloor}\right)^{\otimes k} \otimes \left(\alpha_{\lfloor n/m\rfloor}\right)^{\otimes k'} \\
&= \left(\alpha_{\lfloor n/m\rfloor}\right)^{\otimes (k+k')} \\
&\overset{*}{\geq} \left(\alpha_{\lfloor \lfloor (w+z)n(1-\delta)\rfloor /(k+k')\rfloor}\right)^{\otimes (k+k')} \\
&\overset{*}{\geq} \alpha_{\lfloor (z+w)n(1-2\delta)\rfloor}.
\end{aligned}$$

The $\geq$ direction is proven similarly.

□

Definition 1.26 (Equivalence classes of resources). *Denote by $\widetilde{\alpha}$ the equivalence class of α, i.e. the set of all α' such that $\alpha' \equiv \alpha$. Define $\widetilde{\mathcal{R}}$ to be the set of equivalence classes of resources in $\mathcal{R}$. Define the relation $\geq$ on $\widetilde{\mathcal{R}}$ by $\widetilde{\alpha} \geq \widetilde{\beta}$ iff $\alpha' \geq \beta'$ for all $\alpha' \in \widetilde{\alpha}$ and $\beta' \in \widetilde{\beta}$. Define the operation $+$ on $\widetilde{\mathcal{R}}$ such that $\widetilde{\alpha} + \widetilde{\beta}$ is the union of $\widetilde{\alpha' + \beta'}$ over all $\alpha' \in \widetilde{\alpha}$ and $\beta' \in \widetilde{\beta}$. Define the operation $\cdot$ on $\widetilde{\mathcal{R}}$ such that $z\widetilde{\alpha}$ is the union of $\widetilde{z\alpha'}$ over all $\alpha' \in \widetilde{\alpha}$.*

Lemma 1.27. *For resources in $\mathcal{R}$:*

1. *$\widetilde{\alpha} \geq \widetilde{\beta}$ iff $\alpha \geq \beta$*

2. *$\widetilde{\alpha} + \widetilde{\beta} = \widetilde{\alpha + \beta}$*

3. *$z\widetilde{\alpha} = \widetilde{z\alpha}$*

Proof. Regarding the first item: it suffices to show the "if" direction. Indeed, for any $\alpha' \in \widetilde{\alpha}$ and $\beta' \in \widetilde{\beta}$

$$\alpha' \geq \alpha \geq \beta \geq \beta',$$

by Theorem 1.22. Regarding the second item: it suffices to show that if $\alpha' \equiv \alpha$, $\beta' \equiv \beta$ then $\alpha' + \beta' \equiv \alpha + \beta$. This follows from Theorem 1.22. Similarly, for the third item it suffices to show that if $\alpha' \equiv \alpha$ then $z\alpha' \equiv z\alpha$, which is true by Theorem 1.22. □

We now state a number of additional properties of $\widetilde{\mathcal{R}}$, each of which can be easily verified.

Theorem 1.28. *The relation $\geq$ forms a partial order on the set $\widetilde{\mathcal{R}}$:*

1. *$\widetilde{\alpha} \geq \widetilde{\alpha}$ (reflexivity)*

2. *if $\widetilde{\alpha} \geq \widetilde{\beta}$ and $\widetilde{\beta} \geq \widetilde{\gamma}$ then $\widetilde{\alpha} \geq \widetilde{\gamma}$ (transitivity)*

3. *if $\widetilde{\alpha} \geq \widetilde{\beta}$ and $\widetilde{\beta} \geq \widetilde{\alpha}$ then $\widetilde{\alpha} = \widetilde{\beta}$ (antisymmetry)*

□

Theorem 1.29. *The following properties hold for the set $\widetilde{\mathcal{R}}$ with respect to $+$ and multiplication by positive real numbers.*

1. *$(zw)\widetilde{\alpha} = z(w\widetilde{\alpha})$*

2. *$(z + w)\widetilde{\alpha} = z\widetilde{\alpha} + w\widetilde{\alpha}$*

3. *$z(\widetilde{\alpha} + \widetilde{\beta}) = z\widetilde{\alpha} + z\widetilde{\beta}$*

4. *$1\,\widetilde{\alpha} = \widetilde{\alpha}$*

□

Theorem 1.30. *For equivalence classes in $\widetilde{\mathcal{R}}$:*

1. *if $\widetilde{\alpha}_1 \geq \widetilde{\alpha}_2$ and $\widetilde{\beta}_1 \geq \widetilde{\beta}_2$ then $\widetilde{\alpha}_1 + \widetilde{\beta}_1 \geq \widetilde{\alpha}_2 + \widetilde{\beta}_2$*
2. *if $\widetilde{\alpha} \geq \widetilde{\beta}$ then $z\widetilde{\alpha} \geq z\widetilde{\beta}$*

□

Warning: Lemma 1.27 has essentially allowed us to replace resources with their equivalence classes and $\equiv$ with $=$. Henceforth we shall equate the two, and drop the $\sim$ superscript. The one exception to this rule is when writing relative resources as $(\beta : \gamma)$ where β is a proper dynamic resource and γ is a proper static resource; in this case replacing $(\beta : \gamma)$ with its equivalence class is well-defined, but replacing β and γ with their equivalence classes wouldn't make sense.

1.3 General resource inequalities

In this section, we describe several resource inequalities that will serve as useful basic tools for manipulating and combining other resource inequalities.

Lemma 1.31. *Let β and β' be proper dynamic resources, and γ and γ' static test resources. The following resource inequalities hold:*

1. *$\beta \geq (\beta : \gamma)$*
2. *$(\beta : \gamma) + \gamma \geq \beta(\gamma)$*
3. *if $\gamma \supseteq \gamma'$ then $(\beta : \gamma') \geq (\beta : \gamma)$*
4. *$\beta : \gamma + \beta' : (\beta\gamma) \geq (\beta' \circ \beta) : \gamma$.*

Proof. Immediate from definitions. □

Lemma 1.32 (Closure). *For resources in $\mathcal{R}$, if $w_0 > 0$ and $w\alpha \geq \beta$ for every $w > w_0$ then $w_0\alpha \geq \beta$.*

Proof. The statement is equivalent to

$$w_0\alpha \geq (1-\delta)\beta, \ \ \forall \delta > 0,$$

which by definition implies the statement for $\delta = 0$. □

The case of $w_0 = 0$ is special and corresponds to the use of a sublinear amount of a resource.

Definition 1.33 (Sublinear o terms). *We write*

$$\alpha + o\gamma \geq \beta$$

if for every $w > 0$

$$\alpha + w\gamma \geq \beta.$$

At the other extreme we might consider the case when we are allowed an unlimited amount of some resource, typically when proving converse theorems.

Definition 1.34 (∞ terms). *We write*

$$\alpha + \infty\gamma \geq \beta$$

if for any $\delta > 0$, there exists k such that for any $\epsilon > 0$ there exists n_1, n_2 and a ϵ-valid protocol **P** *satisfying*

$$\left\| \mathbf{P}[(\alpha_{\lfloor n_1/k \rfloor} + \gamma_{\lfloor n_2/k \rfloor})^{\times k}] - \beta_{\lfloor (1-\delta)n \rfloor} \right\| \leq \epsilon.$$

This means that we can use an amount of γ that increases arbitrarily quickly with n. Note that $\infty\gamma$ cannot be defined as a resource, since it violates Eq. (1.11).

Definition 1.35 (Negative terms). *For any $z < 0$, define the statement*

$$\alpha + z\gamma \geq \beta$$

to mean that

$$\alpha \geq \beta + (-z)\gamma.$$

Similarly, $\alpha \geq \beta + z\gamma$ means that $\alpha + (-z)\gamma \geq \beta$.

Again $-\gamma$ is obviously not a resource, but the above definition lets us treat it as such.

We now return to sublinear terms. In general we cannot neglect sublinear resources; e.g. in entanglement dilution, they are both necessary[HL04, HW03] and sufficient[LP99]. However, this situation only occurs when they cannot be generated from the other resources being used in the protocol.

Lemma 1.36 (Removal of o terms). *For $\alpha, \beta, \gamma \in \mathcal{R}$, if*

$$\begin{array}{rcl} \alpha + o\gamma & \geq & \beta \\ z\alpha & \geq & \gamma \end{array}$$

for some real $z > 0$, then

$$\alpha \geq \beta.$$

Proof. For any $w > 0$

$$(1 + zw)\alpha \geq \alpha + w\gamma \geq \beta,$$

and the lemma follows by the Closure Lemma (1.32). □

One place that sublinear resources often appear is as catalysts, meaning they are used to enable a protocol without themselves being consumed. Repeating the protocol many times reduces the cost of the catalyst to sublinear:

Lemma 1.37 (Cancellation). *For $\alpha, \beta, \gamma \in \mathcal{R}$, if*

$$\alpha + \gamma \geq \beta + \gamma,$$

then $\alpha + o\gamma \geq \beta$.

Proof. Combine N copies of the inequality (using part 1 of Theorem 1.22) to obtain

$$\gamma + N\alpha \geq \gamma + N\beta.$$

Divide by N:

$$N^{-1}\gamma + \alpha \geq N^{-1}\gamma + \beta \geq \beta.$$

As N^{-1} is arbitrarily small, the result follows. □

Often we will find it useful to use shared randomness as a catalyst. The condition for this to be possible is that the randomness be incoherently decoupled:

Lemma 1.38 (Recycling common randomness). *If α and β are resources for which*

$$\alpha + R\,[c\,c] \geq \beta,$$

and the $[c\,c]$ is incoherently decoupled in the above resource inequality (RI), then

$$\alpha + o\,[c\,c] \geq \beta.$$

Proof. Since $[c\,c]$ is asymptotically independent of the β resource, by definitions 1.12 and 1.21 it follows that

$$\alpha + R\,[c\,c] \geq \beta + R\,[c\,c].$$

An application of the cancellation lemma (1.37) yields the desired result. □

Corollary 1.39. *If $\alpha \geq [c\,c]$ and β is pure then*

$$\alpha + R\,[c\,c] \geq \beta$$

can always be derandomized to

$$\alpha \geq \beta.$$

Proof. It suffices to notice that for a pure output resource β, equation (1.9) is automatically satisfied. □

The following theorem tells us that in proving channel coding theorems one only needs to consider the case where the input state is maximally mixed. A similar result was shown in [BKN00] (see also [KW04, YDH05]), though with quite different techniques and formalism.

Theorem 1.40 (Absolutization). *The following resource inequalities hold:*

1. $[q \to q : \tau] \geq [q \to q]$
2. $[\![c \to c : \tau]\!] \geq [q \to qq]$
3. $[c \to c : \tau] \geq [c \to c]$

Proof. The lemma is a direct consequence of Lemma 1.11. We shall prove case 1., as the proofs of 2. and 3. are identical. By Lemma 1.11, we know that

$$[q \to q] : [\tau] + 2[c\,c] \geq [q \to q] + 2[c\,c].$$

By the cancellation lemma,

$$[q \to q] : [\tau] + o[c\,c] \geq [q \to q].$$

Since

$$[q \to q] : [\tau] \geq [c\,c],$$

by Lemma 1.36 the o term can be dropped, and we are done. □

Finally, we note how convex combinations of static resources can be thought of as states conditioned on classical variables.

Theorem 1.41. *Consider some static i.i.d. resource $\alpha = \langle \sigma \rangle$, where*

$$\sigma^{AX_ABX_B} = \sum_x p_x |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_B} \otimes \rho_x^{AB}.$$

Namely, Alice and Bob share an ensemble of bipartite states, and they both have the classical information about which state they hold. Denote $\alpha_x = \langle \rho_x \rangle$. Then

$$\alpha \geq \sum_x p_x \alpha_x.$$

Proof. Recall the notion of the typical set[CT91, CK81] $\mathcal{T}$ such that for any $\epsilon, \delta > 0$ and sufficiently large n, $p^{\otimes n}(\mathcal{T}) \geq 1 - \epsilon$ and for any $x^n \in \mathcal{T}$,

$$|n_x - p_x n| \leq \delta n,$$

where n_x is the number of occurrences of the symbol x in x^n. Then

$$\left\| \sigma^{\otimes n} - \sum_{x^n \in \mathcal{T}} |x^n\rangle\langle x^n|^{X_A} \otimes p^{\otimes n}(x^n) |x^n\rangle\langle x^n|^{X_B} \otimes \rho_{x^n} \right\|_1 \leq \epsilon.$$

The state that we want to simulate is $\mathbf{S}[\sum_x p_x \alpha_x] = (\omega_n)_n$ with

$$\omega_n = \bigotimes_x \rho_x^{\otimes \lfloor p_x n \rfloor}.$$

For any $x^n \in \mathcal{T}$ there is, clearly, a unitary $U^A_{x^n} \otimes U^B_{x^n}$ that maps ρ_{x^n} to $\omega_{([1-\delta]n-1)} \otimes \hat{\rho}_{x^n}$ exactly for some state $\hat{\rho}_{x^n}$. Performing

$$\left(\sum_{x^n} |x^n\rangle\langle x^n|^{X_A} \otimes U^A_{x^n} \right) \otimes \left(\sum_{x^n} |x^n\rangle\langle x^n|^{X_B} \otimes U^B_{x^n} \right)$$

and tracing out subsystems thus brings $\sigma^{\otimes n}$ ϵ-close to $\omega_{([1-\delta]n-1)}$. Hence the claim. □

In fact, the above result could be strengthened to the equality

$$\alpha = \sum_x p_x \alpha_x + H(X_A)_\sigma [cc], \tag{1.23}$$

but we will not need this fact, so leave the proof as an exercise for the reader. However, we will show how a similar statement to Theorem 1.41 can be made about relative resources.

Theorem 1.42. *Consider some channel $\mathcal{N}$ with input Hilbert space A and a state σ of the form*

$$\sigma^{RAX_AX_B} = \sum_x p_x |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_B} \otimes \phi_x^{RA}.$$

Namely, Alice has an ensemble of states $|\phi_x\rangle$, and both parties have the classical information identifying the state. Then

$$\sum_x p_x \langle \mathcal{N} : \phi_x^A \rangle \geq \langle \mathcal{N} : \sigma \rangle.$$

□

Proof. We will only give an outline of the simulation procedure; the proof of correctness is essentially the same as for the last theorem. Given $\sigma^{\otimes m}$ with $m = (1-\delta)n - 1$, Alice will locally prepare $\hat{\rho}^{x^m}$ conditioned on x^m from $\sigma^{\otimes m}$ (which is possible since ϕ_x^{RA} can be locally prepared by Alice), perform the inverse of the map $U^A_{x^m}$ from the last theorem and then apply $\mathcal{N}^{\otimes n}$. □

1.4 Known coding theorems expressed as resource inequalities

There have been a number of quantum and classical coding theorems discovered to date, typically along with so-called converse theorems which prove that the coding theorems cannot be improved upon. The theory of resource inequalities has been developed to provide an underlying unifying principle. This direction was initially suggested in [DW03b].

We shall state theorems such as Schumacher compression, the classical reverse Shannon theorem, the instrument compression theorem, the classical-quantum Slepian-Wolf theorem, the HSW theorem, and CR concentration as resource inequalities. Then we will show how some of these can be used as building blocks, yielding transparent and concise proofs of some derivative results.

We shall work within the QQ formalism.

Schumacher compression. The quantum source compression theorem was proven by Schumacher in [JS94, Sch95]. Given a quantum state $\rho^{A'}$, define $\sigma^B := \mathrm{id}^{A'\to B}(\rho^{A'})$. Then the following resource inequality (RI) holds:

$$(H(B)_\sigma + \delta)[q \to q] \geq \langle \mathrm{id}^{A'\to B} : \rho^{A'} \rangle \tag{1.24}$$

if and only if $\delta \geq 0$.

Note that this formulation simultaneously expresses both the coding theorem and the converse theorem.

Entanglement concentration. The problem of entanglement concentration was solved in [BBPS96], and is, in a certain sense, a static counterpart to Schumacher's compression theorem. Entanglement concentration can be thought of as a coding theorem which says that given a pure bipartite quantum state $|\phi\rangle^{AB}$ the following RI holds:

$$\langle \phi^{AB} \rangle \geq H(B)_\phi \, [q\, q]. \tag{1.25}$$

The reverse direction is known as *entanglement dilution* [BBPS96], and thanks to Lo and Popescu [LP99] it is known that

$$H(B)_\phi \, [q\, q] + o\, [c \to c] \geq \langle \phi^{AB} \rangle. \tag{1.26}$$

Were it not for the $o\, [c \to c]$ term, we would have the equality $\langle \phi^{AB} \rangle = H(B)_\phi \, [q\, q]$. However, it turns out that the $o[c \to c]$ term cannot be avoided[HL04, HW03]. This means that the strongest equality we can state has a sublinear amount of classical communication on both sides:

$$H(B)_\phi \, [q\, q] + o\, [c \to c] = \langle \phi^{AB} \rangle + o\, [c \to c]. \tag{1.27}$$

Note how Eq. (1.27) states the converse in a form that is in some ways stronger than Eq. (1.24), since it implies the transformation is not only optimal, but also asymptotically reversible. We can also state a converse when more classical communication is allowed, though no longer as a resource equality:

$$\langle \phi^{AB} \rangle + \infty\, [c \to c] \geq (H(B)_\phi - \delta)\, [q\, q]$$

iff $\delta \geq 0$; and similarly for entanglement dilution.

Shannon compression. Shannon's classical compression theorem was proven in [Sha48]. Given a classical state ρ^{X_A} and defining

$$\sigma^{X_B} = \overline{\mathrm{id}}^{X_A \to X_B}(\rho^{X_A}),$$

Shannon's theorem says that

$$(H(X_B)_\sigma + \delta)[c \to c] \geq \langle \overline{\mathrm{id}}^{X_A \to X_B} : \rho^{X_A} \rangle, \tag{1.28}$$

if and only if $\delta \geq 0$.

Common randomness concentration. This is the classical analogue of entanglement concentration, and a static counterpart to Shannon's compression theorem. It states that, if Alice and Bob have a copy of the same random variable X, embodied in the classical bipartite state

$$\rho^{X_A X_B} = \sum_x p_x |x\rangle\langle x|^{X_A} \otimes |x\rangle\langle x|^{X_B},$$

then

$$\langle \rho^{X_A X_B} \rangle \geq H(X_B)_\rho \, [c\, c]. \tag{1.29}$$

Incidentally, common randomness dilution can do without the o term:

$$H(X_B)_\rho \, [c\, c] \geq \langle \rho^{X_A X_B} \rangle.$$

Thus we obtain a simple resource equality:

$$H(X_B)_\rho \, [c \to c] = \langle \rho^{X_A X_B} \rangle.$$

Classical reverse Shannon theorem (CRST). This theorem was proven in [BSST02, Win02], and it generalizes Shannon's compression theorem to compress probability distributions of classical states instead of pure classical states. Given a classical channel $\overline{\mathcal{N}} : X_{A'} \to Y_B$ and a classical state $\rho^{X_{A'}}$, the CRST states that

$$I(X_A; Y_B)_\sigma [c \to c] + H(X_A|Y_B)_\sigma [c\, c] \geq \langle \overline{\mathcal{N}} : \rho^{X_{A'}} \rangle, \tag{1.30}$$

where

$$\sigma^{X_A Y_B} = \overline{\mathcal{N}} \circ \overline{\Delta}^{X_{A'} \to X_{A'} X_A} (\rho^{X_{A'}}).$$

Moreover, given a modified classical channel $\overline{\mathcal{N}}' : X_{A'} \to Y_A Y_B$ which also provides Alice with a copy of the channel output,

$$\overline{\mathcal{N}}' = \overline{\Delta}^{Y_B \to Y_A Y_B} \circ \overline{\mathcal{N}},$$

the following stronger RI also holds:

$$I(X_A; Y_B)_\sigma [c \to c] + H(X_A|Y_B)_\sigma [c\, c] \geq \langle \overline{\mathcal{N}}' : \rho^{X_{A'}} \rangle, \tag{1.31}$$

In fact, this latter RI can be reversed to obtain the equality

$$I(X_A; Y_B)_\sigma [c \to c] + H(X_A|Y_B)_\sigma [c\, c] = \langle \overline{\mathcal{N}}' : \rho^{X_{A'}} \rangle. \tag{1.32}$$

However, in the case without feedback, the best we can do is a tradeoff curve between cbits and rbits, with Eq. (1.30) representing the case of unlimited randomness consumption. The full tradeoff will be given by an RI of the following form

$$a \, [c \to c] + b \, [c\, c] \geq \langle \overline{\mathcal{N}} : \rho^{X_{A'}} \rangle$$

where (a, b) range over some convex set $CR(\overline{\mathcal{N}})$. It can be shown[Wyn75, BW05] that $(a, b) \in CR(\overline{\mathcal{N}})$ iff there exist channels $\overline{\mathcal{N}}_1 : X_{A'} \to W_{C'}, \overline{\mathcal{N}}_2 : W_{C'} \to Y_B$ such that $\overline{\mathcal{N}} = \overline{\mathcal{N}}_2 \circ \overline{\mathcal{N}}_1$ and $a \geq I(X_A; W_C)_\omega, b \geq I(X_A Y_B; W_C)_\omega$, where

$$\omega^{X_A W_C Y_B} := \overline{\mathcal{N}}_2 \circ \overline{\Delta}^{W_{C'} \to W_{C'} W_C} \circ \overline{\mathcal{N}}_1 \circ \overline{\Delta}^{X_{A'} \to X_{A'} X_A} (\rho^{X_{A'}}).$$

Classical compression with quantum side information. This problem was solved in [DW03a, Win99b], and is a generalization of Shannon's classical compression theorem in which Bob has quantum side information about the source. Suppose Alice and Bob are given an ensemble

$$\rho^{X_A B} = \sum_x p_x |x\rangle\langle x|^{X_A} \otimes \rho_x^B,$$

and Alice wants to communicate X_A to Bob, which would give them the state

$$\sigma^{X_B B} := \overline{\mathrm{id}}^{X_A \to X_B}(\rho^{X_A B}).$$

To formalize this situation, we use the Source as one of the protagonists in the protocol, so that the coding theorem inputs a map from the Source to Alice and Bob $\langle \overline{\mathrm{id}}^{S_X \to X_A} \otimes \mathrm{id}^{S_B \to B} : \rho^{S_X S_B}\rangle$ and outputs a map from the Source entirely to Bob. The coding theorem is then

$$\langle \overline{\mathrm{id}}^{S_X \to X_A} \otimes \mathrm{id}^{S_B \to B} : \rho^{S_X S_B}\rangle + (H(X_B|B)_\sigma + \delta)[c \to c] \geq \langle \overline{\mathrm{id}}^{S_X \to X_B} \otimes \mathrm{id}^{S_B \to B} : \rho^{S_X S_B}\rangle, \quad (1.33)$$

which holds iff $\delta \geq 0$. This formulation ensures that we work with well-defined resources instead of using the natural-seeming, but incorrect $\langle \mathrm{id}^{X_A \to X_B} : \rho^{X_A B}\rangle$ (which violates Eqns. (1.10) and (1.11)).

Of course, with no extra resource cost Alice could keep a copy of X_A.

Instrument compression theorem. This theorem was proven in [Win04], and is a generalization of the CRST. Given a remote instrument $\mathbf{T} : A' \to AX_B$, and a quantum state $\rho^{A'}$, the following RI holds:

$$I(R; X_B)_\sigma[c \to c] + H(X_B|R)_\sigma[c\,c] \geq \langle \mathbf{T} : \rho^{A'}\rangle, \quad (1.34)$$

where

$$\sigma^{RAX_B} = \mathbf{T}(\psi^{RA'})$$

and $|\psi\rangle\langle\psi|^{RX_A} \supseteq \rho^{X_A}$. Moreover, given a modified remote instrument which also provides Alice with a copy of the instrument output,

$$\mathbf{T}' = \overline{\Delta}^{X_B \to X_A X_B} \circ \mathbf{T},$$

the RI still holds:

$$I(R; X_B)_\sigma[c \to c] + H(X_B|R)_\sigma[c\,c] \geq \langle \mathbf{T}' : \rho^{A'}\rangle. \quad (1.35)$$

Only this latter RI is known to be optimal (up to a trivial substitution of $[c \to c]$ for $[c\,c]$); indeed

$$a[c \to c] + b[c\,c] \geq \langle \mathbf{T}' : \rho^{A'}\rangle. \quad (1.36)$$

iff $a \geq I(R; X_B)_\sigma$ and $a + b \geq H(X_B)_\sigma$.

By contrast, only the communication rate of Eq. (1.34) is known to be optimal; examples are known in which less randomness is necessary.

Remote state preparation (RSP) Instrument compression can be thought of as a generalization of the CRST from $\{c \to c\}$ channels to $\{q \to c\}$ channels. In contrast, remote state preparation (proved in [BHL+05]) generalizes the CRST to $\{c \to q\}$ channels.

Let $\mathcal{E} = \sum_i p_i\, |i\rangle\langle i|^{X_A} \otimes |\psi_i\rangle\langle\psi_i|^{AB}$ be an ensemble of bipartite states. Define the corresponding $\{c \to q\}$ channel $\mathcal{N}_\mathcal{E}$ by

$$\mathcal{N}_\mathcal{E}(|i\rangle\langle j|^{X_A}) = \delta_{ij}\, |i\rangle\langle i|^{X_A} \otimes |\psi_i\rangle\langle\psi_i|^{AB}. \quad (1.37)$$

This means that $\mathcal{N}_\mathcal{E}$ measures the input in the standard basis and maps outcome i to the joint state ψ_i^{AB}. Thus, $\mathcal{E} = \mathcal{N}_\mathcal{E}(\mathcal{E}^{X_A})$, where $\mathcal{E}^{X_A}$ is the classical input state $\sum_i p_i\, |i\rangle\langle i|^{X_A}$.

The coding theorem of RSP states that

$$I(X_A; B)_\mathcal{E}[c \to c] + H(B)[q\,q] \geq \langle \mathcal{N}_\mathcal{E} : \mathcal{E}^{X_A}\rangle, \quad (1.38)$$

meaning that Alice can use the resources on the LHS to prepare a sequence of states $|\psi_{i_1}\rangle \cdots |\psi_{i_n}\rangle$ of her choosing, with high fidelity on average if she chooses i^n according to $p^{\otimes n}$. Note that since Alice holds the purification of Bob's state, this is stronger than the ability to simulate a $\{c \to q\}$ channel that gives Bob mixed states. The cbit cost is optimal in either case, since HSW coding (Eq. (1.42), below) yields $\langle \mathcal{N}_{\mathcal{E}} : \mathcal{E}^{X_A} \rangle \geq I(X_A; B)_{\mathcal{E}}[c \to c]$ even if Alice's half of ψ_i^{AB} is discarded. However, the entanglement cost of Eq. (1.38) is only known to be optimal for the setting when Alice holds the purification of Bob's output. Determining the minimal resources necessary to perform *visible mixed-state data compression* has been a long-standing open problem in quantum information theory.[BCF$^+$01, KI01, Win02]

Ref. [BHL$^+$05] also proved a stronger "single-shot" version of RSP, the simplest form of which is that $n(1 + o(1))$ cbits and n ebits can be used to prepare an arbitrary n qubit state. It it interesting to note that this does not form an asymptotic resource (as given in Definition 1.15) because it fails to satisfy Eq. (1.11).[*]

Teleportation and super-dense coding. Teleportation [BBC$^+$93] and super-dense coding [BW92] are finite protocols, and we have discussed them already in the introduction. In a somewhat weaker form they may be written as resource inequalities. Teleportation (TP):

$$2\,[c \to c] + [q\,q] \geq [q \to q]. \tag{1.39}$$

Super-dense coding (SD):

$$[q \to q] + [q\,q] \geq 2\,[c \to c]. \tag{1.40}$$

Finally, entanglement distribution:

$$[q \to q] \geq [q\,q]. \tag{1.41}$$

All of these protocols are optimal (we neglect the precise statements), but composing them with each other (e.g. trying to reverse teleportation by using super-dense coding) is wasteful. We will give a resolution to this problem in Chapter 3 by using coherent classical communication.

Holevo-Schumacher-Westmoreland (HSW) theorem. The direct part of this theorem was proven in [Hol98, SW97] and the converse in [Hol73]. Together they say that given a quantum channel $\mathcal{N} : A' \to B$, for any ensemble

$$\rho^{X_A A'} = \sum_x p_x |x\rangle\langle x|^{X_A} \otimes \rho_x^{A'}$$

the following RI holds:

$$\langle \mathcal{N} : \rho^{A'} \rangle \geq (I(X_A; B)_\sigma - \delta)[c \to c], \tag{1.42}$$

iff $\delta \geq 0$, where

$$\sigma^{X_A B} = \mathcal{N}^{A' \to B}(\rho^{X_A A'}).$$

Shannon's noisy channel coding theorem This theorem was proven in [Sha48] and today can be understood as a special case of the HSW theorem. One version of the theorem says that given a classical channel $\overline{\mathcal{N}} : X_{A'} \to Y_B$ and any classical state $\rho^{X_{A'}}$ the following RI holds:

$$\langle \overline{\mathcal{N}} : \rho^{X_{A'}} \rangle \geq (I(X_A; Y_B)_\sigma - \delta)[c \to c], \tag{1.43}$$

[*] This has a number of interesting implications. For example, "single-shot" RSP is not amenable to the sort of cbit-ebit tradeoffs that are possible in the ensemble case[DB01, HJW02, BHL$^+$05]. In fact, the $\exp(n)$ cbit cost for simulating single-shot RSP of n qubits is one of the few known examples where infinite, or super-linear, resources are useful. Also, the RSP capacities of channels appear to be different for single-shot and ensemble RSP[Leu04].

iff $\delta \geq 0$ and where

$$\sigma^{X_A Y_B} := \overline{\mathcal{N}} \circ \overline{\Delta}^{X_{A'} \to X_{A'} X_A}(\rho^{X_A}). \tag{1.44}$$

If we optimize over all input states, then we find that

$$\langle \overline{\mathcal{N}} \rangle \geq C[c \to c] \tag{1.45}$$

iff there exists an input $\rho^{X_{A'}}$ such that $C \geq I(X_A; Y_B)_\sigma$, with σ given by Eq. (1.44).

Entanglement-assisted capacity theorem. This theorem was proven in [BSST02, Hol02]. The direct coding part of the theorem says that, given a quantum channel $\mathcal{N} : A' \to B$, for any quantum state $\rho^{A'}$ the following RI holds:

$$\langle \mathcal{N} : \rho^{A'} \rangle + H(R)_\sigma [q\,q] \geq I(R;B)_\sigma \, [c \to c], , \tag{1.46}$$

where

$$\sigma^{RB} = \mathcal{N}(\psi^{RA'})$$

for an arbitrary ψ satisfying $|\psi\rangle\langle\psi|^{RA'} \supseteq \rho^{A'}$.

The only converse proven in [BSST02, Hol02] was for the case of infinite entanglement: they found that $\langle CN \rangle + \infty[q\,q] \geq C[c \to c]$ iff $C \leq I(R;B)_\sigma$ for some appropriate σ. [Sho04b] gave a full solution to the tradeoff problem for entanglement-assisted classical communication which we will present an alternate derivation of in Section 4.2.7.

Quantum capacity (LSD) theorem. This theorem was conjectured in [Sch96, SN96], a heuristic (but not universally accepted) proof given by Lloyd [Llo96] and finally proven by Shor [Sho02] and with an independent method by Devetak [Dev05a]. The direct coding part of the theorem says that, given a quantum channel $\mathcal{N} : A' \to B$, for any quantum state $\rho^{A'}$ the following RI holds:

$$\langle \mathcal{N} : \rho^{A'} \rangle \geq (I(R\,\rangle B)_\sigma - \delta)\, [q \to q], \tag{1.47}$$

iff $\delta \geq 0$ and where

$$\sigma^{RB} = \mathcal{N}(\psi^{RA'})$$

for any $\psi^{RA'}$ satisfying $|\psi\rangle\langle\psi|^{RA'} \supseteq \rho^{A'}$.

Noisy super-dense coding theorem. This theorem was proven in [HHH+01]. The direct coding part of the theorem says that, given a bipartite quantum state ρ^{AB}, the following RI holds:

$$\langle \rho^{AB} \rangle + H(A)_\rho \, [q \to q] \geq I(A;B)_\rho \, [c \to c]. \tag{1.48}$$

A converse was proven in [HHH+01] only for the case when an infinite amount of $\langle \rho^{AB} \rangle$ is supplied, but we will return to this problem and provide a full trade-off curve in Section 4.2.2.

Entanglement distillation. The direct coding theorem for one-way entanglement distillation is embodied in the *hashing inequality*, proved in [DW05a, DW04]: given a bipartite quantum state ρ^{AB},

$$\langle \rho^{AB} \rangle + I(A;E)_\psi \, [c \to c] \geq I(A\,\rangle B)_\psi \, [q\,q], \tag{1.49}$$

where $|\psi\rangle\langle\psi|^{ABE} \supseteq \rho^{AB}$.

Again, the converse was previously only known for the case when an unlimited amount of classical communication was available[Sch96, SN96, DW05a, DW04]. In Section 4.2.5 we will give an expression for the full trade-off curve.

Noisy teleportation. This RI was discovered in [DHW04]. Given a bipartite quantum state ρ^{AB},

$$\langle \rho^{AB} \rangle + I(A;B)_\rho \, [c \to c] \geq I(A\rangle B)_\rho \, [q \to q].$$

Indeed, letting $|\psi\rangle\langle\psi|^{ABE} \supseteq \rho^{AB}$,

$$\begin{aligned} \langle \rho^{AB} \rangle + I(A;B)_\psi \, [c \to c] &= \langle \rho^{AB} \rangle + I(A;E)_\psi \, [c \to c] + 2I(A\rangle B)_\psi [c \to c] \\ &\geq I(A\rangle B)_\psi \, [q\, q] + 2I(A\rangle B)_\psi [c \to c] \\ &\geq I(A\rangle B)_\psi \, [q \to q]. \end{aligned}$$

The first inequality follows from Eq. (1.49) and the second from teleportation.

Classical-quantum communication trade-off for remote state preparation. The main coding theorem of [HJW02] has two interpretations. Viewed as a statement about quantum compression with classical side information, it says that, given an ensemble

$$\rho^{X_{A'} A'} = \sum_x p_x |x\rangle\langle x|^{X_{A'}} \otimes \rho_x^{A'},$$

for any classical channel $\overline{\mathcal{N}} : X_{A'} \to Y_B$, the following RI holds:

$$H(B|Y_B)_\sigma [q \to q] + I(X_A; Y_B)_\sigma [c \to c] \geq \langle \mathrm{id}^{A' \to B} : \rho^{X_{A'} A'} \rangle. \tag{1.50}$$

where

$$\sigma^{X_A Y_B B} = ((\overline{\mathcal{N}}^{X_{A'} \to Y_B} \circ \overline{\Delta}^{X_{A'} \to X_{A'} X_A}) \otimes \mathrm{id}^{A' \to B}) \rho^{X_{A'} A'}.$$

Conversely, if $a[q \to q] + b[c \to c] \geq \langle \mathrm{id}^{A' \to B} : \rho^{X_{A'} A'} \rangle$ then there exists a classical channel $\overline{\mathcal{N}} : X_A \to Y_B$ with corresponding state σ such that $a \geq H(B|Y_B)_\sigma$ and $b \geq I(X_A; Y_B)_\sigma$.

We shall now show how the proof from [HJW02] may be written very succinctly in terms of previous results. Define $\overline{\mathcal{N}}' = \overline{\Delta}^{Y_B \to Y_A Y_B} \circ \overline{\mathcal{N}}$. By the Classical Reverse Shannon Theorem (Eq. (1.31)) and part 3 of Lemma 1.31,

$$I(X_A; Y_B)_\sigma [c \to c] + H(X_A|Y_B)_\sigma [c\, c] \geq \langle \overline{\mathcal{N}}' : \rho^{X_{A'} A'} \rangle.$$

On the other hand, Schumacher compression (Eq. (1.24)) and Theorem 1.42 imply

$$H(B|Y_B)_\sigma [q \to q] \geq \langle \mathrm{id}^{A' \to B} : \overline{\mathcal{N}}'(\rho^{X_A A'}) \rangle.$$

Adding the two equations and invoking part 2 of Lemma 1.31 gives

$$H(B|Y_B)_\sigma [q \to q] + I(X_A; Y_B)_\sigma [c \to c] + H(X_A|Y_B)_\sigma [c\, c] \geq \langle \mathrm{id}^{A' \to B} : \rho^{X_{A'} A'} \rangle.$$

Finally, derandomizing via Corollary 1.39 gives the desired result (Eq. 1.50).

The result of [HJW02] may be also viewed as a statement about remote state preparation. Suppose we are given a classical state $\rho^{X_{A''}}$ and a $\{c \to q\}$ map $\mathcal{N}'_{\mathcal{E}} : X_{A''} \to B$, $\mathcal{N}'_{\mathcal{E}} = \mathrm{id}^{A' \to B} \circ \mathcal{N}_{\mathcal{E}}$, where $\mathcal{N}_{\mathcal{E}}$ has Kraus representation $\{|\phi_x\rangle^{A^* A'} \langle x|^{X_{A''}}\}_x$. Then for any classical channel $\overline{\mathcal{N}} : X_A \to Y_B$, the following RI holds:

$$H(B|Y_B)_\sigma [q \to q] + I(X_A; Y_B)_\sigma [c \to c] \geq \langle \mathcal{N}'_{\mathcal{E}} : \rho^{X_{A''}} \rangle, \tag{1.51}$$

where $\sigma^{X_A Y_B B}$ is defined as above and

$$\rho^{X_{A'} A^* A'} = (\mathcal{N}_{\mathcal{E}} \circ \overline{\Delta}^{X_{A''} \to X_{A''} X_{A'}}) \rho^{X_{A''}}.$$

This follows from adding (Eq. (1.50)) to

$$\begin{aligned}\langle \mathrm{id}^{A'\to B} : \rho^{X_{A'}A'}\rangle &\geq \langle \mathrm{id}^{A'\to B} : \rho^{X_{A'}A^*A'}\rangle + \langle (\mathcal{N}_{\mathcal{E}} \circ \overline{\Delta}^{X_{A''}\to X_{A''}X_{A'}}) : \rho^{X_{A''}}\rangle \\ &\geq \langle \mathcal{N}'_{\mathcal{E}} : \rho^{X_{A''}}\rangle.\end{aligned}$$

The first inequality follows from part 3 of Lemma 1.31 and the locality of the map $\mathcal{N}_{\mathcal{E}}$. The second is an application of part 4 of Lemma 1.31.

Common randomness distillation. This theorem was originally proven in [DW03b]. Given an ensemble

$$\rho^{X_A B} = \sum_x p_x |x\rangle\langle x|^{X_A} \otimes \rho_x^B,$$

the following RI holds:

$$\langle \rho^{X_A B}\rangle + H(X_A|B)_\rho[c \to c] \geq H(X_A)_\rho[c\,c]. \tag{1.52}$$

Armed with our theory of resource inequalities, the proof becomes extremely simple.

$$\begin{aligned}\langle \rho^{X_A B}\rangle + H(X_A|B)_\rho[c \to c] &\geq \langle \rho^{X_A B}\rangle + \langle \overline{\Delta}^{X_A\to X_A X_B} : \rho^{X_A B}\rangle \\ &\geq \langle \rho^{X_A X_B B}\rangle \\ &\geq \langle \rho^{X_A X_B}\rangle \\ &\geq H(X_A)_\rho[c\,c].\end{aligned}$$

The first inequality is by classical compression with quantum side information (Eq. (1.33)), the second by Lemma 1.31, part 2, and the fourth by common randomness concentration (Eq. (1.29)).

1.5 Discussion

This chapter has laid the foundations of a formal approach to quantum Shannon theory in which the basic elements are asymptotic resources and protocols mapping between them. Before presenting applications of this approach in the next three chapters, we pause for a moment to discuss the limitations of our formalism and possible ways it may be extended.

The primary limitation is that our approach is most successful when considering one-way communication and when dealing with only one noisy resource at a time. These, and other limitations, suggest a number of ways in which we might imagine revising the notion of an asymptotic resource we have given in Definition 1.15. For example, if we were to explore unitary and/or bidirectional resources more carefully, then we would need to reexamine our treatments of depth and of relative resources. Recall that in Definition 1.16 we (1) always simulate the depth-1 version of the output resource, (2) are allowed to use a depth-k version of the input resource where k depends only on the target inefficiency and not the target error. These features were chosen rather delicately in order to guarantee the convergence of the error and inefficiency in the Composability Theorem (1.22), which in turn gets most of its depth blow-up from the double-blocking of the Sliding Lemma (1.10). However, it is possible that a different model of resources would allow protocols which deal with depth differently. This won't make a difference for one-way resources due to the Flattening Lemma (1.17), but there is evidence that depth is an important resource in bidirectional communication[KNTSZ01]; on the other hand, it is unknown how quickly depth needs to scale with n.

Relative resources are another challenge for studying bidirectional communication. As we discussed in Section 1.2.2, if ρ^{AB} cannot be locally duplicated then $\langle \mathcal{N} : \rho^{AB}\rangle$ fails to satisfy Eq. (1.10) therefore is not a valid resource. The problem is that being able to simulate n uses of a channel on n copies of a correlated or entangled state is not necessarily stronger than the ability to simulate $n-1$ uses of the channel on $n-1$ copies of the state. The fact that many bidirectional problems in classical information theory[Sha61] remain unsolved is an indication that the quantum versions of these problems will be difficult. On the other hand, it is possible that special cases, such as unitary gates or Hamiltonians, will offer simplifications not possible in the classical case.

Another challenge to our definition of a resource comes from unconventional "pseudo-resources" that resemble resources in many ways but fail to satisfy the quasi-i.i.d. requirement (Eq. (1.11)). For example, the ability to remotely prepare an arbitrary n qubit state (in contrast with the ensemble version in Eq. (1.38)) cannot be simulated by the ability to remotely prepare k states of $n(1+\delta)/k$ qubits each. There are many fascinating open questions surrounding this "single-shot" version of RSP; for example, is the RSP capacity of a channel ever greater than its quantum capacity?[*] Another example comes from the "embezzling states" of [vDH03]. The n-qubit embezzling state can be prepared from n cbits and n ebits (which are also necessary[HW03]) and can be used as a resource for entanglement dilution and for simulating noisy quantum channels on non-i.i.d. inputs[BDH$^+$05]; however, it also cannot be prepared from k copies of the $n(1+\delta)/k$-qubit embezzling state. These pseudo-resources are definitely useful and interesting, but it is unclear how they should fit into our resource formalism.

Other extensions of the theory will probably require less modification. For example, it will not a priori be hard to extend the theory to multi-user scenarios. Resources and capacities can even be defined in non-cooperative situations pervasive in cryptography (see e.g. [LNC03]), which will mostly require a more careful enumeration of different cases. We can also consider privacy to be a resource. Our definitions of decoupled classical communication are a step in this direction; also there are expressions for the private capacity of quantum channels[Dev05a] and states[DW05a], and there are cryptographic versions of our Composability Theorem[BOM04, Unr04].

[*] Thanks to Debbie Leung for suggesting this question.

Chapter 2

Communication using unitary interactions

In this chapter, we approach bipartite unitary interactions through the lens of quantum Shannon theory, by viewing them as a two-way quantum channels. For example, we might try to find the classical communication capacity of a CNOT $= |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_x$ with control qubit in Alice's laboratory and target qubit in Bob's laboratory. More generally, we will fix a bipartite gate $U \in \mathcal{U}_{d\times d} = \mathcal{U}_{d^2}$ and investigate the rate at which U can generate entanglement, send classical or quantum messages and so on.

This work can be applied both to computation (in a model where local operations are easy and interactions are expensive) and to the rest of Shannon theory, which will be our primary focus in the next two chapters. Most other work on bipartite unitary gates has been more concerned with computational issues, but in Section 2.1 we survey the literature with an eye toward information theory applications. The main results of this chapter are the capacities of a bipartite unitary gate to create entanglement (in Section 2.2) and to send classical messages in one direction when assisted by an unlimited amount of entanglement (in Section 2.3). Along the way, we also establish some easily computable bounds on and relations between these capacities (in Sections 2.2 and 2.3) and discuss these capacities for some interesting specific gates in Section 2.4. We conclude with a summary and discussion in Section 2.5.

Bibliographical note: Except where other works are cited, most of the results in this chapter are from [BHLS03] (joint work with Charles Bennett, Debbie Leung and John Smolin). However, this thesis reformulates them in the formalism of Chapter 1, which allows many of the definitions, claims and proofs to be greatly simplified.

2.1 Background

2.1.1 Survey of related work

The nonlocal strength of unitary interactions was first discussed within a model of communication complexity, when Nielsen introduced the Schmidt decomposition of a unitary gate (described below) as a measure of its nonlocality[Nie98]. The idea of studying a gate in terms of nonlocal invariants—parameters which are unchanged by local unitary rotations—was first applied to two-qubit gates by [Mak02], which found that the nonlocal properties of these gates are completely described by three real parameters. Later these invariants would be interpreted by [KBG01, KC01] as components of a useful general decomposition of two-qubit gates: for any $U \in \mathcal{U}_{2\times 2}$, there exist $A_1, A_2, B_1, B_2 \in \mathcal{U}_2$ and $\theta_x, \theta_y, \theta_z \in (-\frac{\pi}{4}, \frac{\pi}{4}]$ such that

$$U = (A_1 \otimes B_1) e^{i(\theta_x \sigma_x \otimes \sigma_x + \theta_y \sigma_y \otimes \sigma_y + \theta_z \sigma_z \otimes \sigma_z)} (A_2 \otimes B_2). \tag{2.1}$$

This fact has a number of useful consequences, but the only one we will use in this work is that the nonlocal part $\exp(i \sum_j \theta_j \sigma_j \otimes \sigma_j)$ is symmetric under exchange of Alice and Bob, implying that $\langle U \rangle = \langle \text{SWAP} U \text{SWAP} \rangle$ for any $U \in \mathcal{U}_{2\times 2}$. This symmetry no longer holds[BCL+02] (and similar decompositions generally do not exist) for $\mathcal{U}_{d\times d}$ with $d > 2$.

Other early work considered the ability of unitary gates to communicate and create entanglement. [CLP01] showed that $\langle \text{CNOT} \rangle + [qq] \geq [c \to c] + [c \leftarrow c]$ and [CGB00] proved that $2 \log d([c \to c] + [c \leftarrow c] + [qq]) \geq \langle U \rangle$ for any $U \in \mathcal{U}_{d\times d}$. The first discussion of asymptotic capacity was in [DVC+01], which found the rate at which Hamiltonians can generate entanglement. Their technique would be adopted mostly unchanged by [LHL03, BHLS03] to find the entanglement capacity of unitary gates. In general it is difficult to exactly calculate the entanglement capability of Hamiltonians and gates, but [CLVV03] finds the rate at which two-qubit Hamiltonians of the form $H = \alpha\sigma_x \otimes \sigma_x + \beta\sigma_y \otimes \sigma_y$ can generate entanglement.

Instead of reducing gates and Hamiltonians to standard resources, such as cbits and ebits, one can consider the rates at which Hamiltonians and gates can simulate one another. The question of when this is possible is related to the issue of computational universality, which we will not review here; rather, we consider optimal simulations in which fast local operations are free. [BCL+02] found the optimal rate at which a two-qubit Hamiltonian can simulate another, if the time evolution is interspersed by fast local unitaries that do not involve ancilla systems. [VC02b] showed that adding local ancilla systems improves this rate, but that classical communication does not. Hamiltonian simulation is further improved when we allow the ancilla systems to contain entanglement that is used catalytically[VC02a].

The question of optimally generating two-qubit unitary interactions using a given nonlocal Hamiltonian was solved (without ancillas) in [VHC02] and the proof was greatly simplified in [HVC02]. A more systematic approach to the problem was developed in [KBG01], which considers systems of many qubits and applies its techniques to nuclear magnetic resonance. Recently, generic gates on n qubits were shown by [Nie05] to require one- and two-qubit Hamiltonians to be applied for $\mathcal{O}(\exp(n))$ time. Hopefully this work will lead to useful upper bounds on the strengths of Hamiltonians, which so far have been difficult to obtain.

Finally, one can also consider the reverse problem of simulating a nonlocal Hamiltonian or gate using standard resources such as cbits and ebits. This problem has so far resisted optimal solutions, except in a few special cases, such as Gottesman's[Got99] simulation of the CNOT using $[c \to c] + [c \leftarrow c] + [qq]$. For general $d \times d$ unitary gates, a simple application of teleportation yields $2 \log d([c \to c] + [c \leftarrow c] + [qq]) \geq \langle U \rangle$ [CGB00] (and see also Proposition 2.8). Unfortunately this technique cannot be used to efficiently simulate evolution under a nonlocal Hamiltonian for time t, since allowing Alice and Bob to intersperse fast local Hamiltonians requires breaking the simulated action of H into t/ϵ serial uses of $e^{-iH\epsilon}$ for $\epsilon \to 0$. This ends up requiring classical communication on the order of t^2 in order to achieve constant error. However, we would like the cost of a simulation to be linear in the time the Hamiltonian is applied, so that we can discuss *simulation rates* that are asymptotically independent of the time the Hamiltonian is applied. If classical communication is given for free, then [CDKL01] shows how to simulate a general Hamiltonian for time t using $\mathcal{O}(t)$ entanglement. This result was improved by Kitaev[Kit04], who showed how to use $\mathcal{O}(t)([q \to q] + [q \leftarrow q])$ to simulate a Hamiltonian for time t. However, though these constructions are efficient, their rates are far from optimal.

2.1.2 Schmidt decompositions of states and operators

Here we review the familiar Schmidt decomposition of bipartite quantum states[Per93, NC00], and explain the analogous, but less well-known, operator Schmidt decomposition for bipartite operators[Nie98].

Proposition 2.1 (Schmidt decomposition). *Any bipartite pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be written as $|\psi\rangle = \sum_{i=1}^m \sqrt{\lambda_i} |\alpha_i\rangle_{\rm A} |\beta_i\rangle_{\rm B}$, where $\lambda_i > 0$, $\sum_i \lambda_i = 1$ (i.e. λ is a probability distribution* *with full support) and $|\alpha_i\rangle \in \mathcal{H}_\mathrm{A}$ and $|\beta_i\rangle \in \mathcal{H}_\mathrm{B}$ are orthogonal sets of vectors (i.e. $\langle\alpha_i|\alpha_{i'}\rangle = \langle\beta_i|\beta_{i'}\rangle = \delta_{ii'}$). Since these vectors are orthogonal, $m \leq \min(\dim\mathcal{H}_\mathrm{A}, \dim\mathcal{H}_\mathrm{B})$.*

Furthermore, the Schmidt rank $\mathrm{Sch}(U) := m$ is unique, as are the Schmidt coefficients λ_i, up to a choice of ordering. Therefore unless otherwise specified we will take the λ_i to be nonincreasing. Also, for any other decomposition $|\psi\rangle = \sum_{i=1}^{l} |\alpha'_i\rangle_A|\beta'_i\rangle_B$ (with $|\alpha'_i\rangle, |\beta'_i\rangle$ not necessarily orthogonal or normalized), we must have $l \geq \mathrm{Sch}(U)$.

Our proof follows the approach of [NC00].

Proof. The key element of the proof is the singular value decomposition (SVD). Choose orthonormal bases $\{|j\rangle\}_{1\leq j\leq d_A}$ and $\{|k\rangle\}_{1\leq k\leq d_B}$ for $\mathcal{H}_\mathrm{A}$ and $\mathcal{H}_\mathrm{B}$ respectively, where $d_A = \dim\mathcal{H}_\mathrm{A}$ and $d_B = \dim\mathcal{H}_\mathrm{B}$. Then $|\psi\rangle$ can be written as $|\psi\rangle = \sum_{j,k} a_{jk}|j\rangle_A|k\rangle_B$, where a is a $d_A \times d_B$ matrix. The SVD states that there exists a set of positive numbers $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}$ and isometries $u : \mathbb{C}^m \rightarrow \mathbb{C}^{d_B}$ and $v : \mathbb{C}^{d_A} \rightarrow \mathbb{C}^m$ such that $a = u \cdot \sqrt{\mathrm{diag}(\vec{\lambda})} \cdot v$. Let $|\alpha_i\rangle_\mathrm{A} := \sum_j u_{ji}|j\rangle_\mathrm{A}$ and $|\beta_i\rangle_\mathrm{B} := \sum_k v_{ik}|k\rangle_\mathrm{B}$. Since u and v are isometries, it follows that $\{|\alpha_i\rangle\}$ and $\{|\beta_i\rangle\}$ are orthonormal sets.

To prove the second set of claims, note that the Schmidt coefficients are just the singular values of the matrix $a_{jk} = \langle\psi| \cdot |j\rangle|k\rangle$; since singular values are unique, so are Schmidt coefficients. Finally if $|\psi\rangle = \sum_{i=1}^{l} |\alpha'_i\rangle_A|\beta'_i\rangle_B$, then $a_{jk} = \sum_{i=1}^{l}\langle\alpha'_i|j\rangle\langle\beta'_i|k\rangle$ and $\mathrm{Sch}(\psi) = \mathrm{rank}\, a \leq l$. □

Schmidt decomposition and entanglement manipulation: The Schmidt coefficients are central to the study of bipartite pure state entanglement. For example, two states can be transformed into one another via local unitary transformations if and only if they have the same Schmidt coefficients. Thus, we usually choose entanglement measures on pure states to be functions only of their Schmidt coefficients.

Moreover, the intuitive requirement that entanglement be nonincreasing under local operations and classical communication (LOCC) is equivalent to the mathematical requirement that entanglement measures be Schur-concave functions of a state's Schmidt coefficients. (A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is Schur-concave iff $v \prec w \Rightarrow f(v) \geq f(w)$[Bha97]). The proof is as follows: Suppose a bipartite pure state $|\psi\rangle$ can be transformed by LOCC into $|\varphi_i\rangle$ with probability p_i (i.e. the state $\sum_i p_i|ii\rangle\langle ii|_{A_1B_1} \otimes |\varphi_i\rangle\langle\varphi_i|_{A_2B_2}$). Then [Nie99a] showed that this transformation is possible if and only if there exist $\vec{\lambda}$ and $\vec{\mu}_i$ such that $\vec{\lambda} = \sum_i p_i\vec{\mu}_i$ where $\vec{\lambda}$ is the set of Schmidt coefficients of $|\psi\rangle$ (ordered arbitrarily) and the $\vec{\mu}_i$ are the Schmidt coefficients for $|\varphi_i\rangle$ (again in an arbitrary ordering). As a consequence, if $E(|\psi\rangle)$ is an entanglement measure that is a Schur-concave function of the Schmidt coefficients of $|\psi\rangle$, then the expectation of E is nonincreasing under LOCC; i.e. $E(|\psi\rangle) \geq \sum_i p_i E(|\varphi_i\rangle)$[Nie99a, Nie99b]. This general principle unifies many results about entanglement not increasing under LOCC. If we take E to be the standard entropy of entanglement $E(\psi) = H(\mathrm{Tr}_B\,\psi)$, then we find that its expectation doesn't increase under LOCC; similarly for the min-entropy $E_\infty(\psi) = -\log\|\mathrm{Tr}_B\,\psi\|_\infty$. Since $(E_0(\psi))^\alpha = \mathrm{Sch}(\psi)^\alpha$ is Schur-concave for all $\alpha \geq 0$, we also find that Schmidt number has zero probability of increasing under any LOCC transformation (of course, this result follows more directly from the relation $\vec{\lambda} = \sum_i p_i\vec{\mu}_i$).

Operator-Schmidt decomposition: A similar Schmidt decomposition exists for bipartite linear operators $M \in \mathcal{L}(\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B})^*$. Define the Hilbert-Schmidt inner product on $\mathcal{L}(\mathcal{H})$ by $(X, Y) := \mathrm{Tr}\, X^\dagger Y/\dim\mathcal{H}$ for any $X, Y \in \mathcal{L}(\mathcal{H})$. For example, a complete orthonormal basis for the space of one-qubit operators is the set of Pauli matrices, $\{I, X, Y, Z\}$.

Let $d_A = \dim\mathcal{H}_\mathrm{A}$ and $d_B = \dim\mathcal{H}_\mathrm{B}$. Then any $M \in \mathcal{L}(\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B})$ can be Schmidt decomposed into

$$M = \sum_{i=1}^{\mathrm{Sch}(M)} \sqrt{\lambda_i} A_i \otimes B_i \tag{2.2}$$

* We use Nielsen's definition of operator Schmidt number from [Nie98]. In [TH00], Terhal and Horodecki defined an alternative notion of Schmidt number for bipartite density matrices which we will not use.

where $\mathrm{Sch}(M) \leq \min(d_A^2, d_B^2)$, $\mathrm{Tr}\, A_i^\dagger A_j = d_A \delta_{ij}$ and $\mathrm{Tr}\, B_i^\dagger B_j = \delta_{ij}$. Normalization means that $\mathrm{Tr}\, M^\dagger M = d_A d_B \sum_i \lambda_i$; typically M is unitary, so $\sum_i \lambda_i = \mathrm{Tr}\, M^\dagger M / d_A d_B = 1$.

A simple example is the CNOT gate which has operator-Schmidt decomposition

$$\textsc{cnot} = \frac{1}{\sqrt{2}}\, |0\rangle\langle 0| \otimes I + \frac{1}{\sqrt{2}}\, |1\rangle\langle 1| \otimes X \tag{2.3}$$

and hence has Schmidt coefficients $\{1/\sqrt{2}, 1/\sqrt{2}\}$, and $\mathrm{Sch}(\textsc{cnot}) = 2$. The SWAP gate for qubits has operator-Schmidt decomposition

$$\textsc{swap} = \frac{1}{4}\,(I \otimes I + X \otimes X + Y \otimes Y + Z \otimes Z) \tag{2.4}$$

and hence $\mathrm{Sch}(\textsc{swap}) = 4$.

Most facts about the Schmidt decomposition for bipartite states carry over to bipartite operators: in particular, if $M = A_1 \otimes B_1 + \ldots + A_m \otimes B_m$, then $m \geq \mathrm{Sch}(M)$. This implies a useful lemma (originally due to [Nie98], but further discussed in [NDD⁺03]):

Lemma 2.2 (Submultiplicity of Schmidt number). *Let U and V be bipartite operators and $|\psi\rangle$ a bipartite state. Then*

1. $\mathrm{Sch}(UV) \leq \mathrm{Sch}(U)\,\mathrm{Sch}(V)$
2. $\mathrm{Sch}(U|\psi\rangle) \leq \mathrm{Sch}(U)\,\mathrm{Sch}(|\psi\rangle)$

Proof. If $U = \sum_i \sqrt{u_i} A_i \otimes B_i$ and $V = \sum_j \sqrt{v_j} C_j \otimes D_j$ are Schmidt decompositions, then $UV = \sum_{i,j} \sqrt{u_i v_j} A_i C_j \otimes B_i D_j$ is a decomposition of UV into $\mathrm{Sch}(U)\,\mathrm{Sch}(V)$ terms. Therefore $\mathrm{Sch}(UV) \leq \mathrm{Sch}(U)\,\mathrm{Sch}(V)$.

Claim (b) is similar. If $|\psi\rangle = \sum_j \sqrt{\lambda_j} |a_j\rangle \otimes |b_j\rangle$, then $U|\psi\rangle = \sum_{i,j} \sqrt{u_i \lambda_j} A_i |a_j\rangle \otimes B_i |b_j\rangle$ is a decomposition with $\mathrm{Sch}(U)\,\mathrm{Sch}(|\psi\rangle)$ terms. Thus $\mathrm{Sch}(U|\psi\rangle) \leq \mathrm{Sch}(U)\,\mathrm{Sch}(|\psi\rangle)$. □

Part (b) of the above Lemma provides an upper bound for how quickly the Schmidt number of a state can grow when acted on by a bipartite unitary gate. It turns out that this bound is saturated when the gate acts on registers that are maximally entangled with local ancilla systems. This is proven by the next Lemma, a simple application of the Jamiolkowski state/operator isomorphism[Jam72] that was first pointed out by Barbara Terhal in an unpublished comment.

Lemma 2.3. *Given Hilbert spaces* $\mathrm{A}, \mathrm{A}', \mathrm{B}, \mathrm{B}'$ *with* $d_A := \dim \mathrm{A} = \dim \mathrm{A}'$ *and* $d_B := \dim \mathrm{B} = \dim \mathrm{B}'$*, let* $M \in \mathcal{L}(\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B})$ *have Schmidt decomposition* $M = \sum_i \sqrt{\lambda_i} A_i \otimes B_i$*. Then the state* $|\Phi(M)\rangle := (M_{AB} \otimes I_{A'B'})|\Phi_{d_A}\rangle_{AA'}|\Phi_{d_B}\rangle_{BB'}$ *also has Schmidt coefficients* $\{\lambda_i\}$*.*

Proof. If we define $|a_i\rangle = (A_i \otimes I)|\Phi_{d_A}\rangle$ and $|b_i\rangle = (B_i \otimes I)|\Phi_{d_B}\rangle$, then $|\Phi(M)\rangle$ can be written as

$$|\Phi(M)\rangle = \sum_i \sqrt{\lambda_i} |a_i\rangle |b_i\rangle. \tag{2.5}$$

Note that $\langle a_i | a_j \rangle = \mathrm{Tr}(A_i^\dagger A_j \otimes I)\Phi_{d_A} = \mathrm{Tr}\, A_i^\dagger A_j / d_A = \delta_{ij}$ and similarly $\langle b_i | b_j \rangle = \delta_{ij}$. Thus Eq. (2.5) is a Schmidt decomposition of $|\Phi(M)\rangle$, and since the Schmidt coefficients are unique, $|\Phi(M)\rangle$ has Schmidt coefficients $\{\lambda_i\}$. □

2.2 Entanglement capacity of unitary gates

In this section, we investigate the entanglement generating capacity of a unitary interaction. Fix a gate $U \in \mathcal{U}_{d \times d}$ (the generalization to $d_A \times d_B$ is straightforward) and let $\langle U \rangle$ denote the corresponding asymptotic resource.* Then we define the *entanglement capacity* $E(U)$ to be the largest E such that

$$\langle U\rangle \geq E[qq] \tag{2.6}$$

For a bipartite pure state $|\psi\rangle^{AB}$, we will also use $E(|\psi\rangle)$ to indicate the entropy of entanglement of $|\psi\rangle$; i.e. $H(A)_\psi = H(B)_\psi$ in the language of the last chapter.

We will start by stating some easily computable bounds on $E(U)$, then prove a general expression for the capacity and conclude by discussing some consequences.

Simple bounds on entanglement capacity: We can establish some useful bounds on $E(U)$ merely by knowing the Schmidt coefficients of U. The following proposition expresses these bounds.

Proposition 2.4. *If U has Schmidt decomposition $U = \sum_i \sqrt{d_A d_B \lambda_i} A_i \otimes B_i$, then*

$$H(\lambda) = \sum_i -\lambda_i \log \lambda_i \leq E(U) \leq H_0(\lambda) = \log \operatorname{Sch}(U). \tag{2.7}$$

Proof. The lower bound follows from Lemma 2.3 and entanglement concentration (recall from Eq. (1.25) that if ψ is a bipartite pure state then $\langle\psi\rangle \geq E(\psi)[qq]$). Thus, $\langle U\rangle \geq \langle\Phi(U)\rangle \geq H(\lambda)[qq]$, where $\Phi(U)$ is defined as in Eq. (2.5) and we have used the fact that $E(\Phi(U)) = H(\lambda)$.

To prove the upper bound, we use Lemma 2.2 to show that n uses of U together with LOCC can generate only mixtures of pure states with Schmidt number $\leq \operatorname{Sch}(U)^n = \exp(nH_0(\lambda))$. Since approximating $|\Phi\rangle^{\otimes nE(1-\delta)}$ to accuracy ϵ requires a mixture of pure states with expected Schmidt number $\geq (1-\epsilon)\exp(nE(1-\delta))$, asymptotically we must have $H_0(\lambda) \geq E(U)$. □

As a corollary, any nonlocal U has a nonzero $E(U)$. A similar, though less quantitative, result holds for communication as well.

Proposition 2.5. *If U is nonlocal then $\langle U\rangle \geq C[c \to c]$ for some $C > 0$.*

We state this here since we will need it for the proof of the next theorem, but defer the proof of Proposition 2.5 until Section 2.3 so as to focus on entanglement generation in this chapter.

General formula for entanglement capacity: The main result on the entanglement capacity is the following method of expressing it in terms of a single use of U:

Theorem 2.6.

$$E(U) = \Delta E_U := \sup_{|\psi\rangle \in \mathcal{H}_{\mathrm{A\,A'\,B\,B'}}} E\left((U_{AB} \otimes I_{A'B'})|\psi\rangle\right) - E(|\psi\rangle) \tag{2.8}$$

where the supremum ranges over Hilbert spaces A', B' of any finite dimension.

In other words, the asymptotic entanglement capacity $E(U)$ is equal to the largest *single-shot* increase of entanglement ΔE_U, if we are allowed to start with an arbitrary pure (possibly entangled) state. This result was independently obtained in [LHL03] and is based on a similar result for Hamiltonians in [DVC+01]. Here we restate the proof of [BHLS03] in the language of asymptotic resources.

Proof. $E(U) \leq \Delta E_U$ [converse]: Consider an arbitrary protocol that uses U n times in order to generate $\approx_\epsilon \Phi^{\otimes nE(U)(1-\delta)}$. We will prove a stronger result, in which even with unlimited classical communication U cannot generate more than ΔE_U ebits per use. Since communication is free, we assume that instead of discarding subsystems, Alice and Bob perform complete measurements and classically communicate their outcomes. Thus, we always work with pure states.

*Note that our definition of $\langle U\rangle$ differs slightly from the definition in [BHLS03]; whereas [BHLS03] allowed n sequential uses of U interspersed by local operations (i.e. the depth n resource $U^{\times n}$), we follow Definition 1.16 and allow only $(U^{\otimes n/k})^{\times k}$ where k depends only on the target inefficiency and not the desired accuracy of the protocol.

Since LOCC cannot increase expected entanglement and Alice and Bob start with a product state, their final state must have expected entanglement $\leq n\Delta E_U$. However, by Fannes' inequality (Lemma 1.1) the output state must have entanglement $\geq nE(U)(1-\delta)(1-\epsilon)-\eta(\epsilon)$. Thus $\forall\epsilon,\delta>0$ we can choose n sufficiently large that $E(U)(1-\delta)(1-\epsilon)-\eta(\epsilon)/n \leq \Delta E_U$, implying that $E(U)\leq\Delta E_U$.*

$E(U)\geq\Delta E_U$ *[coding theorem]:* Assume $\Delta E_U>0$; otherwise the claim is trivial. Recall from Eqns. (1.25) and (1.26) our formulation of entanglement concentration $\langle\psi\rangle\geq E(\psi)[qq]$ and dilution $E(\psi)[qq]+o[c\to c]\geq\langle\psi\rangle$. Then

$$\langle U\rangle + E(\psi)[qq] \geq \langle U\rangle + o[c\to c] + E(\psi)[qq] \geq \langle U\rangle + \langle\psi\rangle \geq \langle U(\psi)\rangle \geq E(U|\psi\rangle)[qq], \tag{2.9}$$

where we have used Proposition 2.5 in the first inequality, entanglement dilution in the second inequality, and entanglement concentration in the last inequality. Using the Cancellation Lemma (1.37), we find that $\langle U\rangle + o[qq] \geq E(U|\psi\rangle) - E(|\psi\rangle)[qq]$, and the sublinear $[qq]$ term can be removed due to Proposition 2.4 and the fact that $\Delta E_U>0$ implies $\mathrm{Sch}(U)>1$. Thus $\langle U\rangle\geq E(U|\psi\rangle)-E(|\psi\rangle)[qq]$ for all ψ. Taking the supremum over ψ and using the Closure Lemma (1.32) yields the desired result. □

The problem of finding $E(U)$ is now reduced to calculating the supremum in Eq. (2.8). To help understand the properties of Eq. (2.8), we now consider a number of possible variations on it, as well as some attempts at simplification.

- *Restricting the size of the ancilla appears hard:* Solving Eq. (2.8) requires optimizing over ancilla systems A′ and B′ of unbounded size. Unfortunately, we don't know if the supremum is achieved for any finite dimensional ancilla size, so we can't give an algorithm with bounded running time that reliably approximates $E(U)$. On the one hand, we know that ancilla systems are sometimes necessary. The two-qubit SWAP gate can generate no entanglement without entangled ancillas, and achieves its maximum of 2 ebits when acting on $|\Phi\rangle_{\mathrm{A\,A'}}|\Phi\rangle_{\mathrm{B\,B'}}$; a separation that is in a sense maximal. On the other hand, some gates, such as CNOT, can achieve their entanglement capacity with no ancillas. Less trivially, [CLVV03] proved that two-qubit Hamiltonians of the form $H=\alpha X\otimes X+\beta Y\otimes Y$ can achieve their entanglement capacity without ancilla systems, though this no longer holds when a $Z\otimes Z$ term is added.

  It is reasonable to assume that even when ancilla are necessary, it should suffice to take them to be the same size as the input systems. Indeed, no examples are known where achieving the entanglement capacity requires $\dim\mathrm{A}'>\dim\mathrm{A}$ or $\dim\mathrm{B}'>\dim\mathrm{B}$. On the other hand, there is no proof that the capacity is achieved for any finite-dimensional ancilla; we cannot rule out the possibility that there is only an infinite sequence of states that converges to the capacity.

- *Infinite dimensional ancilla don't help:* Though we cannot put an upper bound on the necessary dimensions of A′ and B′, we can assume that they are finite dimensional. In other words, we will show that ΔE_U is unchanged if we modify the sup in Eq. (2.8) to optimize over $|\psi\rangle\in\mathcal{H}_{\mathrm{A\,A'\,B\,B'}}$ s.t. $E(\psi)<\infty$ and $\dim\mathcal{H}_{\mathrm{A'}}=\dim\mathcal{H}_{\mathrm{B'}}=\infty$. Denote this modified supremum by $\Delta E_U'$. We will prove that $\Delta E_U=\Delta E_U'$.

  First, we state a useful lemma.

  Lemma 2.7. *Any bipartite state $|\psi\rangle$ with $E(\psi)<\infty$ can be approximated by a series of states $|\varphi_1\rangle,|\varphi_2\rangle,\ldots$, each with finite Schmidt number and obeying $\|\psi-\varphi_n\|_1\log\mathrm{Sch}(\varphi_n)\to 0$ as $n\to\infty$. (In other words, the error converges to zero faster than $1/\log\mathrm{Sch}(\varphi_n)$.)*

  Proof. Schmidt decompose $|\psi\rangle$ as $|\psi\rangle=\sum_{i=1}^{\infty}\sqrt{\lambda_i}|i\rangle|i\rangle$ and define the normalized state $|\varphi_n\rangle=\sum_{i=1}^{n}\sqrt{\lambda_i}|i\rangle|i\rangle/\sqrt{\sum_{i=1}^{n}\lambda_i}$. Let $\delta_n:=\frac{1}{2}\|\psi-\varphi_n\|_1=\sum_{i>n}\lambda_i$. Now, use the fact that $E(\psi)<\infty$

*A more formal (and general) version of this argument will also appear in the proof of Theorem 3.7 in Section 3.4.2.

and $\lambda_n \leq 1/n$ to obtain

$$E(\psi) - \sum_{i=1}^{n} \lambda_i \log(1/\lambda_i) = \sum_{i=n+1}^{\infty} \lambda_i \log(1/\lambda_i) \geq \sum_{i=n+1}^{\infty} \lambda_i \log(1/\lambda_n) = \delta_n \log(1/\lambda_n) \geq \delta_n \log n \tag{2.10}$$

Since the term on the left converges to 0 as $n \to \infty$, we also have that $\delta_n \log n \to 0$ as $n \to \infty$. Using $n = \mathrm{Sch}(\varphi_n)$ and $\delta_n = \frac{1}{2}\|\psi - \varphi_n\|_1$, our desired result follows. □

Now $\forall \epsilon > 0, \exists |\psi\rangle \in \mathcal{H}_{\mathrm{A\,A'\,B\,B'}}$ with $\dim \mathcal{H}_{A'} = \dim \mathcal{H}_{B'} = \infty$ such that $E(U|\psi\rangle) - E(|\psi\rangle) > \Delta E'_U - \epsilon$. By Lemma 2.7, we can choose $|\varphi\rangle$ with $\mathrm{Sch}(\varphi) < \infty$ (and thus can belong to $\mathcal{H}_{\mathrm{A\,A'\,B\,B'}}$ with $\dim \mathrm{A'}, \dim \mathrm{B'} < \infty$) such that $\|\psi - \varphi\|_1 \log \mathrm{Sch}(\varphi) \leq \epsilon$. By Fannes' inequality (Lemma 1.1), $|E(\psi) - E(\varphi)| \leq \epsilon + \eta(\epsilon)/\log \mathrm{Sch}\,\varphi$ and $|E(U|\psi\rangle) - E(U|\varphi\rangle)| \leq (\epsilon + \eta(\epsilon))(1 + (\log \mathrm{Sch}(\varphi))/(\log \mathrm{Sch}(U)))$ (since $\mathrm{Sch}(U|\varphi\rangle) \leq \mathrm{Sch}(U)\,\mathrm{Sch}(\varphi)$). Combining these, we find that $E(U|\varphi\rangle) - E(|\varphi\rangle) \to \Delta E'_U$ as $\epsilon \to 0$, implying that $\Delta E_U = \Delta E'_U$.

- *Sometimes it helps to start with entanglement:* Subtracting one entropy from another in Eq. (2.8) is rather ugly; it would be nice if we could eliminate the second term (and at the same time restrict $\dim \mathrm{A'} \leq \dim \mathrm{A}$ and $\dim \mathrm{B'} \leq \dim \mathrm{B}$) by maximizing only over product state inputs. However, this would result in a strictly lower capacity for some gates. This is seen most dramatically for Hamiltonian capacities, for which $\frac{d}{dt}E(e^{-iHt}|\alpha\rangle|\beta\rangle) = 0$ for any $|\alpha\rangle \in \mathcal{H}_{AA'}, \beta \in \mathcal{H}_{\mathrm{B\,B'}}$, due to the quantum Zeno effect: after a small amount of time t, the largest Schmidt coefficient is $1 - \mathcal{O}(t^2)$. The same principle applies to the gate $U = e^{-iHt}$ for t sufficiently small: the entanglement capacity is $\mathcal{O}(t)$ (because $\mathcal{O}(1/t)$ uses of U give a gate far from the identity with $\mathcal{O}(1)$ entanglement capacity), though the most entanglement that can be created from unentangled inputs by one use of U is $\mathcal{O}(t^2 \log(1/t))$.

  As a corollary, the lower bound of Proposition 2.4 is not tight for all gates.

- *Mixed states need not be considered:* We might also try optimizing over density matrices rather than pure states. For this to be meaningful, we need to replace the entropy of entanglement with a measure of mixed-state entanglement[BDSW96], such as entanglement of formation $E_f(\rho) := \min\{\sum_i p_i E(\psi_i) : \rho = \sum_i p_i \psi_i\}$, entanglement cost $E_c(\rho) := \inf_m \frac{1}{m} E_f(\rho^{\otimes m}) = \inf\{e : e[qq] + \infty[c \to c] \geq \langle\rho\rangle\}$, or distillable entanglement $D(\rho) := \sup\{e : \langle\rho\rangle + \infty[c \to c] + \infty[c \leftarrow c] \geq e[qq]\}$ [BDSW96].

  We claim that $\Delta E_U = \sup_\rho E_f(U(\rho)) - E_f(\rho) = \sup_\rho E_c(U(\rho)) - E_c(\rho) = \sup_\rho D(U(\rho)) - E_c(\rho)$. To prove this for E_f, decompose an arbitrary ρ^{AB} into pure states as $\rho = \sum_i p_i \psi_i$ s.t. $E_f(\rho) = \sum_i p_i E(\psi_i)$. Now we use the convexity of E_f to show that $E_f(U(\rho)) = E_f(\sum_i p_i U(\psi_i)) \leq \sum_i p_i E(U(\psi_i))$, implying that

  $$E_f(U(\rho)) - E_f(\rho) \leq \sum_i p_i \left[E(U(\psi_i)) - E(\psi_i)\right] \leq \max_i E(U(\psi_i)) - E(\psi_i).$$

  Thus, any increase in E_f can be achieved by a pure state.

  A similar, though slightly more complicated, argument applies for E_c. For any $\epsilon > 0$ and any ρ, there exists m sufficiently large that $E_c(\rho) + \epsilon \geq \frac{1}{m}\sum_i p_i E(\psi_i)$ for some $\{p_i, \psi_i\}$ such that $\rho^{\otimes m} = \sum_i p_i \psi_i$. Using first the definition of E_c and then convexity, we have $E_c(U(\rho)) \leq \frac{1}{m} E_f(U(\rho)^{\otimes m}) \leq \frac{1}{m}\sum_i p_i E(U^{\otimes m}(\psi_i))$. Thus,

  $$\begin{aligned} E_c(U(\rho)) - E_c(\rho) - \epsilon &\leq \frac{1}{m}\sum_i p_i \left[E(U^{\otimes m}(\psi_i)) - E(\psi_i)\right] \leq \max_i (E(U^{\otimes m}(\psi_i)) - E(\psi_i))/m \\ &\leq \max_i \max_{j \in \{1,\ldots,m\}} E((U^{\otimes j} \otimes I^{\otimes m-j})(\psi_i)) - E((U^{\otimes j-1} \otimes I^{\otimes m-j+1})(\psi_i)) \leq \Delta E_U. \end{aligned}$$

This proof implicitly uses the fact that $E(U)$ is (sub)additive; i.e. $E(U^{\otimes 2}) = 2E(U)$.

Finally, $\Delta E_U = \sup_\rho D(U(\rho)) - E_c(\rho)$ because of the E_c result from the last paragraph and the fact that $D(\rho) \leq E_c(\rho)$. This case corresponds to the operationally reasonable scenario of paying $E_c(\rho)[qq]$ for the input state and getting $D(U(\rho))[qq]$ from the output state. Of course, this case also follows from the fact that classical communication doesn't help entanglement capacity.

Contrasting the entanglement capacity of unitary gates and noisy quantum channels: The problem of generating entanglement with a unitary gate turns out to have a number of interesting differences from the analogous problem of using a noisy quantum channel to share entanglement. Here we survey some of those differences.

- *Free classical communication doesn't help:* In the proof of the converse of Theorem 2.6, we observed that unlimited classical communication in both directions doesn't increase the entanglement capacity. For noisy quantum channels, it is known that forward communication doesn't change the entanglement capacity[BDSW96], though in some cases back communication can improve the capacity (e.g. back communication increases the capacity of the 50% erasure channel from zero to 1/2) and two-way communication appears to further improve the capacity[BDSS04].

- *Quantum and entanglement capacities appear to be different:* A noisy quantum channel $\mathcal{N}$ has the same capacity to send quantum data that it has to generate entanglement (i.e. $\langle \mathcal{N} \rangle \geq Q[q \to q]$ iff $\langle \mathcal{N} \rangle \geq Q[qq]$)[BDSW96], though with free classical back communication this is no longer thought to hold[BDSS04]. Since unitary gates are intrinsically bidirectional, we might instead ask about their total quantum capacity $Q_+(U) := \max\{Q_1 + Q_2 : \langle U \rangle \geq Q_1[q \to q] + Q_2[q \leftarrow q]\}$ and ask whether it is equal to $E(U)$. All that is currently known is the bound $Q_+(U) \leq E(U)$, which is saturated for gates like CNOT and SWAP. However, in Section 2.4.3, I will give an example of a gate that appears to have $Q_+(U) < E(U)$, though this conjecture is supported only by heuristic arguments.

- *Entanglement capacities are strongly additive:* For any two bipartite gates U_1 and U_2, we have $E(U_1 \otimes U_2) \geq E(U_1) + E(U_2)$, since we can always run the optimal entanglement generating protocols of U_1 and U_2 in parallel. On the other hand, $E(U_1 \otimes U_2) = \sup_\psi E((U_1 \otimes U_2)|\psi\rangle) - E(|\psi\rangle) = \sup_\psi [E((U_1 \otimes U_2)|\psi\rangle) - E((U_1 \otimes I)|\psi\rangle)] + [E((U_1 \otimes I)|\psi\rangle) - E(|\psi\rangle)] \leq \Delta E_{U_2} + \Delta E_{U_1} = E(U_2) + E(U_1)$. Thus $E(U_1 \otimes U_2) = E(U_1) + E(U_2)$.

  In contrast, quantum channel capacities (equivalently either for quantum communication or entanglement generation) appear to be superadditive[SST01].

- *Entanglement capacities are always nonzero:* If U is a nonlocal gate (i.e. cannot be written as $U = U_A \otimes U_B$), then according to Proposition 2.4, $E(U) > 0$. On the other hand, there exist nontrivial quantum channels with zero entanglement capacity: classical channels cannot create entanglement and bound entangled channels cannot be simulated classically, but also cannot create any pure entanglement.

2.3 Classical communication capacity

Nonlocal gates can not only create entanglement, but can also send classical messages both forward (from Alice to Bob) and backwards (from Bob to Alice). Therefore, instead of a single capacity, we need to consider an achievable classical rate region. Define $\mathrm{CC}(U) := \{(C_1, C_2) : \langle U \rangle \geq C_1[c \to c] + C_2[c \leftarrow c]\}$. Some useful special cases are the forward capacity $C_\to(U) = \max\{C_1 : (C_1, 0) \in \mathrm{CC}(U)\}$, backward capacity $C_\leftarrow(U) = \max\{C_2 : (0, C_2) \in \mathrm{CC}(U)\}$ and bidirectional capacity $C_+(U) =$

$\max\{C_1 + C_2 : (C_1, C_2) \in \mathrm{CC}(U)\}$. (By Lemma 1.9 CC($U$) is a closed set, so these maxima always exist.)

We can also consider the goal of simultaneously transmitting classical messages and generating entanglement. Alternatively, one might want to use some entanglement to help transmit classical messages. We unify these scenarios and others by considering the three-dimensional rate region $\mathrm{CCE}(U) := \{(C_1, C_2, E) : \langle U\rangle \geq C_1[c \to c] + C_2[c \leftarrow c] + E[qq]\}$. When some of C_1, C_2 and E are negative, it means that the resource is being consumed; for example, if $E < 0$ and $C_1, C_2 \geq 0$, then the resource inequality $\langle U\rangle + (-E)[qq] \geq C_1[c \to c] + C_2[c \leftarrow c]$ represents entanglement-assisted communication. Some useful limiting capacities are $C^E_{\to}(U) := \max\{C_1 : (C_1, 0, -\infty) \in \mathrm{CCE}(U)\}$, $C^E_{\leftarrow}(U) := \max\{C_2 : (0, C_2, -\infty) \in \mathrm{CCE}(U)\}$ and $C^E_{+}(U) := \max\{C_1 + C_2 : (C_1, C_2, -\infty) \in \mathrm{CCE}(U)\}$.

To get a sense of what these capacity regions can look like, Fig. 2-1 contains a schematic diagram for the achievable region CC(U) and the definitions of the various capacities when we set $E = 0$. We present all the *known* properties and intentionally show the features that are not ruled out, such as the asymmetry of the region, and the nonzero curvature of the boundary.

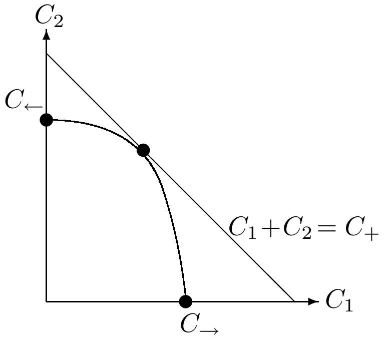


Figure 2-1: Example of a possible achievable rate region CC(U), with the limiting capacities of $C_{\to}, C_{\leftarrow}$ and C_+ indicated.

There are much simpler examples – the unassisted achievable region for CNOT and SWAP are similar triangles with vertices $\{(0,0),(0,1),(1,0)\}$ and $\{(0,0),(0,2),(2,0)\}$ respectively (see Section 2.4.1).

In general, little is known about the unassisted achievable region of (C_1, C_2) besides the convexity and the monotonicity of its boundary. The most perplexing question is perhaps whether the region has reflective symmetry about line $C_1 = C_2$, which would imply that $C_{\to}(U) = C_{\leftarrow}(U)$. Eq. (2.1) shows that any two-qubit gate or Hamiltonian is locally equivalent to one with Alice and Bob interchanged, so that the achievable region is indeed symmetric. In higher dimensions, on the other hand, [BCL$^+$02] shows that there are Hamiltonians (and so unitary gates) that are intrinsically asymmetric. However, it remains open whether the achievable rate pairs are symmetric, or more weakly, whether $C_{\to} = C_{\leftarrow}$.

The rest of this section is as follows:

Section 2.3.1 proves some basic facts about the achievable classical communication region. Then we establish some bounds on communication rates similar to, but weaker than, the bounds on entanglement rate in Proposition 2.4.

Section 2.3.2 proves a capacity formula for $C^E_{\to}(U)$ (or equivalently $C^E_{\leftarrow}(U)$) that parallels the formula in Theorem 2.6. This formula will be improved in the next chapter when we introduce coherent classical communication.

Section 2.3.3 discusses relations between the classical communication and the entanglement generation capacities of unitary gates.

Section 2.3.4 explores the difficulties involved in proving capacity theorems for bidirectional communication.

2.3.1 General facts about the achievable classical communication rate region

We begin with some basic facts about CCE.

- *Monotonicity: If* $(C_1, C_2, E) \in \mathrm{CCE}(U)$ *then* $(C_1 - \delta_1, C_2 - \delta_2, E - \delta_3) \in \mathrm{CCE}(U)$ *for any* $\delta_1, \delta_2, \delta_3 \geq 0$. This is because we can always choose to discard resources.

- *Convexity:* $\mathrm{CCE}(U)$ *is a convex set.* This follows from time-sharing (part 2 of Theorem 1.22 and part 3 of Lemma 1.25.

- *Classical feedback does not help: If* $(C_1, C_2, E) \in \mathrm{CCE}(U)$*, then* $(C_1, 0, E) \in \mathrm{CCE}(U)$ *and* $(0, C_2, E) \in \mathrm{CCE}(U)$*.* We mention this fact now, but defer the proof until Chapter 3.

  Combining this with monotonicity and the fact that classical feedback doesn't improve entanglement capacity, we obtain as a corollary that $\mathrm{CCE}(U) \subseteq [-\infty, C^E_{\rightarrow}(U)] \times [-\infty, C^E_{\leftarrow}(U)] \times [-\infty, E(U)] \subseteq [\infty, 2\log d] \times [\infty, 2\log d] \times [\infty, 2\log d]$. This second inclusion depends on Proposition 2.8, proven below.

- *No more than* $E(U^\dagger)$ *ebits are ever needed: If* $(C_1, C_2, E) \in \mathrm{CCE}(U)$*, then* $(C_1, C_2, -E(U^\dagger)) \in \mathrm{CCE}(U)$. A proof of this will be sketched in Section 2.3.3, and it also follows from Theorem 3.1 in the next chapter.

- *Shared randomness does not help: If* $\langle U\rangle + \infty[cc] \geq C_1[c \to c] + C_2[c \leftarrow c] + E[qq]$*, then* $(C_1, C_2, E) \in \mathrm{CCE}(U)$*.*

  This is due to a standard derandomization argument (further developed in [CK81, DW05b]). Let r denote the shared randomness and let $x := (a, b)$ run over all possible messages sent by Alice and Bob with n uses of U (a set of size $\leq \exp(Cn)$ for $C := C_1 + C_2$). If $e_{x,r}$ is the corresponding probability of error, then our error-correcting condition is that $\max_x \mathbb{E}_r e_{x,r} \leq \epsilon$. Now sample m copies of the shared randomness, $(r_1, \ldots, r_m) =: \vec{r}$, where m is a parameter we will choose later. According to Hoëffding's inequality[Hoë63], we have

  $$\Pr_{\vec{r}}\left[\frac{1}{m}\sum_{i=1}^{m} e_{x,r} \geq 2\epsilon\right] \leq \exp(-m\epsilon^2/2), \tag{2.11}$$

  for any particular value of x. We apply the union bound over all $\leq \exp(Cn)$ values of x to obtain

  $$\Pr_{\vec{r}}\left[\max_x \frac{1}{m}\sum_{i=1}^{m} e_{x,r} \geq 2\epsilon\right] \leq \exp(Cn - m\epsilon^2/2). \tag{2.12}$$

  Thus, if we choose $m > 2Cn/\epsilon^2$, then there exists a choice of $\vec{r}$ with maximum error $\leq 2\epsilon$. If Alice and Bob preagree on $\vec{r}$, then they need only $\log m$ bits of shared randomness to agree on which r_i to use. Since $\log m = \mathcal{O}(\log n + \log(1/\epsilon))$, this randomness can be generated by a negligible amount of extra communication.

We now state an upper bound, originally due to [CGB00].

Proposition 2.8. *If* $U \in \mathcal{U}_{d\times d}$*, then* $C^E_{\rightarrow}(U) \leq 2\log d$ *and* $C^E_{\leftarrow}(U) \leq 2\log d$*.*

Proof. The proof is based on simulating U with teleportation: Alice teleports her input to Bob using $2\log d[c \to c] + \log d[qq]$, Bob applies U locally (and hence for free), and then Bob teleports Alice's half of the state back using $2\log d[c \leftarrow c] + \log d[qq]$. Thus we obtain the resource inequality

$$2\log d\left([c \to c] + [c \leftarrow c] + [qq]\right) \geq \log d\left([q \to q] + [q \leftarrow q]\right) \geq \langle U\rangle \tag{2.13}$$

Allowing free entanglement and back communication yields $2\log d[c \to c] + \infty[q \leftarrow q] \geq C^E_{\to}(U)[c \to c]$. Causality[Hol73] implies that $C^E_{\to}(U) \leq 2\log d$. A similar argument proves that $C^E_{\leftarrow}(U) \leq 2\log d$. □

It is an interesting open question whether any good bounds on classical capacity can be obtained as functions of a gate's Schmidt coefficients, as we found with Proposition 2.4 for the case of entanglement generation.

We now prove Proposition 2.5, which stated that any nonlocal U has a nonzero classical capacity. An alternate proof can be found in [BGNP01].

Proof of Proposition 2.5. Let E_0 the amount of entanglement created by applying U to the AB registers of $|\Phi_d\rangle_{AA'}|\Phi_d\rangle_{BB'}$. If U is nonlocal, then $E_0 > 0$ according to Proposition 2.4.

Alice can send a noisy bit to Bob with the following t-use protocol. Bob inputs $|\Phi_d\rangle^{\otimes t}_{BB'}$ to all t uses of U. To send "0" Alice inputs $|\Phi_d\rangle^{\otimes t}_{AA'}$ to share tE_0 ebits with Bob, i.e. inputting a fresh copy of $|\Phi_d\rangle$ each time. To send "1", Alice inputs $|0\rangle_A$ to the first use of U, takes the output and uses it as the input to the second use, and so on. Alice only interacts a d-dimensional register throughout the protocol, so their final entanglement is no more than $\log d$. Thus different messages from Alice result in very different amounts of entanglement at the end of the protocol.

Let ρ_0 and ρ_1 denote Bob's density matrices when Alice sends 0 or 1 respectively. Using Fannes' inequality (Lemma 1.1), $tE_0 - \log d \leq \log d \; \|\rho_0 - \rho_1\|_1 + \frac{\log e}{e}$. If we choose $t > (\log d + \frac{\log e}{e})/E_0$, then $\rho_0 \neq \rho_1$ and Bob has a nonzero probability of distinguishing ρ_0 from ρ_1 and thereby identifying Alice's message. Thus the t-use protocol simulates a noisy classical channel with nonzero capacity and $C_{\to}(U) > 0$. □

2.3.2 Capacity theorem for entanglement-assisted one-way classical communication

We conclude the section with a general expression for $C^E_{\to}(U)$. Though we will improve it in Chapter 3 to characterize the entire one-way tradeoff region $\mathrm{CE}(U) := \{(C,E) : (C,0,E) \in \mathrm{CCE}(U)\}$, the proof outlines useful principles which we will later use.

First, we recall some notation from our definition of remote state preparation (RSP) in Section 1.4. Let

$$\mathcal{E} = \sum_i p_i \, |i\rangle\langle i|^{X_A} \otimes |\psi_i\rangle\langle\psi_i|^{A_1A_2B_1B_2} \tag{2.14}$$

be an ensemble of bipartite states $|\psi_i\rangle$, where Alice holds the index i, U acts on A_1B_1 and A_2, B_2 are ancilla spaces. Thus we can define $U(\mathcal{E})$ by

$$U(\mathcal{E}) := \sum_i p_i \, |i\rangle\langle i|^{X_A} \otimes (U^{A_1B_1} \otimes \mathbb{1}^{B_1B_2})(|\psi_i\rangle\langle\psi_i|^{A_1A_2B_1B_2}) \tag{2.15}$$

We will use A to denote the composite system A_1A_2 and B to denote B_1B_2. As in Section 1.4, define the $\{c \to q\}$ channel $\mathcal{N}_{\mathcal{E}}$ by $\mathcal{N}_{\mathcal{E}}(|i\rangle\langle i|) = |i\rangle\langle i| \otimes \psi_i$, so that that $\mathcal{E} = \mathcal{N}_c E(\mathcal{E}^{X_A})$. Defining $\mathcal{N}_{U(\mathcal{E})}$ similarly, we can use Lemma 1.31 to show that

$$\langle \mathcal{N}_{\mathcal{E}} : \mathcal{E}^{X_A}\rangle + \langle U\rangle \geq \langle U \circ \mathcal{N}_{\mathcal{E}} : \mathcal{E}^{X_A}\rangle = \langle \mathcal{N}_{U(\mathcal{E})} : \mathcal{E}^{X_A}\rangle. \tag{2.16}$$

Recall from HSW coding (Eq. (1.42)) that

$$\langle \mathcal{N}_{\mathcal{E}} : \mathcal{E}^{X_A}\rangle \geq I(X_A; B)_{\mathcal{E}}[c \to c], \tag{2.17}$$

while RSP (Eq. (1.38)) states that

$$I(X_A;B)_{\mathcal{E}}[c\to c]+H(B)[q\,q]\geq\langle\mathcal{N}_{\mathcal{E}}:\mathcal{E}^{X_A}\rangle. \tag{2.18}$$

In the presence of free entanglement, these resource inequalities combine to become an equality:

$$I(X_A;B)_{\mathcal{E}}[c\to c]+\infty[q\,q]=\langle\mathcal{N}_{\mathcal{E}}:\mathcal{E}^{X_A}\rangle+\infty[q\,q]. \tag{2.19}$$

This remarkable fact can be thought of as a sort of reverse Shannon theorem for $\{c\to q\}$ channels, stating that when entanglement is free (in contrast to the CRST, which requires free rbits), any $\{c\to q\}$ channel on a fixed source is equivalent to an amount of classical communication given by its capacity.

Recall the similar equality for partially entangled states in the presence of a sublinear amount of classical communication: $\langle\psi^{AB}\rangle+o[c\to c]=H(B)_\psi[q\,q]+o[c\to c]$. By analogy with entanglement generation in Theorem 2.6, we will use the resource equality in Eq. (2.19) to derive a capacity theorem for classical communication in the presence of unlimited entanglement.

Theorem 2.9.

$$C_{\to}^{E}(U)=\Delta\chi_U:=\sup_{\mathcal{E}}\left[I(X_A;B)_{U(\mathcal{E})}-I(X_A;B)_{\mathcal{E}}\right] \tag{2.20}$$

where the supremum is over all ensembles $\mathcal{E}$ of the form in Eq. (2.14).

The proof closely follows the proof of Theorem 2.6.

Proof. We begin with the converse, proving that $C_{\to}^{E}(U)\leq\Delta\chi_U$. Alice and Bob begin with a fixed input state, which can be thought of as an ensemble $\mathcal{E}_0$ with $I(X_A;B)_{\mathcal{E}}=0$. Local operations (which for simplicity, we can assume are all isometries) cannot increase $I(X_A;B)$, so after n uses of U the mutual information must be $\leq n\Delta\chi_U$. (For a generalized and more formal verson of this argument, see the proof of Theorem 3.7 in Section 3.4.2.) The bound $C_{\to}^{E}(U)\leq\Delta\chi_U$ then follows from Fannes' inequality.

Coding theorem: For any ensemble $\mathcal{E}$, we have $\langle U\rangle+I(X_A;B)_{\mathcal{E}}[c\to c]+\infty[qq]\geq\langle U\rangle+\langle\mathcal{N}_{\mathcal{E}}:\mathcal{E}^{X_A}\rangle+\infty[qq]\geq\langle\mathcal{N}_{U(\mathcal{E})}:\mathcal{E}^{X_A}\rangle+\infty[qq]\geq I(X_A;B)_{U(\mathcal{E})}[c\to c]+\infty[qq]$. Using the Cancellation Lemma (1.37) and taking the supremum over $\mathcal{E}$, we find that $\langle U\rangle+o[c\to c]+\infty[qq]\geq\Delta\chi_U[c\to c]+\infty[qq]$. Finally, we can use Proposition 2.5 and Lemma 1.36 to eliminate the sublinear classical communication cost. □

Although the coding theorem is formally very similar to the coding theorem for entanglement generation, its implementation looks rather different. Achieving the bound in Theorem 2.6 is rather straightforward: 1) n_1 copies are created of some state $\psi^{A_1A_2B_1B_2}$ s.t. $\Delta E_U\approx H(B)_{U(\psi)}-H(B)_\psi$, 2) $U^{\otimes n_1}$ is applied to $\psi^{\otimes n_1}$, 3) entanglement is concentrated from $(U|\psi\rangle)^{\otimes n_1}$, 4) $\approx n_1H(B)_\psi$ ebits are used to recreate $\psi^{\otimes n_1}$ and $\approx n_1(H(B)_{U(\psi)}-H(B)_\psi)\approx n_1\Delta E_U$ ebits are set aside as output, 5) steps 2-4 are repeated n_2 times to make the cost of the catalyst vanish. The coding scheme for entanglement-assisted classical communication is similar, but has some additional complications because different parts of the message are not interchangeable. The resulting protocol involves a peculiar preprocessing step in which Alice runs through the entire protocol backwards before U is used for the first time; for this reason, we call it the "looking-glass protocol." The procedure is as follows:

1. Choose an ensemble $\mathcal{E}=\sum_i p_i\,|i\rangle\!\langle i|\otimes\psi_i$ with $I(X_A;B)_{U(\mathcal{E})}-I(X_A;B)_{\mathcal{E}}\approx\Delta\chi_U$.

2. The message is broken into n_1 blocks $M_1,\ldots,M_{n_1}$, each of length $\approx n_2\Delta\chi_U$. Initialize R_{n_1} to be an arbitrary string of length $\approx n_2I(X_A;B)_{\mathcal{E}}$.

3. For $k=n_1,n_1-1,\ldots,1$:

(a) Encode the string (R_k, M_k) ($\approx n_2 I(X_A; B)_{U(\mathcal{E})}$ bits) into an element of $(U(\mathcal{E}))^{\otimes n_2}$, say $U|\psi_{x_{k,1}}\rangle \otimes \cdots \otimes U|\psi_{x_{k,n_2}}\rangle$ for some p-typical string $x_k^{n_2}$. This is accomplished via HSW coding.

(b) Alice now wishes to use RSP to send $|\psi_{x_k^{n_2}}\rangle := |\psi_{x_{k,1}}\rangle \otimes \cdots \otimes |\psi_{x_{k,n_2}}\rangle$ to Bob. She performs the RSP measurement on some shared entanglement and obtains an outcome with $\approx n_2 I(X_A; B)_{\mathcal{E}}$ bits, which she doesn't send to Bob directly, but instead stores in the register R_{k-1}.

4. Finally, Alice sends R_0 to Bob using $\approx n_2 I(X_A; B)_{\mathcal{E}}[c \to c]$.

5. For $k = 1, \ldots, n_1$:

(a) Bob uses R_{k-1} to perform his half of RSP and reconstruct his half of $|\psi_{x_k^{n_2}}\rangle$.

(b) Alice and Bob apply U n_2 times to obtain $\approx U^{\otimes n_2}|\psi_{x_k^{n_2}}\rangle$.

(c) Bob performs HSW decoding to obtain (M_k, R_k) with a high probability of success.

It might seem that errors and inefficiencies from the many HSW and RSP steps accumulate dangerously over the many rounds of the looking-glass protocol. In [BHLS03], the protocol was carefully analyzed and the errors and inefficiency were shown to converge to zero. However, the validity of the composite protocol follows even more directly from the Composability Theorem (1.22); remarkably, this permits a proof that is much more compact and intuitive than even the description of the above protocol, let alone a verification of its correctness.

As a corollary of Theorem 2.9, entanglement-assisted capacities are additive (i.e. $C^E_{\rightarrow}(U_1 \otimes U_2) = C^E_{\rightarrow}(U_1) + C^E_{\rightarrow}(U_2)$). The proof is basically the same as the proof that $E(U)$ is additive.

Another corollary we can obtain is an optimal coding theorem for entanglement-assisted one-way quantum communication: $Q^E_{\rightarrow}(U) := \max\{Q : \langle U\rangle + \infty[qq] \geq Q[q \to q]\} = C^E_{\rightarrow}(U)/2$. This is because when entanglement is free, teleportation and super-dense coding imply that 2 cbits are equivalent to 1 qubit.

2.3.3 Relations between entanglement and classical communication capacities

One of the most interesting properties of unitary gates as communication channels is that their different capacities appear to be closely related. In this section we prove that $C_+(U) \leq E(U)$ and then discuss some similar bounds.

Proposition 2.10. *If $(C_1, C_2, E) \in \mathrm{CCE}(U)$ then $E(U) \geq C_1 + C_2 + E$.*

Using the fact that back communication does not improve capacities (proved in the next chapter), we can improve this bound to $E(U) \geq \max(C_1, 0) + \max(C_2, 0) + E$.

This claim is significant for two reasons. First is that it implies that it may be easier to connect different unitary gate capacities than it has been to relate different capacities of noisy channels. It is directly useful in finding gate capacities and raises the intriguing question of whether the converse inequality of Proposition 2.5 (that $E(U) > 0 \Rightarrow C_{\rightarrow}(U) > 0$) can be strengthened, and ultimately whether $C_+(U) = E(U)$.

The fact that $C_+(U) \leq E(U)$ has a deeper implication as well, which is that not all classical communication is created equal. While normally $[c \to c] \not\geq [qq]$, a cbit sent through unitary means *can* be converted into entanglement. This suggests that using unitary gates to communicate gives us something stronger than classical bits; a resource that we will formally define in the next chapter as *coherent bits* or cobits. The consequences will be productive not only for the study of unitary gate capacities, but also for many other problems in quantum Shannon theory.

Proof of Proposition 2.10. Assume for now that $E \geq 0$. For any n, there is a protocol $\mathcal{P}_n$ that uses U n times to send $C_1^{(n)}$ cbit($\rightarrow$) $+C_2^{(n)}$ cbit($\leftarrow$) and create $E^{(n)}$ ebits with $C_1^{(n)} \geq n(C_1 - \delta_n)$, $C_2^{(n)} \geq n(C_2 - \delta_n)$, $E^{(n)} \geq n(E - \delta_n)$ and error $\leq \epsilon_n$, where $\delta_n, \epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. We analyze the protocol using the QP formalism, in which $\mathcal{P}_n$ is an isometry such that for any $a \in \{0,1\}^{C_1^{(n)}}, b \in \{0,1\}^{C_2^{(n)}}$,

$$\begin{aligned} |\varphi_{ab}\rangle \quad &:= \quad \mathcal{P}_n|a\rangle_{A_1}|b\rangle_{B_1} \\ \text{and} \quad & F\left(\,|b\rangle_{A_1}|a\rangle_{B_1}|\Phi\rangle_{A_2B_2}^{\otimes E^{(n)}}, \mathrm{Tr}_{A_3B_3}\,|\varphi_{ab}\rangle\langle\varphi_{ab}|_{A_{1,2,3}B_{1,2,3}}\,\right) = 1 - \epsilon_{ab} \geq 1 - \epsilon_n\,. \end{aligned} \tag{2.21}$$

for some $\epsilon_{ab} \leq \epsilon_n$. By Uhlmann's Theorem[Uhl76], there exist normalized (though not necessarily orthogonal) states $|\gamma_{ab}\rangle$ and $|\eta_{ab}\rangle$ satisfying

$$|\varphi_{ab}\rangle = \sqrt{1-\epsilon_n}|b\rangle_{A_1}|a\rangle_{B_1}|\Phi\rangle_{A_2B_2}^{E^{(n)}}|\gamma_{ab}\rangle_{A_3B_3} + \sqrt{\epsilon_n}|\eta_{ab}\rangle_{A_{1,2,3}B_{1,2,3}}. \tag{2.22}$$

Note that we have changed ϵ_{ab} to ϵ_n by an appropriate choice of $|\eta_{ab}\rangle$. This will simplify the analysis later.

To generate entanglement, Alice and Bob will apply $\mathcal{P}_n$ to registers A_1B_1 that are maximally entangled with local ancillas A_4B_4; i.e. the states $|\Phi\rangle_{A_1A_4}^{\otimes C_1^{(n)}} = 2^{-C_1^{(n)}/2}\sum_a |a\rangle_{A_1}|a\rangle_{A_4}$ and $|\Phi\rangle_{B_1B_4}^{\otimes C_2^{(n)}} = 2^{-C_2^{(n)}/2}\sum_b |b\rangle_{B_1}|b\rangle_{B_4}$. The resulting output state is

$$|\overline{\varphi}_n\rangle_{AB} = \sqrt{1-\epsilon_n}|\psi_n\rangle_{AB} + \sqrt{\epsilon_n}|\delta_n\rangle_{AB}, \tag{2.23}$$

where

$$|\psi_n\rangle_{AB} = 2^{-(C_1^{(n)}+C_2^{(n)})/2}\sum_{a,b}|b\rangle_{A_1}|a\rangle_{A_4}|a\rangle_{B_1}|b\rangle_{B_4}|\Phi\rangle_{A_2B_2}^{\otimes E^{(n)}}|\gamma_{ab}\rangle_{A_3B_3}. \tag{2.24}$$

A similar expression exists for $|\delta_n\rangle_{AB}$, but it is not needed, so we omit it. Note that every Schmidt coefficient of $|\psi_n\rangle$ is $\leq \exp(-(C_1^{(n)} + C_2^{(n)} + E^{(n)}))$, so $E(|\psi_n\rangle) \geq C_1^{(n)} + C_2^{(n)} + E^{(n)}$.

We will use Fannes' inequality (Lemma 1.1) to relate $E(|\overline{\varphi}_n\rangle)$ to $E(|\psi_n\rangle)$. From Eq. (2.23), we have $|\langle\overline{\varphi}_n|\psi_n\rangle| \geq \sqrt{1-\epsilon_n}$. Applying the relation between fidelity and trace distance in Eq. (1.4), we find $\|\varphi_n - \psi_n\|_1 \leq 2\sqrt{\epsilon_n}$. Also, $|\overline{\varphi}_n\rangle$ was created with n uses of U, so $\mathrm{Sch}(|\overline{\varphi}_n\rangle) \leq (\mathrm{Sch}(U))^n \leq d^{2n}$. Thus

$$\begin{aligned} |E(|\psi_n\rangle) - E(|\varphi_n\rangle)| \quad &\leq \quad (2n\log d)\,2\sqrt{\epsilon_n} + \eta(2\sqrt{\epsilon_n}) \\ E(|\overline{\varphi}_n\rangle) \quad &\geq \quad n\left(C_1 + C_2 + E - 3\delta_n - 4\sqrt{\epsilon_n}\log d - \frac{\eta(2\sqrt{\epsilon_n})}{n}\right) \end{aligned} \tag{2.25}$$

Therefore as $n \rightarrow \infty$, $\frac{1}{n}E(|\overline{\varphi}_n\rangle) \rightarrow C_1 + C_2 + E$. Since $n\langle U\rangle \overset{*}{\geq} \langle\overline{\varphi}_n\rangle$, it follows that $\langle U\rangle \geq (C_1 + C_2 + E)[qq]$.

We omit the quite similar proof of the $E < 0$ case; however, note that this case also follows from the more general Theorem 3.1, which will be proved in Section 3.5. □

A similar bound exists for the entanglement-assisted capacity: $C_+^E(U) \leq E(U) + E(U^\dagger)$. This result is proved in [BS03a], though some preliminary steps are found in [BHLS03, BS03b]. Here we give a sketch of the argument and explain its evolution through [BS03b, BHLS03, BS03a].

As in Proposition 2.10, Alice and Bob will input halves of maximally entangled states into a communication protocol $\mathcal{P}_n$ that uses U n times. This creates $\approx nC_+^E(U)$ ebits. However, the entanglement assistance leads to two additional complications. First, we need to bound the amount of entanglement that $\mathcal{P}_n$ uses to communicate. Say that $\mathcal{P}_n$ starts with $E^{(n)}$ ebits. Then its entanglement consumption is no greater than $\max_{a,b}\left[E^{(n)} - E(\mathcal{P}_n|a\rangle_A|b\rangle_B|\Phi\rangle_{AB}^{E^{(n)}})\right] \leq nE(U^\dagger)$ (using $\Delta E_{U^\dagger} = E(U^\dagger)$ from Theorem 2.6). Here $E(U^\dagger)$ can be thought of as an *entanglement destroying capacity* of U if we recognize that unitarily disentangling a state is a nonlocal task. For $U \in \mathcal{U}_{2\times2}$, we always have

$E(U) = E(U^\dagger)$, but for $d > 2$, numerical evidence suggests that equality no longer holds[CLS02]. Since $\mathcal{P}_n$ uses no more than $nE(U^\dagger)$ ebits, we have $\langle U\rangle + E(U^\dagger)[qq] \geq C_+^E(U)[qq]$ and thus $E(U) \geq C_+^E(U) - E(U^\dagger)$, implying the desired result. More generally, for any $(C_1, C_2, E) \in \text{CCE}(U)$ this result implies that $(C_1, C_2, -E(U^\dagger)) \in \text{CCE}(U)$; i.e. more than $E(U^\dagger)$ ebits are never needed for any communication protocol.

The argument outlined above follows the presentation of [BS03b]. However, we also need to address the second problem introduced by free entanglement. For the inefficiency caused by communication errors to vanish as in Proposition 2.10, we need to ensure that the logs of the Schmidt numbers of the states we work with grow at most linearly with n. Equivalently, we need to show that the parameter $E^{(n)}$ from the previous paragraph can be chosen to be $\leq Kn$ for some constant K. In [BHLS03], the explicit construction of Theorem 2.9 was used to achieve this bound for one-way communication, and thereby to prove the weaker result that $C_{\rightarrow}^E(U) \leq E(U) + E(U^\dagger)$.

Finally [BS03a] proves an exponential bound on Schmidt rank for general bidirectional protocols, by applying HSW coding in both directions to $\mathcal{P}_n$. Specifically, for any input of Bob's, Alice can consider $\mathcal{P}_n$ to be a channel that communicates $n(C_1 - \delta_n)$ bits with error $\leq \epsilon_n$; such a channel has HSW capacity $\approx n(C_1 - \delta_n)(1 - \epsilon_n) = nC_1 - o(n)$. Similarly, Bob can code for a channel to Alice that has capacity $nC_2 - o(n)$. These block codes require k blocks of $\mathcal{P}_n$ with $k \gg \exp(n)$, but now the total error goes to zero as $k \to \infty$, while the entanglement cost $kE^{(n)}$ grows linearly with k. So the desired capacity is achieved by taking $k \to \infty$ before n. A refined version of this argument will be presented in the proof of Theorem 3.1 in Section 3.5.

Technically, HSW coding is not quite appropriate here, since Alice's channel weakly depends on Bob's input and vice versa. Thus, a small modification of [BS03a]'s proof is necessary. The correct coding theorem to use for bidirectional channels was given in 1961 by Shannon[Sha61] and can be used to obtain the result claimed in [BS03a] (see also [CLL05] for a generalization of Shannon's 1961 result to noisy bidirectional quantum channels). Unlike the HSW theorem and Shannon's original noisy channel coding theorem[Sha48], the two-way coding theorem only achieves low average error instead of low maximum error. For entanglement generation, average error is sufficient, but in the next chapter we will show (in Theorem 3.1) that maximum error can also be made small for bidirectional protocols. In fact, the average error and maximum error conditions appear to be asymptotically equivalent in general, given some mild assumptions[DW05b, CK81].

2.3.4 Challenges for bidirectional communication

We conclude our discussion of classical communication using unitary gates in this section, by reviewing attempts to extend Theorem 2.6 to the case of bidirectional communication and pointing out the difficulties that arise.

There is no bidirectional analogue of HSW coding, even classically. In [Sha61], Shannon considers communication with noisy bidirectional channels—a model in some ways simpler, but in other ways more complex, than unitary gates—and establishes upper and lower bounds that do not always coincide. We briefly restate those bounds here. Define a bidirectional channel $N(A_{\text{out}}B_{\text{out}}|A_{\text{in}}B_{\text{in}})$ where A_{in} is Alice's input, B_{in} is Bob's input, A_{out} is Alice's output and B_{out} is Bob's output. For any probability distribution on the inputs $A_{\text{in}}B_{\text{in}}$, consider the rate pair $I(A_{\text{in}}; B_{\text{out}}|B_{\text{in}})[c \to c] + I(B_{\text{in}}; A_{\text{out}}|A_{\text{in}})[c \leftarrow c]$. [Sha61] proves that this rate pair is

- achievable if we maximize over *product* distributions on $A_{\text{in}}B_{\text{in}}$ (i.e. $I(A_{\text{in}}; B_{\text{in}}) = 0$) ; and
- an upper bound if we maximize over arbitrary distributions on $A_{\text{in}}B_{\text{in}}$ (i.e. if $\langle N\rangle \geq C_1[c \to c] + C_2[c \leftarrow c]$, then there exists a joint distribution on $A_{\text{in}}B_{\text{in}}$ such that $C_1 = I(A_{\text{in}}; B_{\text{out}}|B_{\text{in}})$ and $C_2 = I(B_{\text{in}}; A_{\text{out}}|A_{\text{in}})$).

Using the chain rule[CK81] we can rewrite these quantities suggestively as $C_1 = I(A_{\text{in}}; B_{\text{in}}B_{\text{out}}) - I(A_{\text{in}}; B_{\text{in}})$ and $C_2 = I(B_{\text{in}}; A_{\text{in}}A_{\text{out}}) - I(A_{\text{in}}; B_{\text{in}})$. In this form, they resemble Eq. (2.20): for communication from Alice to Bob we measure the difference between the output correlation $I(A_{\text{in}}; B_{\text{in}}B_{\text{out}})$

and the input correlation $I(A_{\rm in}; B_{\rm in})$ and a similar expression holds for communication from Bob to Alice. This has led [BS03a] to conjecture that a bidirectional version of $\Delta\chi_U$ (defined in Eq. (2.20)) should describe the two-way classical capacity of a unitary gate. However, even in the classical case, Shannon's inner and outer bounds on the capacity region (corresponding to uncorrelated or correlated inputs respectively) are in general different.

This highlights the difficulties in coding for bidirectional channels. The messages both parties send may interfere with each other, either positively or negatively. The best known protocols reduce the bidirectional channel to a pair of one-way channels for which Alice and Bob code independently. However, we cannot rule out the case in which Alice and Bob use correlated channel inputs to improve the rate.

The same general concerns apply to quantum bidirectional channels, including unitary gates, although not all of the corresponding bounds have been proven. Some promising steps towards this goal are in [YDH05, Yar05], which derive capacity expressions for quantum channels with two inputs and one output.

Reversible RSP is not possible for all bidirectional ensembles. The crucial ingredient in the proof of Theorem 2.9 was the equivalence for any ensemble $\mathcal{E}$ (given unlimited entanglement) between the induced $\{c \to q\}$ map $\mathcal{N}_\mathcal{E}$ and the standard resource $I(X_A; B)_\mathcal{E}[c \to c]$. Now suppose $\mathcal{E}$ is a bidirectional ensemble $\sum_{i,j} p_i q_j \, |i\rangle\langle i|^{X_A} \otimes |j\rangle\langle j|^{Y_B} \otimes |\psi_{ij}\rangle^{AB}\}$ which has a corresponding $\{cc \to qq\}$ channel $\mathcal{N}_\mathcal{E}$ mapping $|i\rangle^A|j\rangle^B$ to $|\psi_{ij}\rangle^{AB}$. To extend Theorem 2.9 to the bidirectional case, we would begin by trying to find pairs (C_1, C_2) such that $\langle\mathcal{N}_\mathcal{E}\rangle + \infty[qq] = C_1[c \to c] + C_2[c \leftarrow c] + \infty[qq]$. It turns out that there are ensembles for which no such equivalence exists. In fact, classical communication cannot reversibly simulate any ensemble whose classical capacity region is not just a rectangle. The proof of this is trivial: if $\langle\mathcal{N}_\mathcal{E}\rangle + \infty[qq] = C_1[c \to c] + C_2[c \leftarrow c] + \infty[qq]$ then $\langle\mathcal{N}_\mathcal{E}\rangle + \infty[qq] \geq R_1[c \to c] + R_2[c \leftarrow c]$ if and only if $R_1 \leq C_1$ and $R_2 \leq C_2$.

One simple example of an ensemble that cannot be reversibly simulated is the ensemble corresponding to the AND channel: $|\psi_{ij}\rangle^{AB} = |i \wedge j\rangle^A |i \wedge j\rangle^B$, where $i, j \in \{0, 1\}$ and $i \wedge j$ is the logical AND operation. Clearly $(1, 0) \in \mathrm{CC(AND)}$ and $(0, 1) \in \mathrm{CC(AND)}$; i.e. the AND ensemble can send one bit from Alice to Bob or one bit from Bob to Alice. (The channel is effectively classical, so we need not consider entanglement) If AND were reversibly simulatable, then we would expect $(1, 1) \in \mathrm{CC(AND)}$. However, $(1, \epsilon) \notin \mathrm{CC(AND)}$ for any $\epsilon > 0$. Suppose Bob sends zero with probability p and one with probability $1 - p$. When Bob sends zero, the channel output is $|00\rangle$ regardless of Alice's input. Alice can only communicate to Bob during the $1 - p$ fraction of time that he sends one, so she can only send $1 - p$ bits to him. Thus we must have $p = 0$. Since Bob always sends one, he cannot communicate any information to Alice.

One might object to the AND example by pointing out that simulating a relative resource is a more reasonable goal, since the capacities of ensembles like AND vary with the probability distribution of Alice and Bob's inputs. In fact, even in the one-way case the HSW/RSP equivalence in Eq. (2.19) is only proven for relative resources.* However, one can construct ensembles where reversible simulation is impossible even if the probability distribution of the input is fixed. We construct one such ensemble (or channel) as follows:

Alice and Bob both input $m + 1$ bit messages, (a_1, a_2) and (b_1, b_2), where a_1 and b_1 are single bits and a_2 and b_2 are m-bit strings. The channel $\mathcal{N}$ computes the following string: $(a_1 \oplus b_1, (a_1 \oplus b_1 ? a_2 : b_2))$ and gives Alice and Bob both a copy of it. The notation $(a_1 \oplus b_1 ? a_2 : b_2)$ means that the channel outputs a_2 if $a_1 \oplus b_1 = 1$ and b_2 if $a_1 \oplus b_1 = 0$. We choose the input probability distributions to be uniform for both parties. Alice and Bob are allowed to agree upon any sort of block coding protocol they wish as long as they still send each input approximately the same number of times.

First, we argue that $(m, 0), (0, m) \in \mathrm{CC}(\mathcal{N})$. The protocol to achieve $(m, 0)$ is as follows: Alice sets $a_1 = 0$ for the first $n/2$ rounds and $a_1 = 1$ for the last $n/2$ rounds. Likewise, Bob sets $b_1 = 0$

*Actually, the quantum reverse Shannon theorem[BDH$^+$05] gives a reversible simulation of *unrelativized* $\{c \to q\}$ channels, though this appears not to be possible for general $\{q \to q\}$ channels, or for the coherent version of $\{c \to q\}$ channels that we will consider in the next chapter.

for the first $n/2$ rounds and $b_2 = 1$ for the last $n/2$ rounds. The other two registers are set uniformly at random. This satisfies the criteria of p_i and q_j being uniform, although it is a very particular coding scheme. Since $a_1 \oplus b_1$ is always zero, it is always Alice's message a_2 which is broadcast to both parties. Thus, this transmits m bits to Bob per use of $\mathcal{N}$. If Bob instead sets $b_1 = 1$ for the first $n/2$ rounds and $b_2 = 0$ for the last $n/2$ rounds, then the communication direction is reversed.

If $\mathcal{N}$ with uniformly distributed inputs had a rectangular rate region, then (m, m) would also be achievable. However any achievable (C_1, C_2) must satisfy $C_1 + C_2 \leq m + 2$, since there is a natural multi-round simulation for $\mathcal{N}$ that uses $m + 2$ total cbits. Choosing $m > 2$ yields a non-rectangular rate region and hence a channel that cannot be efficiently simulated, even with a fixed input probability distribution.

Arguably, even this example does not go far enough, since we could talk about simulating $\mathcal{N}$ with respect to a bipartite test state ρ^{AB}. However, it is hard to define a corresponding asymptotic resource; the natural choice of $\langle \mathcal{N} : \rho^{AB} \rangle = (\mathcal{N}^{\otimes n} : \rho^{\otimes n})_{n=1}^{\infty}$ violates Eq. (1.11) since extra input test states ρ^{AB} can no longer be created for free locally. On the other hand, $\langle \mathcal{N} : \rho_1^A \otimes \rho_2^B \rangle$ is a well-defined resource for which there may be a reversible simulation, but since it cannot contain any correlations between Alice and Bob it is hard to imagine using it in a protocol analogous to the one in Theorem 2.9.

Combined, these facts mean that we are likely to need new methods and possibly new ways of thinking about resources to find the two-way capacity regions of unitary gates.

2.4 Examples

There are only a handful of examples of unitary gates where any capacities can be computed exactly. On the other hand, some more complicated gates appear to give separations between quantities like $C_{\rightarrow}$ and $C_{\leftarrow}$ or C_{+} and E, though we will only be able to offer incomplete proofs for these claims. This section will describe what is known about the capacities of all of these examples. Many of the results on the two-qubit gates SWAP, CNOT and DCNOT are taken from [CLP01].

2.4.1 SWAP, CNOT and double CNOT

We begin by reviewing three well-known gates in $\mathcal{U}_{2\times 2}$.

SWAP: The SWAP gate on two qubits is in a sense the strongest two qubit gate; i.e. for any two-qubit U, $\langle \textsc{swap} \rangle = [q \to q] + [q \leftarrow q] \geq \langle U \rangle$. The proof follows the lines of Proposition 2.8: any U can be simulated by sending Alice's input to Bob using $[q \to q]$, Bob performing U locally, and then Bob sending Alice's qubit back with $[q \leftarrow q]$. Thus, we would expect it to saturate all of the upper bounds we have found on capacities.

In fact the capacity region is

$$\mathrm{CCE}(\textsc{swap}) = \{(C_1, C_2, E) : C_1 \leq 2, C_2 \leq 2, E \leq 2, \max(C_1, 0) + \max(C_2, 0) + E \leq 2\}. \tag{2.26}$$

The first two upper bounds follow from Proposition 2.8 and the last two upper bounds from

$$\max(C_1, 0) + \max(C_2, 0) + E \leq E(\textsc{swap}) \leq \log \mathrm{Sch}(\textsc{swap}) = 2.$$

To show that this entire region is achievable, we can apply Proposition 2.10 to the single point $(2, 2, -2) \in \mathrm{CCE}(\textsc{swap})$. This in turn follows from applying super-dense coding in both directions to obtain $\textsc{swap} + 2[qq] = [q \to q] + [qq] + [q \leftarrow q] + [qq] \geq 2[c \to c] + 2[c \leftarrow c]$.

There are more direct proofs for some of the other extreme points of the capacity region; the interested reader should try the exercise of finding a simple alternate proof of $\textsc{swap} \geq 2[c \to c]$ (see Eq. (63) of [CLP01] for the answer).

CNOT: It turns out that the capacity region of CNOT is exactly one half the size of the capacity region for SWAP: $\mathrm{CCE}(\textsc{cnot}) = \{(C_1, C_2, E) : C_1 \leq 1, C_2 \leq 1, E \leq 1, \max(C_1, 0) + \max(C_2, 0) + E \leq$

1}. (We will later see that this is no accident but rather a consequence of the asymptotic equivalence $2\langle\textsc{cnot}\rangle = \langle\textsc{swap}\rangle$.)

The first three upper bounds follow from a simulation due to Gottesman[Got99],

$$[c \to c] + [c \leftarrow c] + [qq] \geq \langle\textsc{cnot}\rangle \tag{2.27}$$

and causality. Then applying Proposition 2.10 yields the last bound.

In terms of achievability, $\langle\textsc{cnot}\rangle \geq [c \to c]$ is obvious, and $\langle\textsc{cnot}\rangle \geq [c \leftarrow c]$ follows from $(H \otimes H)\textsc{cnot}(H \otimes H)|0\rangle|b\rangle = \textsc{swap}\,\textsc{cnot}\,\textsc{swap}|0\rangle|b\rangle = |b\rangle|b\rangle$. However, just as the entire SWAP capacity region follows from $(2, 2, -2)$, the entire CNOT region follows from the inequality $\textsc{cnot} + [qq] \geq [c \to c] + [c \leftarrow c]$, which is achieved by a protocol due to [CLP01]:

$$(Z^a H \otimes I)\textsc{cnot}(X^a \otimes Z^b)|\Phi_2\rangle_{AB} = |b\rangle_A|a\rangle_B \tag{2.28}$$

Double CNOT: The double CNOT is formed by applying two CNOTs consecutively: first one with Alice's qubit as control and Bob's as target, and then one with Bob's qubit as control and Alice's as target. Equivalently we can write $\textsc{dcnot} = \textsc{swap}\,\textsc{cnot}\,\textsc{swap}\,\textsc{cnot}$. For $a, b \in \{0, 1\}$, we have $\textsc{dcnot}|a\rangle|b\rangle = |b\rangle|a \oplus b\rangle$.

The double CNOT seems weaker than two uses of a CNOT, but it turns out to have the same capacity region as the SWAP gate, or as $(\textsc{cnot})^{\times 2}$:

$$\text{CCE}(\textsc{dcnot}) = \text{CCE}(\textsc{swap}) = \{(C_1, C_2, E) : C_1 \leq 2, C_2 \leq 2, E \leq 2, \max(C_1, 0) + \max(C_2, 0) + E \leq 2\}. \tag{2.29}$$

The upper bounds are the same as for SWAP, and achievability is shown in [CLP01]. Specifically, they give a protocol for the point $(2, 2, -2) \in \text{CCE}(\textsc{dcnot})$, from which all other points follow.

Relations among SWAP, CNOT and Double CNOT: If we were to judge the strengths of the SWAP, CNOT and DCNOT gates solely based on their capacity regions, then it would be reasonable to conclude that

$$\langle\textsc{swap}\rangle = 2\langle\textsc{cnot}\rangle = \langle\textsc{dcnot}\rangle. \tag{2.30}$$

However, it has been historically difficult to construct efficient maps between these gates. [CLP01] has conjectured that $2\textsc{cnot} \not\geq \textsc{swap}$, and since $2\textsc{cnot} \geq \textsc{dcnot}$, this would imply that $\textsc{dcnot} \not\geq \textsc{swap}$. Moreover, [HVC02] shows that DCNOT is takes less time than SWAP to simulate using nonlocal Hamiltonians, implying that it somehow has less nonlocal power. A cute side effect of coherent classical communication, which we will introduce in the next chapter, will be a concise proof of Eq. (2.30), confirming the intuition obtained from capacity regions.

Of course, this simple state of affairs appears to be the exception rather than the rule. We now consider two examples of gates whose capacity regions appear to be less well behaved.

2.4.2 A gate for which $C_{\leftarrow}(U)$ may be less than $C_{\to}(U)$

In this section we introduce a gate $U_{\text{XOXO}} \in \mathcal{U}_{d\times d}$ that appears to have $C_{\leftarrow}(U) < C_{\to}(U)$ when d is sufficiently large. Define U_{XOXO} as follows:

$$\begin{aligned}
U_{\text{XOXO}}|x0\rangle &= |xx\rangle & \forall 0 \leq x < d \\
U_{\text{XOXO}}|xx\rangle &= |x0\rangle & \forall 0 \leq x < d \\
U_{\text{XOXO}}|xy\rangle &= |xy\rangle & \forall x \neq y \neq 0
\end{aligned}$$

The first two lines are responsible for the gate's affectionate nickname, "XOXO." The $d = 2$ case corresponds to a CNOT, which is locally equivalent to a symmetric gate, though as d increases U_{XOXO} appears to be quite asymmetric.

Bounds on capacities for $U_{\mathbf{XOXO}}$

If Alice inputs $|a\rangle$ and Bob inputs $|0\rangle$, then Bob will obtain a copy of Alice's input a. Thus $C_{\rightarrow}(U_{\text{XOXO}}) \geq \log d$.

Define $S_x \in \mathcal{L}(\mathcal{H}_B)$ by

$$S_x|y\rangle = \begin{cases} |0\rangle & \text{if } x = y \\ |x\rangle & \text{if } 0 = y \\ |y\rangle & \text{otherwise} \end{cases}$$

Then $U_{\text{XOXO}} = \sum_x |x\rangle\langle x| \otimes S_x$, so $\text{Sch}(U_{\text{XOXO}}) \leq d$. Thus $E(U_{\text{XOXO}}) \leq \log d$. Combining this with $C_{\rightarrow}(U_{\text{XOXO}}) \geq \log d$ yields $\log d \leq C_{\rightarrow}(U_{\text{XOXO}}) \leq C_{+}(U_{\text{XOXO}}) \leq E(U_{\text{XOXO}}) \leq \log \text{Sch}(U_{\text{XOXO}}) \leq \log d$. Thus these must all be equalities, and we have

$$C_{\rightarrow}(U_{\text{XOXO}}) = C_{+}(U_{\text{XOXO}}) = E(U_{\text{XOXO}}) = \log \text{Sch}(U_{\text{XOXO}}) = \log d$$

These are the only capacities that know how to determine exactly. However, we can bound a few other capacities.

Suppose Alice and Bob share a d-dimensional maximally entangled state $|\Phi_d\rangle = \frac{1}{\sqrt{d}}\sum_x |x\rangle|x\rangle$. Using such a state Bob can communicate $\log d$ bits to Alice. The protocol is as follows. Let $b \in \{0, \ldots, d-1\}$ be the message Bob wants to send and let $\omega = \exp(2\pi i/d)$. First Bob applies the unitary transformation $\sum_x \omega^{bx}|x\rangle\langle x|$ to his half of $|\Phi_d\rangle$, leaving them with the state $\frac{1}{\sqrt{d}}\sum_x \omega^{bx}|x\rangle|x\rangle$. Then they apply the gate U_{XOXO} to obtain the product state $\frac{1}{\sqrt{d}}\sum_x \omega^{bx}|x\rangle|0\rangle$. Alice can now apply the inverse Fourier transform $\frac{1}{\sqrt{d}}\sum_{xy}|x\rangle\langle y|\omega^{-xy}$ to recover Bob's message.

Thus $C^E_{\leftarrow}(U_{\text{XOXO}}) \geq \log d$. This yields a lower bound for $C_{\leftarrow}(U_{\text{XOXO}})$ as well, since one possible communication strategy for Bob is to use U_{XOXO} once to create a copy of $|\Phi_d\rangle$ and a second time to send $\log d$ bits to Alice, using up the copy of $|\Phi_d\rangle$.

So $\frac{1}{2}\log d \leq C_{\leftarrow}(U_{\text{XOXO}}) \leq \log d$. We would like to know whether $C_{\leftarrow}(U_{\text{XOXO}}) < C_{\rightarrow}(U_{\text{XOXO}}) = \log d$. We cannot prove this expression asymptotically, but can show that if Alice and Bob share no entanglement and are initially uncorrelated, Alice's mutual information with Bob's message is strictly less than $\log d$ after a single use of U_{XOXO}.

Bounding the one-shot rate of $U_{\mathbf{XOXO}}$

Proposition 2.11. *If Alice and Bob share no entanglement and input uncorrelated states into U_{XOXO}, Alice's mutual information with Bob's message is less than $(1-\epsilon)\log d + \mathcal{O}(1)$ for some constant $\epsilon > 0$.*

Proof. Let α, β, γ be small positive parameters that we will choose later.

Assume Alice begins with fixed input $|\psi^A\rangle^A = \sum_i a_i|i\rangle^{A_1}\sum_j A_{ij}|j\rangle^{A_2}$ where $\sum_i |a_i|^2 = \sum_j |A_{ij}|^2 = 1$ and A denotes the composite Hilbert space A_1A_2. Let $R \subseteq \{0, \ldots, d-1\}$ be the set given by

$$R = \left\{i : |a_i|^2 > \alpha\right\}.$$

The normalization condition means that $|R| \leq 1/\alpha$.

Bob will signal to Alice with some ensemble $\mathcal{E} = \sum_x p_x \left|x\middle\rangle\middle\langle x\right|^{X_B} \otimes \left|\psi^B_x\middle\rangle\middle\langle\psi^B_x\right|^B$. We will divide the indices x into three sets S_1, S_2 and S_3, according to various properties of the states $|\psi^B_x\rangle$. Write one such state as $\sum_i b^{(x)}_i|i\rangle^{B_1}\sum_j B^{(x)}_{ij}|j\rangle_{B^2}$, where again B denotes the composite Hilbert space B_1B_2,

U_{XOXO} acts on A_1B_1 and A_2B_2 are ancilla systems. Now define S_1, S_2 and S_3 by

$$S_1 = \left\{x : |b_0^{(x)}|^2 \geq \beta\right\} \tag{2.31}$$

$$S_2 = \left\{x : |b_0^{(x)}|^2 < \beta \text{ and } \sum_{i\in R} |b_i^{(x)}|^2 \geq \gamma\right\} \tag{2.32}$$

$$S_3 = \left\{x : |b_0^{(x)}|^2 < \beta \text{ and } \sum_{i\in R} |b_i^{(x)}|^2 < \gamma\right\} \tag{2.33}$$

Without loss of generality, we can introduce a second classical register for Bob, Y_B, that records which of the S_y the index x belongs to. If we also include Alice's fixed input state, then $\mathcal{E}$ becomes

$$\mathcal{E} = \sum_{y\in\{1,2,3\}} |y\rangle\langle y|^{Y_B} \otimes \sum_{x\in S_y} |x\rangle\langle x|^{X_B} \otimes |\psi_x\rangle\langle\psi_x|^{AB}, \tag{2.34}$$

where $|\psi_x\rangle := |\psi^A\rangle|\psi_x^B\rangle$.

After $U := U_{\text{XOXO}}$ is applied, the parties are left with the ensemble $U(\mathcal{E}) := (U^{A_1B_1} \otimes I^{X_BY_BA_2B_2})(\mathcal{E})$. The mutual information of Alice's state with Bob's message is given by

$$\begin{aligned} I(X_B; A)_{U(\mathcal{E})} &= I(X_BY_B; A)_{U(\mathcal{E})} = I(X_B; A|Y_B)_{U(\mathcal{E})} + I(A; Y_B)_{U(\mathcal{E})} \\ &\leq I(X_B; A|Y_B)_{U(\mathcal{E})} + \log 3 \leq \max_{y\in\{1,2,3\}} I(X_B; A)_{U(\mathcal{E}_y)} + \log 3. \end{aligned} \tag{2.35}$$

Here we have defined the ensemble $\mathcal{E}_y$, for $y \in \{1,2,3\}$ to be the ensemble $\mathcal{E}$ conditioned on $Y_B = y$; i.e.

$$\mathcal{E}_y := \left(\sum_{x\in S_y} p_x\right)^{-1} \sum_{x\in S_y} p_x\, |x\rangle\langle x|^{X_B} \otimes |y\rangle\langle y|^{Y_B} \otimes |\psi_x\rangle\langle\psi_x|^{AB}. \tag{2.36}$$

Thus to prove our proposition it suffices to verify that $I(X_B; A)_{U(\mathcal{E}_y)} < (1-\epsilon)\log d + \mathcal{O}(1)$ for each choice of y.

For cases $y = 1, 2$ we will use the following two facts.

Fact 2.12. *Let ρ be a d-dimensional state and suppose that $\operatorname{Tr}\Pi\rho = p$ for some k-dimensional projector Π. Then measuring $\{\Pi, I - \Pi\}$ yields a state with entropy no greater than*

$$-k\frac{p}{k}\log\frac{p}{k} - (d-k)\frac{1-p}{d-k}\log\frac{1-p}{d-k} = H_2(p) + p\log k + (1-p)\log(d-k) < 1 + \log k + (1-p)\log d. \tag{2.37}$$

Since $H(\rho) \leq H(\Pi\rho\Pi + (1-\Pi)\rho(1-\Pi))$ it follows that $H(\rho) \leq (1-p)\log d + 1 + \log k$. If we treat k as a constant, then this is $(1-p)\log d + \mathcal{O}(1)$.

Fact 2.13. *The mutual information of the output is bounded by entropy of Bob's input as follows:*

$$I(X_B; A)_{U(\mathcal{E}_y)} \leq I(X_B; AB_1)_{U(\mathcal{E}_y)} = I(X_B; AB_1)_{\mathcal{E}_y} \leq H(AB_1)_{\mathcal{E}_y} = H(B_1)_{\mathcal{E}_y}. \tag{2.38}$$

We can now prove that $I(X_B; A)_{U(\mathcal{E}_1)} < (1-\epsilon)\log d + \mathcal{O}(1)$. By the definition of S_1, we have $\langle 0|\mathcal{E}_1^{B_1}|0\rangle \geq \beta$. Now we use first Fact 2.13 and then Fact 2.12 (with the projector $\Pi = |0\rangle\langle 0|^{B_1}$) to obtain

$$I(X_B; A)_{U(\mathcal{E}_1)} \leq H(B_1)_{\mathcal{E}_1} < (1-\beta)\log d + 1. \tag{2.39}$$

This last expression is $\leq (1-\epsilon)\log d + 1$ as long as $\epsilon \leq \beta$.

The case of $y = 2$ will yield to similar analysis. Define $|i'\rangle \in \mathcal{H}_B$ by $|i'\rangle^B = |i\rangle^{B_1} \otimes \sum_j B_{ij}|j\rangle^{B_2}$. Now define a projector $\Pi = \sum_{i\in R} |i'\rangle\langle i'|^B$ so that $\operatorname{Tr}\Pi = |R|$ and $p := \operatorname{Tr}\langle\psi_x^B|\Pi|\psi_x^B\rangle = \sum_{i\in R} |b_i|^2$.

Note that $\operatorname{Tr}\Pi \leq 1/\alpha$ and $p \geq \gamma$. Now we can again use Facts 2.13 and 2.12 to bound $I(X_B; A)_{U(\mathcal{E}_2)} < 1 + \log 1/\alpha + (1-\gamma)\log d$. This is $\leq (1-\epsilon)\log d + \mathcal{O}(1)$ if we choose $\epsilon \leq \gamma$.

Note that these two bounds are independent of U. They simply say that having a lot of weight in a small number of dimensions limits the potential for communication. Case S_3 is the interesting case. Here we will argue that if Bob inputs a state that is not zero and does not match Alice's state well, he will not change Alice's state very much.

Suppose a particular input state can be expressed as

$$|\psi\rangle = \sum_{i,j,k,l} a_i b_j A_{ik} B_{il} |ijkl\rangle^{A_1 A_2 B_1 B_2}.$$

According to the definition of S_3, $|b_0|^2 < \beta$ and $\sum_{i\in R} |b_i|^2 < \gamma$, where $R = \{i : |a_i|^2 \geq \alpha\}$.

After one use of the nonlocal gate U, the new state is

$$|\psi'\rangle := U|\psi\rangle = |\psi\rangle + \sum_{i\neq 0}\sum_{k,l} a_i b_i A_{ik} B_{il} |i0kl\rangle + a_i b_0 A_{ik} B_{0l} |iikl\rangle - a_i b_i A_{ik} B_{il} |iikl\rangle - a_i b_0 A_{ik} B_{0k} |i0kl\rangle$$

Writing $|\psi'\rangle$ in this form is useful for bounding the state change

$$\begin{aligned}
\left\| |\psi'\rangle - |\psi\rangle \right\|^2 &= \sum_{i,k,l} |a_i|^2 |A_{ik}|^2 \left(|b_i B_{il} - b_0 B_{0l}|^2 + |b_0 B_{0l} - b_i B_{il}|^2 \right) \\
&= 2 \sum_{i,l} |a_i|^2 |b_i B_{il} - b_0 B_{0l}|^2 \\
&\leq 4 \sum_{i,l} |a_i|^2 \left(|b_i|^2 |B_{il}|^2 + |b_0|^2 |B_{0l}|^2 \right) \\
&= 4 \sum_{i} |a_i|^2 \left(|b_i|^2 + |b_0|^2 \right) \\
&= 4 \sum_{i\in R} |a_i b_i|^2 + 4 \sum_{i\notin R} |a_i b_i|^2 + 4|b_0|^2
\end{aligned} \tag{2.40}$$

where the inequality on the third line follows from the general bound $|x-y|^2 \leq (|x|+|y|)^2 = 2(x^2+y^2) - (|x|-|y|)^2 \leq 2(x^2+y^2)$.

We can bound each of the three terms in Eq. (2.40) separately. First,

$$\sum_{i\in R} |a_i b_i|^2 \leq \sum_{i\in R} |b_i|^2 < \gamma$$

The second term is

$$\sum_{i\notin R} |a_i b_i|^2 \leq \sum_{i\notin R} \left(\sum_{j\notin R} |a_j|^2 \right) |b_i|^2 < \alpha \sum_{j\notin R} |b_j|^2 \leq \alpha \sum_j |b_j|^2 = \alpha$$

The third term is simply $|b_0|^2 < \beta$.

Thus $\left\| |\psi'\rangle - |\psi\rangle \right\|^2 < 4(\alpha+\beta+\gamma)$. In terms of fidelity, $F(|\psi\rangle, |\psi'\rangle) = |\langle\psi|\psi'\rangle|^2 > 1 - 4(\alpha+\beta+\gamma)$. Converting this to trace distance means that $\frac{1}{2}\| \, |\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'| \, \|_1 < 2\sqrt{(\alpha+\beta+\gamma)}$. Since this holds for each element of $\mathcal{E}_3$ and trace distance is convex it follows that $\frac{1}{2}\|\mathcal{E}_3^A - U(\mathcal{E}_3)^A\|_1 < 2\sqrt{(\alpha+\beta+\gamma)}$. Alice's system is initially in a pure state, so we can do a Schmidt decomposition between A and A' and thus assume that $\dim A' = d$. This also means that $H(\mathcal{E}_3^A) = 0$. Using Fannes' inequality then yields $I(X_B; A)_{U(\mathcal{E}_3)} \leq H(A)_{U(\mathcal{E}_3)} < 4\sqrt{(\alpha+\beta+\gamma)} \cdot 2\log d + (\log e)/e \leq (1-\epsilon)\log d + \mathcal{O}(1)$ as long as $\epsilon \leq 1 - 8\sqrt{(\alpha+\beta+\gamma)}$.

This proves our claim for any $\alpha, \beta, \gamma > 0$ as long as $\epsilon \leq \beta$, $\epsilon \leq \gamma$ and $\epsilon \leq 1 - 8\sqrt{(\alpha+\beta+\gamma)}$. This clearly holds as long as α, β, γ and ϵ are small enough. The largest value of ϵ possible is $\sqrt{\sqrt{33}-\sqrt{32}} \approx 0.2962$, when $\alpha \approx 0$ and $\beta = \gamma = \epsilon$. □

I suspect that the actual asymptotic capacity is closer to $\frac{1}{2}\log d + \mathcal{O}(1)$ for large values of d, but more careful techniques will be required to prove this.

2.4.3 A gate for which $C_+(U)$ may be less than $E(U)$

Another separation that appears plausible is between the total classical capacity $C_+(U)$ and the entanglement capacity $E(U)$. In this section we present an example of a gate U for which it appears that $C_+(U) < E(U)$, though, as with the last section, we cannot actually prove this claim.

The gate is defined (for any d) as follows: $U = I + |\Phi_d\rangle\langle 01| + |01\rangle\langle\Phi_d| - |01\rangle\langle 01| - |\Phi_d\rangle\langle\Phi_d|$. Obviously, $E(U) \geq \log d$, since $U|01\rangle = |\Phi_d\rangle$. This inequality is not quite tight (i.e. $E(U) > \log d$ and probably $E(U) \approx \log d + O(1)$), but this doesn't matter for the argument.

I conjecture that $C_+^E(U) = O(1) < \log d$ for large d. However, the only statement that can readily be proven is that, like the last section, a single use of U for one-way communication on uncorrelated product inputs can create strictly less than $\log d$ bits of mutual information, for d sufficiently large.

The proof is actually almost identical to the proof of the last section, though slightly simpler. If Alice and Bob input product states, then the overlap of their states with $|\Phi_d\rangle$ is $\leq 1/\sqrt{d}$, so this portion of U has little effect. We divide Alice's signal ensemble into a part with a large $|0\rangle$ component (which has low entropy) and a part with a small $|0\rangle$ component (which is nearly unchanged by the action of U). As a result, the total amount of information that Alice can send to Bob (or that Bob can send to Alice) with one use of U, starting from uncorrelated product states, is strictly less than the entanglement capacity. However, this argument is far from strong enough to prove a separation between asymptotic capacities.

2.5 Discussion

We conclude this chapter by restating its key results and discussing some of the major open questions. Most of the gate capacities can be expressed in terms of the three-dimensional region $\mathrm{CCE}(U) := \{(C_1, C_2, E) : \langle U\rangle \geq C_1[c \to c] + C_2[c \leftarrow c] + E[qq]\}$. The two coding theorems (2.6 and 2.9) establish that

- $\max\{E : \langle U\rangle \geq E[qq]\} =: E(U) = \Delta E_U := \sup_\psi H(B)_{U(\psi)} - H(B)_\psi$
- $\max\{C : \langle U\rangle + \infty[qq] \geq C[c \to c]\} =: C_\to^E(U) = \Delta\chi_U := \sup_{\mathcal{E}} \chi(\mathrm{Tr}_A\, U(\mathcal{E})) - \chi(\mathrm{Tr}_A\, \mathcal{E})$

The key bounds (from Propositions 2.4, 2.5, 2.8 and 2.10) are

- $C_+(U) \leq E(U) \leq \log \mathrm{Sch}(U) \leq 2\log d$
- $C_+^E(U) \leq \min(4\log d, E(U) + E(U^\dagger))$
- $C_\to^E(U), C_\leftarrow^E(U) \leq 2\log d$
- $E(U) \geq H(\lambda)[qq]$ where $\{\lambda_i\}$ are Schmidt coefficients of U.
- $C_\to(U) \neq 0 \iff C_\leftarrow(U) \neq 0 \iff E(U) \neq 0 \iff \mathrm{Sch}(U) \neq 1 \iff U$ is nonlocal.

These results suggest a number of open questions.

- Can we find an upper bound on the dimension of the ancillas $A'B'$ that are needed for an optimal input state for entanglement generation? For entanglement-assisted classical communication, how large do the dimensions of $A'B'$ need to be, and how many states are needed in the optimal ensemble? These are important for numerical studies of the capacities.

- Do there exist U such that $C_{\rightarrow}(U) \neq C_{\leftarrow}(U)$? Note that U cannot be a two-qubit gate since the decomposition in Eq. (2.1) implies that two-qubit gates have symmetric capacities. I conjecture that $C_{\rightarrow}(U_{\text{XOXO}}) \neq C_{\leftarrow}(U_{\text{XOXO}})$ for U_{XOXO} defined as in Section 2.4.2.

- Do there exist U such that $C^E_{\rightarrow}(U) \neq C^E_{\leftarrow}(U)$? All of the examples of gates in Section 2.4 satisfy $U = U^\dagger$, but unpublished work with Peter Shor proves that in this case the entanglement-assisted capacity regions are fully symmetric. It seems plausible that this situation would hold in general, but no proof or counterexample is known.

- Do there exist U for which $C_+(U) < E(U)$? I conjecture that this inequality holds for the gate defined in Section 2.4.3.

- Is $E(U) = E(U^\dagger)$? Both quantities relate to how entangling a nonlocal gate is. However, we can only prove the equality when $U = U^T$, by using the fact $E(U) = E(U^*)^*$. This generalizes the proof in Ref. [BS03b] for 2-qubit gates since $U = U^T$ for all 2-qubit gates that are decomposed in the form of Eq. (2.1). Numerical work suggests that the equality does not hold for some U in higher dimensions [CLS02]. More generally, we can ask whether $\text{CCE}(U) = \text{CCE}(U^\dagger)$.

- Is $E(U)$ completely determined by the Schmidt coefficients of U?

- It seems unlikely that classical capacity can be determined by Schmidt coefficients alone, but can we derive better lower and upper bounds on classical capacity based on the Schmidt coefficients of a gate? Specifically, can we show that $C^E_+(U) \leq 2 \log \text{Sch}(U)$, or even better, that $\log \text{Sch}(U) \left([q \rightarrow q] + [q \leftarrow q]\right) \geq \langle U \rangle$? Right now these inequalities are only known to be true when $\text{Sch}(U)$ is maximal (i.e. equal to $d_A d_B$ when $U \in \mathcal{U}_{d_A \times d_B}$).

[*] This is because $\max_\psi E(U|\psi\rangle) - E(|\psi\rangle) = \max_\psi E(U|\psi^*\rangle) - E(|\psi^*\rangle) = \max_\psi E(U^*|\psi\rangle) - E(|\psi\rangle)$.

Chapter 3

Coherent classical communication

3.1 Introduction and definition

One of the main differences between classical and quantum Shannon theory is the number of irreversible, but optimal, resource transformations that exist in quantum Shannon theory. The highest rate that ebits or cbits can be created from qubits is one-for-one: $[q \to q] \geq [q\,q]$ and $[q \to q] \geq [c \to c]$. But the best way to create qubits from cbits and ebits is teleportation: $2[c \to c] + [q\,q] \geq [q \to q]$. These protocols are all asymptotically optimal—for example, the classical communication requirement of teleportation cannot be decreased even if entanglement is free—but composing them is extremely wasteful: $3[q \to q] \geq 2[c \to c] + [q\,q] \geq [q \to q]$. This sort of irreversibility represents one of the main challenges of quantum information theory: resources may be qualitatively equivalent but quantitatively incomparable.

In this chapter we will introduce a new primitive resource: the *coherent bit* or *cobit*. To emphasize its connection with classical communication, we denote the asymptotic resource (defined below) by $[\![c \to c]\!]$.[*] Coherent classical communication will simplify and improve a number of topics in quantum Shannon theory:

- We will find that coherently decoupled cbits can be described more simply and naturally as cobits.

- Replacing coherently decoupled cbits with cobits will make many resource transformations reversible. In particular, teleportation and super-dense coding become each other's inverses, a result previously only known when unlimited entanglement is allowed.

- More generally, we find that many forms of irreversibility in quantum Shannon theory are equivalent to the simple map $[\![c \to c]\!] \geq [c \to c]$.

- We will expand upon Proposition 2.10 to precisely explain how cbits are more powerful when they are sent through unitary means. This has a number of consequences for unitary gate capacities.

- In the next chapter, coherent classical communication will be used to relate many of the different protocols in quantum Shannon theory, give simple proofs of some existing protocols and create some entirely new protocols. These will allow us to determine two-dimensional tradeoff curves for the capacities of channels and states to create or consume cbits, ebits and qubits.

Coherent classical communication can be defined in two ways, which we later show to be equivalent.

[*] Other work[DHW05, Dev05b] uses $[q \to qq]$ to denote cobits, in order to emphasize their central place among isometries from A to AB.

- *Explicit definition in terms of finite resources:*

  Fix a basis for $\mathbb{C}^d$: $\{|x\rangle\}_{x=0}^{d-1}$. First, we recall from Section 1.2.1 the definitions of quantum and classical communication: $\mathrm{id}_d = \sum_x |x\rangle^B\langle x|^{A'}$ (a perfect quantum channel), $\overline{\mathrm{id}}_d = \sum_x |x\rangle^B|x\rangle^E\langle x|^{A'}$ (a perfect classical channel in the QP formalism) and $\overline{\Delta}_d = \sum_x |x\rangle^A|x\rangle^B|x\rangle^E\langle x|^{A'}$ (the classical copying operation in the QP formalism). Then we define a perfect coherent channel as

  $$\Delta_d = \sum_{x=0}^{d-1} |x\rangle^A|x\rangle^B\langle x|^{A'}. \tag{3.1}$$

  It can be thought of as a purification of a cbit in which Alice controls the environment, as a sort of quantum analogue to a feedback channel. The asymptotic resource is then given by $[\![c \to c]\!] := \langle\Delta_2\rangle$.

- *Operational definition as an asymptotic resource:* We can also define a cobit as a cbit sent through unitary, or more generally isometric, means. The approximate version of this statement is that whenever a protocol creates coherently decoupled cbits (cf. Definition 1.14), then a modified version of the protocol will create cobits. Later we will prove a precise form of this statement, known as "Rule O," because it describes how output cbits should be made coherent.

  When C *input cbits* are coherently decoupled (cf. Definition 1.13) we instead find that replacing them with C cobits results in C extra ebits being generated in the output. This input rule is known as "Rule I." Both rules are proved in Section 3.5.

  The canonical example of coherent decoupling is when cbits are sent using a unitary gate. In Theorem 3.1, we show that cbits sent through unitary means can indeed be coherently decoupled, and thereby turned into cobits.

The rest of the chapter is organized as follows.

Section 3.2 will give some simple examples of how cobits can be obtained.

Section 3.3 will then describe how to use coherent classical communication to make quantum protocols reversible and more efficient. It will conclude with a precise statement of Rules I and O.

Section 3.4 will apply these general principles to remote state preparation[BHL$^+$05], which leads to new protocols for super-dense coding of quantum state[HHL04] as well as many new results for unitary gate capacities.

Section 3.5 collects some of the longer proofs from the chapter, in order to avoid interrupting the exposition of the rest of the chapter.

Section 3.6 concludes with a brief discussion.

Bibliographical note: Most of this chapter is based on [Har04], though in Section 3.5 the proofs of Rules I and O are from [DHW05] (joint work with Igor Devetak and Andreas Winter). and the proof of Theorem 3.1 is from [HL05] (joint work with Debbie Leung).

3.2 Sources of coherent classical communication

Qubits and cbits arise naturally from noiseless and dephasing channels respectively, and can be obtained from any noisy channel by appropriate coding [Hol98, SW97, Llo96, Sho02, Dev05a]. Similarly, we will show both a natural primitive yielding coherent bits and a coding theorem that can generate coherent bits from a broad class of unitary operations.

The simplest way to send a coherent message is by modifying super-dense coding (SD). In SD, Alice and Bob begin with $|\Phi_2\rangle$ and want to use id_2 to send a two bit message a_1a_2 from Alice to Bob. Alice encodes her message by applying $Z^{a_1}X^{a_2}$ to her half of $|\Phi_2\rangle$ and then sending it to Bob, who decodes by applying $(H \otimes I)\textsc{cnot}$ to the state, obtaining

$$(H \otimes I)\textsc{cnot}(Z^{a_1}X^{a_2} \otimes I)|\Phi_2\rangle = |a_1\rangle|a_2\rangle$$

Now modify this protocol so that Alice starts with a quantum state $|a_1a_2\rangle$ and applies $Z^{a_1}X^{a_2}$ to her half of $|\Phi_2\rangle$ conditioned on her quantum input. After she sends her qubit and Bob decodes, they will be left with the state $|a_1a_2\rangle^A|a_1a_2\rangle^B$. Thus,

$$[q \to q] + [q\,q] \overset{*}{\geq} 2[\![c \to c]\!] \tag{3.2}$$

In fact, any unitary operation capable of classical communication is also capable of an equal amount of coherent classical communication, though in general this only holds asymptotically. The following theorem gives a general prescription for obtaining coherent communication and proves part of the equivalence of the two definitions of cobits given in the introduction.

Theorem 3.1. *For any bipartite unitary or isometry U, if*

$$\langle U\rangle \geq C_1[c \to c] + C_2[c \leftarrow c] + E[q\,q] \tag{3.3}$$

for $C_1, C_2 \geq 0$ and $E \in \mathbb{R}$ then

$$\langle U\rangle \geq C_1[\![c \to c]\!] + C_2[\![c \leftarrow c]\!] + E[q\,q] \tag{3.4}$$

If we define $\mathrm{C_oC_oE}(U) = \{(C_1, C_2, E) : \langle U\rangle \geq C_1[\![c \to c]\!] + C_2[\![c \leftarrow c]\!] + E[q\,q]\}$, then this theorem states that $\mathrm{CCE}(U)$ and $\mathrm{C_oC_oE}(U)$ coincide on the quadrant $C_1, C_2 \geq 0$.

Here we will prove only the case where $C_2 = 0$, deferring the full bidirectional proof to Section 3.5. By appropriate coding (as in [BS03a]), we can reduce the one-way case of Theorem 3.1 to the following coherent analogue of HSW coding.

Lemma 3.2 (Coherent HSW). *Given a PP ensemble of bipartite pure states*

$$|\mathcal{E}\rangle = \sum_{x\in\mathcal{X}} \sqrt{p_x}|x\rangle^R|x\rangle^{X_A}|\psi_x\rangle^{AB} \tag{3.5}$$

and an isometry

$$U_{\mathcal{E}} = \sum_x |x\rangle\!\langle x|^{X_A} \otimes |\psi_x\rangle^{AB} \tag{3.6}$$

then

$$\langle U_{\mathcal{E}} : \mathcal{E}^{X_A}\rangle \geq I(X_A;B)_{\mathcal{E}}[\![c \to c]\!] + H(B|X_A)[q\,q]. \tag{3.7}$$

Proof. A slightly modified form of HSW coding (e.g. [Dev05a]) holds that for any $\delta > 0, \epsilon > 0$ and every n sufficiently large there exists a code $\mathcal{C} \subset \mathcal{S}^n$ with $|\mathcal{C}| = \exp(n(I(X_A;B)_{\mathcal{E}} - \delta))$, a decoding POVM $\{D_{c^n}\}_{c^n\in\mathcal{C}}$ with error $< \epsilon$ and a type q with $\|p - q\|_1 \leq |\mathcal{X}|/n$ such that every codeword $c^n := c_1 \ldots c_n \in \mathcal{C}$ (corresponding to the state $|\psi_{c^n}\rangle^{AB} := |\psi_{c_1}\rangle^{A_1B_1} \cdots |\psi_{c_n}\rangle^{A_nB_n}$) has type q (i.e. $\forall x, |\{c_j = x\}| = nq_x$). By error $< \epsilon$, we mean that for any $c^n \in \mathcal{C}$, $\langle\psi_{c^n}|(I \otimes D_c)|\psi_{c^n}\rangle > 1 - \epsilon$.

Using Neumark's Theorem[Per93], Bob can make his decoding POVM into a unitary operation U_D defined by $U_D|0\rangle|\phi\rangle = \sum_{c^n} |c^n\rangle\sqrt{D_{c^n}}|\phi\rangle$. Applying this to his half of a codeword $|\psi_{c^n}\rangle$ will yield a state within ϵ of $|c^n\rangle|\psi_{c^n}\rangle$, since measurements with nearly certain outcomes cause almost no disturbance[Win99a].

The communication strategy begins by applying $U_{\mathcal{E}}$ to $|c^n\rangle_{X_A}$ to obtain $|c^n\rangle^{X_A}|\psi_{c^n}\rangle^{AB}$. Bob then decodes unitarily with U_D to yield a state within ϵ of $|c^n\rangle^{X_A}|c^n\rangle^{X_B}|\psi_{c^n}\rangle^{AB}$. Since c^n is of type q, Alice and Bob can coherently permute the states of $|\psi_{c^n}\rangle$ to obtain a state within ϵ of $|c^n\rangle_{X_A}|c^n\rangle_{X_B}|\psi_1\rangle^{\otimes nq_1}\cdots|\psi_{|\mathcal{X}|}\rangle^{\otimes nq_{|\mathcal{X}|}}$. Then they can apply entanglement concentration[BBPS96] to $|\psi_1\rangle^{\otimes nq_1}\cdots|\psi_{|\mathcal{X}|}\rangle^{\otimes nq_{|\mathcal{X}|}}$ to obtain $\approx nH(B|X_A)_{\mathcal{E}}$ ebits without disturbing the coherent message $|c^n\rangle_{X_A}|c\rangle_{X_B}$. □

This will be partially superseded by the full proof of Theorem 3.1. However, it is worth appreciating the key ideas of the proof—making measurements coherent via Neumark's Theorem and finding a way to decouple ancillas shared by Alice and Bob—as they will appear again in the later proofs, but surrounded by more mathematical details.

There are many cases in which no ancillas are produced, so we do not need the assumptions of Lemma 3.2 that communication occurs in large blocks. For example, a CNOT can transmit one coherent bit from Alice to Bob or one coherent bit from Bob to Alice. Recall the protocol given in Eq. (2.28) for $\textsc{cnot} + [q\,q] \overset{*}{\geq} [c \to c] + [c \leftarrow c]$: $(Z^a H \otimes I)\textsc{cnot}(X^a \otimes Z^b)|\Phi_2\rangle^{AB} = |b\rangle^A|a\rangle^B$. This can be made coherent by conditioning the encoding on a quantum register $|a\rangle^{A'}|b\rangle^{B'}$, so that

$$\textsc{cnot} + [q\,q] \overset{*}{\geq} [\![c \to c]\!] + [\![c \leftarrow c]\!] \tag{3.8}$$

3.3 Rules for using coherent classical communication

By discarding her state after sending it, Alice can convert coherent communication into classical communication, so $[\![c \to c]\!] \geq [c \to c]$. Alice can also generate entanglement by inputting a superposition of messages (as in Proposition 2.10), so $[\![c \to c]\!] \geq [q\,q]$. The true power of coherent communication comes from performing both tasks—classical communication and entanglement generation—simultaneously. This is possible whenever the classical message sent is coherently decoupled, i.e. random and nearly independent of the other states at the end of the protocol.

Teleportation [BBC$^+$93] satisfies these conditions, and indeed a coherent version has already been proposed in [BBC98]. Given an unknown quantum state $|\psi\rangle^A$ and an EPR pair $|\Phi_2\rangle^{AB}$, Alice begins coherent teleportation not by a Bell measurement on her two qubits but by unitarily rotating the Bell basis into the computational basis via a CNOT and Hadamard gate. This yields the state $\frac{1}{2}\sum_{ij}|ij\rangle^A X^i Z^j|\psi\rangle^B$. Using two coherent bits, Alice can send Bob a copy of her register to obtain $\frac{1}{2}\sum_{ij}|ij\rangle^A|ij\rangle^B X^i Z^j|\psi\rangle^B$. Bob's decoding step can now be made unitary, leaving the state $(|\Phi_2\rangle^{AB})^{\otimes 2}|\psi\rangle^B$. In terms of resources, this can be summarized as: $2[\![c \to c]\!] + [q\,q] \overset{*}{\geq} [q \to q] + 2[q\,q]$. Canceling the ebits on both sides (possible since $[\![c \to c]\!] \geq [q\,q]$) gives $2[\![c \to c]\!] \geq [q \to q] + [q\,q]$. Combining this relation with Eq. (3.2) yields the equality*

$$2[\![c \to c]\!] = [q \to q] + [q\,q]. \tag{3.9}$$

This has two important implications. First, teleportation and super-dense coding are reversible so long as all of the classical communication is left coherent. Second, cobits are equivalent, as resources, to the existing resources of qubits and ebits. This means that we don't need to calculate quantities such as the cobit capacity of a quantum channel; coherent communication introduces a new tool for solving old problems in quantum Shannon theory, and is not directly a source of new problems.

Another protocol that can be made coherent is Gottesman's method[Got99] for simulating a distributed CNOT using one ebit and one cbit in either direction. At first glance, this appears completely irreversible, since a CNOT can be used to send one cbit forward or backwards, or to create one ebit, but no more than one of these at a time.

*Our use of the Cancellation Lemma means that this equality is only asymptotically valid. [vE05] proves a single-shot version of this equality, but it requires that the two cobits be applied in series, with local unitary operations in between.

Using coherent bits as inputs, though, allows the recovery of 2 ebits at the end of the protocol, so $[\![c \to c]\!] + [\![c \leftarrow c]\!] + [q\,q] \overset{*}{\geq} \langle \text{CNOT} \rangle + 2[q\,q]$, or using entanglement catalytically, $[\![c \to c]\!] + [\![c \leftarrow c]\!] \geq \langle \text{CNOT} \rangle + [q\,q]$. Combined with Eq. (2.28), this yields another equality:

$$\langle \text{CNOT} \rangle + [q\,q] = [\![c \to c]\!] + [\![c \leftarrow c]\!].$$

Another useful bipartite unitary gate is SWAP, which we recall is equivalent to $[q \to q] + [q \leftarrow q]$. Applying Eq. (3.9) then yields

$$2\langle \text{CNOT} \rangle = 1\langle \text{SWAP} \rangle$$

which explains the similar communication and entanglement capacities for these gates found in the last chapter. Previously, the most efficient methods known to transform between these gates gave $3\langle \text{CNOT} \rangle \geq 1\langle \text{SWAP} \rangle \geq 1\langle \text{CNOT} \rangle$.

A similar argument can be applied to DCNOT. Since $\langle \text{DCNOT} \rangle + 2[q\,q] \geq 2[c \to c] + 2[c \leftarrow c]$, it follows (from Theorem 3.1 or direct examination) that $\langle \text{DCNOT} \rangle + 2[q\,q] \geq 2[\![c \to c]\!] + 2[\![c \leftarrow c]\!]$ and that $\langle \text{DCNOT} \rangle \geq [q \to q] + [q \leftarrow q] = \langle \text{SWAP} \rangle$. Combining this with Proposition 2.8, we find that $\langle \text{DCNOT} \rangle = \langle \text{SWAP} \rangle$, a surprising fact in light of the observation of [HVC02] that DCNOT is easier for some nonlocal Hamiltonians to simulate than SWAP. In fact, by the same argument, any gate in $\mathcal{U}_{d \times d}$ with $C_+^E(U) = 4 \log d$ must be equivalent to the $d \times d$ SWAP gate.

The above examples give the flavor of when classical communication can be replaced by coherent communication (i.e. "made coherent.") In general, we require that the classical message be (almost) uniformly random and (almost) coherently decoupled from all other systems, including the environment. This leads us to two general rules regarding making classical communication coherent. When coherently-decoupled cbits are in the input to a protocol, Rule I ("input") says that replacing them with cobits not only performs the protocol, but also has the side effect of generating entanglement. Rule O ("output") is simpler; it says that if a protocol outputs coherently-decoupled cbits, then it can be modified to instead output cobits. Once coherently decoupled cbits are replaced with cobits we can then use Eq. (3.9) to in turn replace cobits with qubits and ebits. Thus, while cobits are conceptually useful, we generally start and finish with protocols involving the standard resources of cbits, ebits and qubits.

Below we give formal statements of rules I and O, deferring their proofs till the end of the chapter.

Theorem 3.3 (Rule I). *If, for some quantum resources $\alpha, \beta \in \mathcal{R}$,*

$$\alpha + R\,[c \to c : \tau] \geq \beta$$

and the classical resource $R\,[c \to c : \tau]$ is coherently decoupled then

$$\alpha + \frac{R}{2}\,[q \to q] \geq \beta + \frac{R}{2}\,[q\,q].$$

Remark: This can be thought of as a coherent version of Lemma 1.38.

The idea behind the proof is that replacing $R[c \to c : \tau]$ with $R[\![c \to c : \tau]\!]$ then gives an extra output of $R[q\,q]$, implying that $\alpha + R[\![c \to c : \tau]\!] \geq \beta + R[q\,q]$. Then $[\![c \to c : \tau]\!]$ can be replaced by $\frac{1}{2}([q \to q] + [q\,q])$ using Eq. (3.9) and Lemma 1.36. To prove this rigorously will require carefully accounting for the errors, which we will do in Section 3.5.

Theorem 3.4 (Rule O). *If, for some quantum resources $\alpha, \beta \in \mathcal{R}$,*

$$\alpha \geq \beta + R\,[c \to c]$$

and the classical resource is decoupled with respect to the RI then

$$\alpha \geq \beta + \frac{R}{2}\,[q\,q] + \frac{R}{2}\,[q \to q].$$

Here the proof is even simpler: $R[c \to c]$ in the output is replaced with $R[\![c \to c]\!]$, which is equivalent to $\frac{R}{2}([q \to q] + [q\,q])$. Again, the details are given in Section 3.5.

In the next chapter, we will show how Rules I and O can be used to obtain a family of optimal protocols (and trade-off curves) for generating cbits, ebits and qubits from noisy channels and states. First, we show a simpler example of how a protocol can be made coherent in the next section.

3.4 Applications to remote state preparation and unitary gate capacities

3.4.1 Remote state preparation

Remote state preparation (RSP) is the task of simulating a $\{c \to q\}$ channel, usually using cbits and ebits. In this section, we show how RSP can be made coherent, not only by applying Rule I to the input cbits, but also by replacing the $\{c \to q\}$ channel by a coherent version that will preserve superpositions of inputs. Finally, we will use this coherent version of RSP to derive the capacity of a unitary gate to send a classical message from Alice to Bob while using/creating an arbitrary amount of entanglement.

Begin by recalling from Section 1.4 our definition of RSP. Let $\mathcal{E} = \sum_i p_i \, |i\rangle\langle i|^{X_A} \otimes |\psi_i\rangle\langle\psi_i|^{AB}$ be an ensemble of bipartite states and $\mathcal{N}_\mathcal{E} : |i\rangle\langle i|^{X_A} \to |i\rangle\langle i|^{X_A} \otimes |\psi_i\rangle\langle\psi_i|^{AB}$ the $\{c \to q\}$ channel such that $\mathcal{N}(\mathcal{E}^{X_A}) = \mathcal{E}$. The main coding theorem of RSP[BHL+05] states that

$$I(X_A; B)_\mathcal{E}[c \to c] + H(B)_\mathcal{E}[q\,q] \geq \langle \mathcal{N}_\mathcal{E} : \mathcal{E}^{X_A} \rangle. \tag{3.10}$$

We will show that the input cbits in Eq. (3.10) are coherently decoupled, so that according to Rule I, replacing them with cobits will perform the protocol and return some entanglement at the same time. This reduces the entanglement cost to $H(B) - I(X_A; B) = H(B|X_A)$, so that

$$I(X_A : B)_\mathcal{E}[\![c \to c]\!] + H(B|X_A)_\mathcal{E}[q\,q] \geq \langle \mathcal{N}_\mathcal{E} : \mathcal{E}^{X_A} \rangle. \tag{3.11}$$

In fact, we can prove an even stronger statement, in which not only is the input coherently decoupled, but there is a sense in which the output is as well. Define a coherent analogue of $\mathcal{N}_\mathcal{E}$, which we call $U_\mathcal{E}$, by

$$U_\mathcal{E} = \sum_i |i\rangle\langle i|^{X_A} \otimes |\psi_i\rangle^{AB}. \tag{3.12}$$

We also replace the QP ensemble $\mathcal{E}$ with the (PP formalism) pure state $|\mathcal{E}\rangle$ given by

$$|\mathcal{E}\rangle = \sum_i \sqrt{p_i} |i\rangle^R |i\rangle^{X_A} |\psi_i\rangle^{AB}. \tag{3.13}$$

We will prove that

$$I(X_A : B)_\mathcal{E}[\![c \to c]\!] + H(B|X_A)_\mathcal{E}[q\,q] \geq \langle U_\mathcal{E} : \mathcal{E}^{X_A} \rangle. \tag{3.14}$$

Since $\langle U_\mathcal{E} : \mathcal{E}^{X_A} \rangle \geq \langle \mathcal{N}_\mathcal{E} : \mathcal{E}^{X_A} \rangle$, this RI implies Eq. (3.11); in particular, the presence of the reference system R ensures that $\mathcal{E}^{X_A}$ is the same in both cases, even if the $|\psi_i\rangle$ are not all orthogonal. Proving Eq. (3.14) will require careful examination of the protocol from [BHL+05], so we defer the details until Section 3.5.

Remark: An interesting special case is when $H(A)_\mathcal{E} = 0$, so that Alice is preparing pure states in Bob's lab rather than entangled states. In this case, $H(B|X_A)_\mathcal{E} = 0$ and Eq. (3.11) becomes simply

$$H(B)_\mathcal{E}[\![c \to c]\!] \geq \langle U_\mathcal{E} : \mathcal{E}^{X_A} \rangle. \tag{3.15}$$

Thus, if we say (following [BHL+05]) that Eq. (3.10) means that "1 cbit + 1 ebit ≥ 1 remote qubit," then Eq. (3.11) means that "1 cobit $\geq$ 1 remote qubit." Here "n remote qubits" mean the ability of Alice to prepare an n-qubit state of her choice in Bob's lab, though we cannot readily define an asymptotic resource corresponding to this ability, since it would violate the quasi-i.i.d. condition (Eq. (1.11)). Despite not being formally defined as a resource, we can think of remote qubits as intermediate in strength between qubits and cbits, just as cobits are; i.e. 1 qubit $\geq$ 1 remote qubit $\geq$ 1 cbit. As resources intermediate between qubits and cbits, remote qubits and cobits have complementary attributes: remote qubits share with qubits the ability to transmit arbitrary pure states, though they cannot create entanglement, while cobits can generate entanglement, but at first glance appear to only be able to faithfully transmit the computational basis states to Bob. Thus it is interesting that in fact 1 cobit $\geq$ 1 remote qubit, and that (due to [BHL$^+$05]) this map is optimal.

Eq. (3.11) yields two other useful corollaries, which we state in the informal language of remote qubits.

Corollary 3.5 (RSP capacity of unitary gates). *If U is a unitary gate or isometry with $\langle U\rangle \geq C[c \to c]$ then $\langle U\rangle \geq C\,$remote qubits$(\to)$.*

Corollary 3.6. (Super-dense coding of quantum states) $[q \to q] + [q\,q] \geq 2\,$remote qubits$(\to)$

More formally, we could say that if $H(B)_{\mathcal{E}} \leq C$ for an ensemble $\mathcal{E}$, then $\langle U\rangle \geq \langle U_{\mathcal{E}} : \mathcal{E}^{X_A}\rangle$, and similarly for Corollary 3.6. We can also express Corollary 3.6 entirely in terms of standard resources as

$$\tfrac{1}{2}I(X_A;B)_{\mathcal{E}}[q \to q] + \left(H(B)_{\mathcal{E}} - \tfrac{1}{2}I(X_A;B)_{\mathcal{E}}\right)[q\,q] \geq \langle U_{\mathcal{E}} : \mathcal{E}^{X_A}\rangle. \tag{3.16}$$

Though this last expression is not particularly attractive, it turns out to be optimal, and in fact to give rise to optimal trade-offs for performing RSP with the three resources of cbits, ebits and qubits[AH03] (see also [AHSW04] for a single-shot version of the coding theorem). We will find this pattern repeated many times in the next chapter; by making existing protocols coherent and using basic information-theoretic inequalities, we obtain a series of optimal tradeoff curves.

Corollary 3.6 was first proven directly in [HHL04] (see also [AHSW04]) and in fact, finding an alternate proof was the original motivation for the idea of coherent classical communication.

Coherent RSP: Now, we explore the consequences of the stronger version of coherent RSP in Eq. (3.14). Just as RSP and HSW coding reverse one another given free entanglement, coherent RSP (Eq. (3.14)) and coherent HSW coding (Lemma 3.2) reverse each other, even taking entanglement into account. Combining them gives the powerful equality

$$I(X_A : B)_{\mathcal{E}}[\![c \to c]\!] + H(B|X_A)[q\,q] = \langle U_{\mathcal{E}} : \mathcal{E}^{X_A}\rangle, \tag{3.17}$$

which improves the original RSP-HSW duality in Eq. (2.19) by eliminating the need for free entanglement. This remarkable statement simultaneously implies entanglement concentration, entanglement dilution, the HSW theorem and remote state preparation and super-dense coding of entangled states.[*]

3.4.2 One-way classical capacities of unitary gates

Here we will use Eq. (3.17) to determine the capacity of a unitary gate V to simultaneously send a classical message and generate or consume entanglement at any finite rate. The proof idea is similar to one in Theorem 2.9; we will use the equivalence between (coherent) ensembles and standard resources (cobits and ebits) to turn a one-shot improvement in mutual information and expected entanglement into an asymptotically efficient protocol. Now that we have an improved version of the duality between RSP and HSW coding, we obtain a precise accounting of the amount of entanglement generated/consumed.

[*] On the other hand, we had to use almost all of these statements in order to prove the result! Still it is nice to see them all unified in a single powerful equation. Also, recent work by Devetak[Dev05b] further generalizes the equalities that can be stated about isometries from A to AB.

Theorem 3.7. *Define* $\mathrm{CE}(V) := \{(C,E) : (C,0,E) \in \mathrm{CCE}(V)\}$ *and*

$$\Delta_{I,E}(V) := \left\{(C,E) : \exists\mathcal{E} \text{ s.t. } I(X_A;B)_{V(\mathcal{E})} - I(X_A;B)_{\mathcal{E}} \geq C \text{ and } H(B|X_A)_{V(\mathcal{E})} - H(B|X_A)_{\mathcal{E}} \geq E\right\}, \tag{3.18}$$

where $\mathcal{E}$ is an ensemble of bipartite pure states in AB conditioned on a classical register X_A.

Then $\mathrm{CE}(V)$ *is equal to the closure of* $\Delta_{I,E}(V)$*.*

Thus the asymptotic capacity using $-E$ ebits of assistance per use of V (or simultaneously outputting E ebits) equals the largest increase in mutual information possible with one use of V if the average entanglement decreases by no more than $-E$. Theorem 2.9 proved this for $E = -\infty$ and our proof here is quite similar. Note that the statement of the theorem is the same whether we consider QP ensembles $\mathcal{E}$ or PP ensembles $|\mathcal{E}\rangle$, though the proof will use the coherent version of RSP in Eq. (3.14).

Proof. Coding theorem: Suppose there exists an ensemble $\mathcal{E}$ with $C = I(X_A;B)_{V(\mathcal{E})} - I(X_A;B)_{\mathcal{E}}$ and $E = H(B|X_A)_{V(\mathcal{E})} - H(B|X_A)_{\mathcal{E}}$. Then

$$\begin{aligned}
\langle V\rangle + \langle U_{\mathcal{E}}\rangle &\geq \langle U_{V(\mathcal{E})}\rangle \\
&\geq I(X_A;B)_{V(\mathcal{E})}[\![c \to c]\!] + H(B|X_A)_{V(\mathcal{E})}[q\,q] \\
&\geq \big(I(X_A;B)_{V(\mathcal{E})} - I(X_A;B)_{\mathcal{E}}\big)\,[\![c \to c]\!] + \big(H(B|X_A)_{V(\mathcal{E})} - H(B|X_A)_{\mathcal{E}}\big)\,[q\,q] + \langle U_{\mathcal{E}}\rangle
\end{aligned}$$

Here the second RI used coherent HSW coding (Lemma 3.2) and the third RI used coherent RSP (Eq. (3.14)). We now use the Cancellation Lemma to show that $\langle V\rangle \geq C[\![c \to c]\!] + E[q\,q]$, implying that $(C,0,E) \in \mathrm{CCE}(V)$.

Converse: We will actually prove a stronger result, in which Bob is allowed unlimited classical communication to Alice. Thus, we will show that if $\langle V\rangle + \infty[c \leftarrow c] \geq C[c \to c] + E[q\,q]$, then there is a sequence of ensembles $\{\widetilde{\mathcal{E}}_n\}$ with $(I(X_A;B)_{V(\widetilde{\mathcal{E}}_n)} - I(X_A;B)_{\widetilde{\mathcal{E}}_n}, H(B|X_A)_{V(\widetilde{\mathcal{E}}_n)} - H(B|X_A)_{\widetilde{\mathcal{E}}_n})$ converging to (C,E) as $n \to \infty$. This will imply that (C,E) is in the closure of $\Delta_{I,E}(V)$.

Let $Y := Y_A Y_B$ the cumulative record of all of Bob's classical messages to Alice. Using the QP formalism, we assume without loss of generality that Bob always transmits his full measurement outcome (cf. Section III of [HL04]) so that Alice and Bob always hold a pure state conditioned on $X_A Y$; i.e. $H(AB|X_A Y) = 0$ and $H(A|X_A Y) = H(B|X_A Y) = I(A\rangle B X_A Y) = I(B\rangle A X_A Y)$.

First consider the case when $E > 0$. Fix a protocol which uses V n times to communicate $\geq n(C - \delta')$ bits and create $\geq n(E - \delta')$ ebits with error $\leq \epsilon$. They start with a product state $\mathcal{E}_0$ for which $I(X_A;B)_{\mathcal{E}_0} = 0$ and $H(B|X_A)_{\mathcal{E}_0} = 0$. Denote their state immediately after j uses of V by $\mathcal{E}_j$. (Without loss of generality, we assume that the n uses of V are applied serially.) Then by Lemma 1.2, $I(X_A;B)_{\mathcal{E}_n} \geq n(C - \delta)$ and $H(B|X_A Y)_{\mathcal{E}_n} \geq n(E - \delta)$ where $\delta = O(\delta + \epsilon) \to 0$ as $n \to \infty$.

Now define the ensemble $\widetilde{\mathcal{E}}_n = \frac{1}{n}\sum_{j=1}^{n} |jj\rangle\langle jj|^{Z_A Z_B} \otimes V^\dagger(\mathcal{E}_j^{ABX_A Y_A Y_B})$. We think of $\hat{X} := X_A Y_A Z_A$ as the message variable and $\hat{B} := B Y_B Z_B$ as representing Bob's system. We will prove that $I(\hat{X};\hat{B})_{V(\widetilde{\mathcal{E}}_n)} - I(\hat{X};\hat{B})_{\widetilde{\mathcal{E}}_n} \geq C - \delta$ and that $H(\hat{B}|\hat{X})_{V(\widetilde{\mathcal{E}}_n)} - H(\hat{B}|\hat{X})_{\widetilde{\mathcal{E}}_n} \geq E - \delta$.

First consider the change in mutual information. Since $Y_A = Y_B$ and $Z_A = Z_B$ (as random variables), $I(\hat{X};\hat{B})_{\widetilde{\mathcal{E}}_n} = I(X_A Y_A Z_A; B Y_B Z_B)_{\widetilde{\mathcal{E}}_n} = I(X_A;B|YZ)_{\widetilde{\mathcal{E}}_n} + H(YZ)_{\widetilde{\mathcal{E}}_n}$ and similarly when we replace $\widetilde{\mathcal{E}}_n$ with $V(\widetilde{\mathcal{E}}_n)$. Since V doesn't act on Y or Z, we have $H(YZ)_{\widetilde{\mathcal{E}}_n} = H(YZ)_{V(\widetilde{\mathcal{E}}_n)}$ and thus

$$\begin{aligned}
I(\hat{X};\hat{B})_{V(\tilde{\mathcal{E}}_n)} - I(\hat{X};\hat{B})_{\tilde{\mathcal{E}}_n} &= I(X_A;B|YZ)_{V(\tilde{\mathcal{E}}_n)} - I(X_A;B|YZ)_{\tilde{\mathcal{E}}_n} \\
&= \frac{1}{n}\sum_{j=1}^{n} I(X_A;B|Y)_{\mathcal{E}_j} - I(X_A;B|Y)_{V^\dagger(\mathcal{E}_j)} \\
&= \frac{1}{n}\left(I(X_A;B|Y)_{\mathcal{E}_n} - I(X_A;B|Y)_{\mathcal{E}_0}\right) + \frac{1}{n}\sum_{j=1}^{n}\left(I(X_A;B|Y)_{\mathcal{E}_{j-1}} - I(X_A;B|Y)_{V^\dagger(\mathcal{E}_j)}\right) \\
&\geq C - \delta + \frac{1}{n}\sum_{j=1}^{n}\left(I(X_A;B|Y)_{\mathcal{E}_{j-1}} - I(X_A;B|Y)_{V^\dagger(\mathcal{E}_j)}\right)
\end{aligned} \tag{3.19}$$

Recall that going from $\mathcal{E}_{j-1}$ to $V^\dagger(\mathcal{E}_j)$ involves local unitaries, a measurement by Bob and classical communication of the outcome from Bob to Alice. We claim that $I(X_A;B|Y)$ does not increase under this process, meaning that the expression inside the sum on the last line is always nonnegative and that $I(\hat{X};\hat{B})_{V(\tilde{\mathcal{E}}_n)} - I(\hat{X};\hat{B})_{\tilde{\mathcal{E}}_n} \geq C - \delta$, implying our desired conclusion. To prove this, write $I(X_A;B|Y)$ as $I(X_A;BY) - I(X_A;Y)$. The $I(X_A;BY)$ term is nonincreasing due to the data-processing inequality[SN96], while $I(X_A;Y)$ can only increase since each round of communication only causes Y to grow.

Now we examine the change in entanglement.

$$\begin{aligned}
H(\hat{B}|\hat{X})_{V(\tilde{\mathcal{E}}_n)} - H(\hat{B}|\hat{X})_{\tilde{\mathcal{E}}_n} &= H(BY_BZ_B|X_AY_AZ_A)_{V(\tilde{\mathcal{E}}_n)} - H(BY_BZ_B|X_AY_AZ_A)_{\tilde{\mathcal{E}}_n} \\
&= H(B|X_AYZ)_{V(\tilde{\mathcal{E}}_n)} - H(B|X_AYZ)_{\tilde{\mathcal{E}}_n} \\
&= \frac{1}{n}\left(H(B|X_AY)_{\mathcal{E}_n} - H(B|X_AY)_{\mathcal{E}_0}\right) + \frac{1}{n}\sum_{j=1}^{n}\left(H(B|X_AY)_{\mathcal{E}_{j-1}} - H(B|X_AY)_{V^\dagger(\mathcal{E}_j)}\right) \\
&\geq E - \delta + \frac{1}{n}\sum_{j=1}^{n}\left(H(B|X_AY)_{\mathcal{E}_{j-1}} - H(B|X_AY)_{V^\dagger(\mathcal{E}_j)}\right)
\end{aligned} \tag{3.20}$$

We would like to show that this last term is positive, or equivalently that $H(B|X_AY)$ is at least as large for $\mathcal{E}_{j-1}$ as it is for $V^\dagger(\mathcal{E}_j)$. This change from $\mathcal{E}_{j-1}$ to $V^\dagger(\mathcal{E}_j)$ involves local unitaries, a measurement by Bob and another classical message from Bob to Alice, which we call Y_j. Also, call the first $j-1$ messages Y^{j-1}. Thus, we would like to show that $H(B|X_AY^{j-1})_{\mathcal{E}_{j-1}} - H(B|X_AY^{j-1}Y_j)_{V^\dagger(\mathcal{E}_j)} \geq 0$. This can be expressed as an average over $H(B)_{\mathcal{E}_{j-1|x,y^{j-1}}} - H(B|Y_j)_{V^\dagger(\mathcal{E}_{j|x,y^{j-1}})}$, where $\mathcal{E}_{j-1|x,y^{j-1}}$ indicates that we have conditioned $\mathcal{E}_{j-1}$ on $X_A = x$ and $Y^{j-1} = y^{j-1}$. This last quantity is positive because of principle that the average entropy of states output from a projective measurement is no greater than the entropy of the original state[Nie99a]. Thus $H(\hat{B}|\hat{X})_{V(\tilde{\mathcal{E}}_n)} - H(\hat{B}|\hat{X})_{\tilde{\mathcal{E}}_n} \geq E - \delta$.

As $n \to \infty$, $\delta \to 0$, proving the theorem.

The case when $E \leq 0$ is similar. We now begin with $H(B|X_AY_A)_{\mathcal{E}_0} \leq n(-E+\delta) = -n(E-\delta)$ and since X_A and Y are classical registers, finish with $H(B|X_AY_A)_{\mathcal{E}_n} \geq 0$. Thus $H(B|X_AY_A)_{\mathcal{E}_n} - H(B|X_AY_A)_{\mathcal{E}_0} \geq n(E-\delta)$. The rest of the proof is the same as the $E > 0$ case. □

3.4.3 Two-way cbit, cobit, qubit and ebit capacities of unitary gates

So far we have two powerful results about unitary gate capacity regions: Theorem 3.1 relates CCE and $\mathrm{C_oC_oE}$ in the $C_1, C_2 \geq 0$ quadrant and Theorem 3.7 gives an expression for CE(U) in terms of a single use of U. Moreover, the proof of Theorem 3.7 also showed that backwards classical communication cannot improve the forward capacity of a unitary gate. This allows us to extend Theorem 3.1 to $C_1 \leq 0$ or $C_2 \leq 0$ as follows:

Theorem 3.8. *For arbitrary real numbers C_1, C_2, E,*

$$(C_1, C_2, E) \in \mathrm{CCE} \Longleftrightarrow (C_1, C_2, E - \min(C_1, 0) - \min(C_2, 0)) \in \mathrm{C_oC_oE}\,. \tag{3.21}$$

This theorem is a direct consequence of the following Lemma, which enumerates the relevant quadrants of the (C_1, C_2) plane.

Lemma 3.9. *For any bipartite unitary or isometry U and $C_1, C_2 \geq 0$,*

$$\begin{aligned}
C_2[c \leftarrow c] + \langle U \rangle &\geq C_1[c \to c] + E[q\,q] \qquad \text{iff} && (3.22)\\
\langle U \rangle &\geq C_1[c \to c] + E[q\,q] \qquad \text{iff} && (3.23)\\
\langle U \rangle &\geq C_1\llbracket c \to c \rrbracket + E[q\,q] \qquad \text{iff} && (3.24)\\
C_2\llbracket c \leftarrow c \rrbracket + \langle U \rangle &\geq C_1\llbracket c \to c \rrbracket + (E{+}C_2)[q\,q] && (3.25)
\end{aligned}$$

and

$$\begin{aligned}
C_1[c \to c] + C_2[c \leftarrow c] + \langle U \rangle &\geq E[q\,q] \quad \text{iff} && (3.26)\\
\langle U \rangle &\geq E[q\,q] \quad \text{iff} && (3.27)\\
C_1\llbracket c \to c \rrbracket + C_2\llbracket c \leftarrow c \rrbracket + \langle U \rangle &\geq (E{+}C_1{+}C_2)[q\,q] && (3.28)
\end{aligned}$$

Basically, the rate at which Alice can send Bob cbits while consuming/generating ebits is not increased by (coherent) classical communication from Bob to Alice, except for a trivial gain of entanglement when the assisting classical communication is coherent.

Proof. Combining (TP) and coherent SD (Eq. (3.2)) yields $2[c \to c] + [q\,q] + [q \to q] + [q\,q] \overset{*}{\geq} [q \to q] + 2\llbracket c \to c \rrbracket$. Canceling the $[q \to q]$ from both sides and dividing by two gives us

$$[c \to c] + [q\,q] \geq \llbracket c \to c \rrbracket\,. \tag{3.29}$$

For the first part of the lemma, recall from the proof of Theorem 3.7 that free backcommunication does not improve the forward capacity of a gate. This means that Eq. (3.22) ⇒ Eq. (3.23). We obtain Eq. (3.23) ⇔ Eq. (3.24) from Theorem 3.1 and Eq. (3.24) ⇒ Eq. (3.25) follows from $\llbracket c \to c \rrbracket \geq [q\,q]$ and composability (Theorem 1.22). Finally, Eq. (3.25) ⇒ Eq. (3.22) because of Eq. (3.29).

For the second part of the theorem, Eq. (3.26) ⇒ Eq. (3.27) follows from Theorem 2.6, Eq. (3.27) ⇒ Eq. (3.28) is trivial and Eq. (3.28) ⇒ Eq. (3.26) is a consequence of Eq. (3.29). □

Quantum capacities of unitary gates: These techniques also allow us to determine the quantum capacities of unitary gates. Define QQE to be the region $\{(Q_1, Q_2, E) : U \geq Q_1[q \to q] + Q_2[q \leftarrow q] + E[q\,q]\}$, corresponding to two-way quantum communication. We can also consider coherent classical communication in one direction and quantum communication in the other; let $\mathrm{QC_oE}$ be the region $\{(Q_1, C_2, E) : U \geq Q_1[q \to q] + C_2\llbracket c \leftarrow c \rrbracket + E[q\,q]\}$ and define $\mathrm{C_oCQE}$ similarly.

As a warmup, we can use the equality $2\llbracket c \to c \rrbracket = [q \to q] + [q\,q]$ to relate $\mathrm{C_oE}$ and QE, defined as $\mathrm{C_oE} = \{(C, E) : (C, 0, E) \in \mathrm{C_oC_oE}\}$ and $\mathrm{QE} = \{(Q, E) : (Q, 0, E) \in \mathrm{QQE}\}$. We claim that

$$(Q, E) \in \mathrm{QE} \Leftrightarrow (2Q, E - Q) \in \mathrm{C_oE}\,. \tag{3.30}$$

To prove Eq. (3.30), choose any $(Q, E) \in \mathrm{QE}$. Then $\langle U \rangle \geq Q[q \to q] + E[q\,q] = 2Q\llbracket c \to c \rrbracket + (E - Q)[q\,q]$, so $(2Q, E - Q) \in \mathrm{C_oE}$. Conversely, if $(2Q, E - Q) \in \mathrm{C_oE}$, then $U \geq 2Q\llbracket c \to c \rrbracket + (E - Q)[q\,q] = Q[q \to q] + E[q\,q]$, so $(Q, E) \in \mathrm{QE}$.

Note that the above argument still works if we add the same resource, such as $Q_2[q \leftarrow q]$, to the right hand side of each resource inequality. Therefore, the same argument that proved Eq. (3.30) also establishes the following equivalences for bidirectional rate regions:

$$\begin{array}{ccc} (Q_1, Q_2, E) \in \mathrm{QQE} & \Longleftrightarrow & (2Q_1, Q_2, E - Q_1) \in \mathrm{C_oCQE} \\ \Updownarrow & & \Updownarrow \\ (Q_1, 2Q_2, E - Q_2) \in \mathrm{QC_oE} & \Longleftrightarrow & (2Q_1, 2Q_2, E - Q_1 - Q_2) \in \mathrm{C_oC_oE} \end{array} . \tag{3.31}$$

Finally, Eq. (3.21) further relates QQE, QCE, CQE, CCE, where QCE and CQE are defined similarly to $\mathrm{QC_oE}$ and $\mathrm{C_oCQE}$ but with incoherent classical communication instead.

Thus once one of the capacity regions (say $\mathrm{C_oC_oE}$) is determined, all other capacity regions discussed above are determined. The main open problem that remains is to find an efficiently computable expression for part of this capacity region. Theorem 3.7 gives a formula for the one-way cbit/ebit tradeoff that involves only a single use of the unitary gate, but we still need upper bounds on the optimal ensemble size and ancilla dimension for it to be practical.

3.5 Collected proofs

In this section we give proofs that various protocols can be made coherent. We start with Rules I and O (from Section 3.3), which gave conditions for when coherently decoupled cbits could be replaced by cobits in asymptotic protocols. Then we show specifically how remote state preparation can be made coherent, proving Eq. (3.14). Finally, we show how two-way classical communication from unitary operations can be made coherent, and prove Theorem 3.1.

3.5.1 Proof of Rule I

In what follows we shall fix ϵ and consider a sufficiently large n so that the protocol is ϵ-valid, ϵ^2-decoupled and accurate to within ϵ.

Whenever the resource inequality features $[c \to c]$ in the input this means that Alice performs a von Neumann measurement on some subsystem A_1 of dimension $D \approx \exp(n(R + \delta))$, the outcome of which she sends to Bob, who then performs an unitary operation depending on the received information.

Before Alice's von Neumann measurement, the joint state of A_1 and the remaining quantum system Q is

$$\sum_x \sqrt{p_x} |x\rangle^{A_1} |\phi_x\rangle^Q,$$

where by ϵ-validity

$$\sum_x |p_x - D^{-1}| \leq \epsilon.$$

Upon learning the measurement outcome x, Bob performs some unitary U_x on his part of Q, almost decoupling it from x:

$$\left\| \sum_x p_x |x\rangle\langle x| \otimes \theta'_x - \sum_x p_x |x\rangle\langle x| \otimes \overline{\theta}' \right\|_1 = \sum_x p_x \|\theta'_x - \overline{\theta}'\|_1 \leq \epsilon^2,$$

where $|\theta'_x\rangle = U_x |\phi_x\rangle$ and $\overline{\theta}' = \sum_x p_x \theta_x$. To simplify the analysis, extend Q to a larger Hilbert space on which there exist purifications $|\overline{\theta}\rangle\langle\overline{\theta}| \supseteq \overline{\theta}'$ and $|\theta_x\rangle\langle\theta_x| \supseteq |\theta'_x\rangle\langle\theta'_x|$ such that (according to

Lemma 1.3) $\|\theta_x - \overline{\theta}\|_1 \leq 2\sqrt{\|\theta'_x - \overline{\theta}'\|_1}$. Then

$$\sum_x p_x \|\theta_x - \overline{\theta}\|_1 \leq \sum_x p_x 2\sqrt{\|\theta'_x - \overline{\theta}'\|_1} \leq 2\sqrt{\sum_x p_x \|\theta'_x - \overline{\theta}'\|_1} \leq 2\epsilon, \tag{3.32}$$

where the second inequality uses the concavity of the square root.

If Alice refrains from the measurement and instead sends A_1 through a *coherent* channel, using $n(R+\delta)$ cobits, the resulting state is

$$\sum_x \sqrt{p_x} |x\rangle^{A_1} |x\rangle^{B_1} |\phi_x\rangle^Q.$$

Bob now performs the *controlled* unitary $\sum_x |x\rangle\langle x|^{B_1} \otimes U_x$, giving rise to

$$|\Upsilon\rangle^{A_1 B_1 Q} = \sum_x \sqrt{p_x} |x\rangle^{A_1} |x\rangle^{B_1} |\theta_x\rangle^Q.$$

We may assume, w.l.o.g., that $\langle\overline{\theta}|\theta_x\rangle$ is real and positive for all x, as this can be accomplished by either Alice or Bob via an x-dependent global phase rotation.

We now claim that $|\Upsilon\rangle^{A_1 B_1 Q}$ is close to $|\Phi_D\rangle^{A_1 B_1} |\overline{\theta}\rangle^Q$. Indeed

$$\langle\Upsilon|\Gamma\rangle|\overline{\theta}\rangle = \sum_x \sqrt{\frac{p_x}{D}} \langle\theta_x|\overline{\theta}\rangle \geq \sum_x \sqrt{\frac{p_x}{D}} \left(1 - \frac{1}{2}\|\theta_x - \theta\|_1\right), \tag{3.33}$$

according to Eq. (1.4). To bound this, we split the sum into two. For the first term, we apply Eq. (1.4) to the diagonal density matrices $\sum_x p_x |x\rangle\langle x|$ and $\sum_x D^{-1} |x\rangle\langle x|$ to obtain

$$\sum_x \sqrt{\frac{p_x}{D}} \geq 1 - \frac{1}{2}\sum_x |p_x - D^{-1}| \geq 1 - \frac{\epsilon}{2} \tag{3.34}$$

The second term is

$$\begin{aligned}
\sum_x \sqrt{\frac{p_x}{D}} \frac{1}{2}\|\theta_x - \theta\|_1 &= \sum_x \frac{1}{2}\left[p_x + \frac{1}{D} - \left(\sqrt{p_x} - \sqrt{1/D}\right)^2\right] \frac{1}{2}\|\theta_x - \theta\|_1 \\
&\leq \sum_x \frac{1}{2}\left(p_x + \frac{1}{D}\right) \frac{1}{2}\|\theta_x - \theta\|_1 \leq \sum_x \frac{1}{2}\left(2p_x + \left|p_x - \frac{1}{D}\right|\right) \frac{1}{2}\|\theta_x - \theta\|_1 \\
&\leq \sum_x p_x \frac{1}{2}\|\theta_x - \theta\|_1 + \sum_x \left|p_x - \frac{1}{D}\right| \leq 2\epsilon.
\end{aligned}$$

Putting this together, we find that

$$\langle\Upsilon|\Gamma\rangle|\overline{\theta}\rangle \geq 1 - 3\epsilon$$

and by Eq. (1.4),

$$\|\Upsilon - \Phi_D \otimes \overline{\theta}\|_1 \leq \sqrt{6\epsilon}$$

Finally, since tracing out subsystems cannot increase trace distance,

$$\|\Upsilon^{A_1 B_1} - \Phi_D\|_1 \leq \sqrt{6\epsilon}$$

Thus, the total effect of replacing cbits cobits is the generation of a state close to Φ_D. This analysis ignores the fact that the cobits are only given up to an error ϵ. However, due to the triangle inequality, this only enters in as an additive factor, and the overall error of $\epsilon + \sqrt{6\epsilon}$ is still asymptotically vanishing. Furthermore, this mapping preserves the ϵ-validity of the original protocol (with respect to the inputs of α) since all we have done to Alice's states is to add purifying systems and add phases, which w.l.o.g. we can assume are applied to these purifying systems.

We have thus shown

$$\alpha + R\,[\![c \to c]\!] \geq \beta + R\,[q\,q].$$

Eq. (3.9) and Lemmas 1.36 and 1.37 give the desired result

$$\alpha + \frac{R}{2}\,[q \to q] \geq \beta + \frac{R}{2}\,[q\,q].$$

□

3.5.2 Proof of Rule O

Again fix ϵ and consider a sufficiently large n so that the protocol is ϵ-valid, ϵ^2-decoupled and accurate to within ϵ. Now the roles of Alice and Bob are somewhat interchanged. Alice performs a unitary operation depending on the classical message x to be sent and Bob performs a von Neumann measurement on some subsystem B_1 which almost always succeeds in reproducing the message. Namely, if we denote by $p_{x'|x}$ the probability of outcome x' given Alice's message was x then, for sufficiently large n,

$$\frac{1}{D}\sum_x p_{x|x} \geq 1 - \epsilon.$$

Again $D = \exp(n(R+\delta))$. Before Bob's measurement, the state of B_1 and the remaining quantum system Q is

$$\sum_{x'} \sqrt{p_{x'|x}}|x'\rangle^{B_1}|\phi_{xx'}\rangle^{Q}.$$

Based on the outcome x' of his measurement, Bob performs some unitary $U_{x'}$ on Q, leaving the state of Q almost decoupled from xx':

$$\left\| \sum_{xx'} D^{-1}p_{x'|x}|x\rangle\langle x| \otimes |x'\rangle\langle x'| \otimes \theta'_{xx'} - \sum_{xx'} D^{-1}p_{x'|x}|x\rangle\langle x| \otimes |x'\rangle\langle x'| \otimes \overline{\theta}' \right\|_1 \leq \epsilon^2,$$

where $|\theta'_{xx'}\rangle = U_{x'}|\phi_{xx'}\rangle$ and $\overline{\theta}' = D^{-1}\sum_{xx'} p_{x'|x}\theta'_{xx'}$. Observe, as before, that we can use Lemma 1.3 to extend Q so that $\overline{\theta} \supseteq \overline{\theta}'$ and $\theta_{xx'} \supseteq \theta'_{xx'}$ are pure states, $\langle\overline{\theta}|\theta_{xx}\rangle$ is real and positive and $\|\theta_{xx'} - \overline{\theta}\|_1 \leq 2\sqrt{\|\theta'_{xx'} - \overline{\theta}'\|_1}$. Again we use the concavity of $x \to \sqrt{x}$ to bound

$$D^{-1}\sum_x p_{x|x}\|\theta_{xx} - \overline{\theta}\|_1 \leq D^{-1}\sum_{xx'} p_{x|x'}\|\theta_{xx'} - \overline{\theta}\|_1 \leq 2\epsilon.$$

We now modify the protocol so that instead Alice performs *coherent* communication. Given a subsystem A_1 in the state $|x\rangle^{A_1}$ she encodes via *controlled* unitary operations, yielding

$$|x\rangle^{A_1}\sum_{x'} \sqrt{p_{x'|x}}|x'\rangle^{B_1}|\phi_{xx'}\rangle^{Q}.$$

Bob refrains from measuring B_1 and instead performs the *controlled* unitary $\sum_{x'} |x'\rangle\langle x'|^{B_1} \otimes U_{x'}$, giving rise to

$$|x\rangle^{A_1}|\Upsilon_x\rangle^{B_1Q} = |x\rangle^{A_1}\left(\sum_{x'}\sqrt{p_{x'|x}}|x'\rangle^{B_1}\otimes|\theta_{xx'}\rangle^Q\right).$$

We claim that this is a good approximation for $R[\![c \to c : \tau]\!] + \langle\overline{\theta}\rangle$, and according to the correctness of the original protocol, $\overline{\theta}$ is close to the output of β_n. To check this, suppose Alice inputs $|\Phi_D\rangle^{RA_1}$ into the communication protocol. We will compare the actual state

$$|\Upsilon\rangle^{RA_1B_1Q} := D^{-\frac{1}{2}}\sum_x |x\rangle^R|x\rangle^{A_1}|\Upsilon_x\rangle^{B_1Q}$$

with the ideal state

$$|\Phi_{\mathrm{GHZ}}\rangle^{RA_1B_1}\otimes|\overline{\theta}\rangle^Q = D^{-\frac{1}{2}}\sum_x |x\rangle^R|x\rangle^{A_1}|x\rangle^{B_1}|\overline{\theta}\rangle^Q.$$

Their inner product is

$$\begin{aligned}\langle\Upsilon|\Phi_{\mathrm{GHZ}}\rangle|\overline{\theta}\rangle &= \frac{1}{D}\sum_x\sqrt{p_{x|x}}\langle\theta_{xx}|\overline{\theta}\rangle \geq \frac{1}{D}\sum_x p_{x|x}\langle\theta_{xx}|\overline{\theta}\rangle \geq \frac{1}{D}\sum_x p_{x|x}\left(1-\tfrac{1}{2}\left\|\theta_{xx}-\overline{\theta}\right\|_1\right)\\ &\geq \frac{1}{D}\sum_x p_{x|x} - \frac{1}{D}\sum_x \tfrac{1}{2}\left\|\theta_{xx}-\overline{\theta}\right\|_1 \geq (1-\epsilon)-\epsilon = 1-2\epsilon\end{aligned}$$

Thus, we can apply Eq. (1.4) to show that

$$\|\Upsilon - \Phi_{GHZ}\otimes\theta\|_1 \leq 2\sqrt{\epsilon}.$$

We have thus shown that

$$\alpha \geq \beta + R\,[q \to q : \tau].$$

Using Theorem 1.40 and Eq. (3.9) gives the desired result

$$\alpha \geq \beta + \frac{R}{2}\,[q\,q] + \frac{R}{2}\,[q \to q].$$

□

3.5.3 Proof of Coherent RSP (Eq. 3.14)

To prove that RSP can be made coherent, we review the proof of Eq. (3.10) from [BHL$^+$05] and show how it needs to be modified. We will assume knowledge of typical and conditionally typical projectors; for background on them, as well as the operator Chernoff bound used in the proof, see [Win99a].

The (slightly modified) proof from [BHL$^+$05] is as follows. Let $\mathcal{E} = \sum_i p_i\,|i\rangle\langle i|^{X_A}\otimes\psi_i^{AB}$ be an ensemble of bipartite states, for which we would like to simulate $\mathcal{N}_\mathcal{E}$ or $U_\mathcal{E}$. Alice is given a string $i^n = (i_1,\ldots,i_n)$ and wants to prepare the joint state $|\psi_{i^n}\rangle^{AB} := |\psi_{i_1}\rangle^{AB}\cdots|\psi_{i_n}\rangle^{AB}$. Let Q_{i^n} be the empirical distribution of i^n, i.e. the probability distribution on i obtained by sampling from i^n. We assume that $\|p - Q_{i^n}\|_1 \leq \delta$, and since our simulation of $U_\mathcal{E}$ will be used in some η-valid protocol, we can do so with error $\leq \eta + \exp(-O(n\delta^2))$. (Here $\eta,\delta \to 0$ as $n\to\infty$.) Thus, the protocol begins by Alice projecting onto the set of i^n with $\|p - Q_{i^n}\|_1 \leq \delta$, in contrast with the protocol in [BHL$^+$05], which begins by having Alice measure Q_{i^n} and send the result to Bob classically.

Define $\Pi^n_{\mathcal{E}^B|i^n,\delta}$ to be the conditionally typical projector for Bob's half of $|\psi_{i^n}\rangle^{AB}$, and let $\Pi^n_{\mathcal{E}^B,\delta}$ be the typical projector for n copies of $\mathcal{E}^B$. These projectors are defined in [Win99a], which also proves that the subnormalized state

$$|\psi'_{i^n}\rangle = (\mathbb{1}\otimes\Pi^n_{\mathcal{E}^B,\delta}\Pi^n_{\mathcal{E}^B|i^n,\delta})|\psi_{i^n}\rangle, \tag{3.35}$$

satisfies $\langle\psi'_{i^n}|\psi'_{i^n}\rangle \geq 1-2\epsilon$, where $\delta,\epsilon\to 0$ as $n\to\infty$. This implies that $\|\psi_{i^n}-\psi''_{i^n}\|_1 \leq 2\sqrt{\epsilon}$, where we define the normalized state $|\psi''_{i^n}\rangle := |\psi'_{i^n}\rangle/\sqrt{\langle\psi'_{i^n}|\psi'_{i^n}\rangle}$. We will now write $|\psi'_{i^n}\rangle$ in a way which suggests how to construct it. Let $|\Phi_D\rangle^{AB}$ be a maximally entangled state with $D := \operatorname{rank}\Pi^n_{\mathcal{E}^B,\delta}$ and $\Phi^B_D = \Pi^n_{\mathcal{E}^B,\delta}/D$. (By contrast, [BHL+05] chooses Φ to be a purification of $\Pi^n_{\sigma,\delta}$ with $\sigma := \sum_x Q_{i^n}(x)\psi^B_x$.) Then $|\psi'_{i^n}\rangle$ can be written as $(M_{i^n}\otimes\mathbb{1})|\Phi_D\rangle$ where $\operatorname{Tr} M^\dagger_{i^n}M_{i^n} = D^{-1}\langle\psi'_{i^n}|\psi'_{i^n}\rangle$. Thus, Alice will apply a POVM composed of rescaled and rotated versions of M_{i^n} to her half of $|\Phi_D\rangle$, and after transmitting the measurement outcome k to Bob, he can undo the rotation and obtain his half of the correct state. The cost of this procedure is $\log D$ ebits and $\log K$ ebits, where we will later specify the number of POVM outcomes K.

We now sketch the proof that this is efficient. From [Win99a], we find the bounds

$$D = \operatorname{rank}\Pi^n_{\mathcal{E}^B,\delta} \leq \exp\left(n(H(B)_{\mathcal{E}}+\delta)\right) \tag{3.36}$$

$$\operatorname{Tr}_A |\psi'_{i^n}\rangle\langle\psi'_{i^n}| \leq \exp\left(-n(H(B|X_A)_{\mathcal{E}}+\delta)\right)\Pi^n_{\mathcal{E}^B,\delta} \tag{3.37}$$

Combining these last two equations and Eq. (3.35) with the operator Chernoff bound[Win99a] means that there exist a set of unitaries $U_1,\ldots,U_K$ such that $\log K \leq n(I(X_A;B)_{\mathcal{E}}+3\delta+o(1))$ and whenever $\|Q_{i^n}-p\|_1\leq\delta$ we have

$$(1-\epsilon)\frac{\Pi^n_{\mathcal{E}^B,\delta}}{D} \leq \frac{1}{K}\sum_{k=1}^{K}\frac{U_k^\dagger M^\dagger_{i^n}M_{i^n}U_k}{\operatorname{Tr} M^\dagger_{i^n}M_{i^n}} \leq (1+\epsilon)\frac{\Pi^n_{\mathcal{E}^B,\delta}}{D}. \tag{3.38}$$

These conditions mean that Alice can construct a POVM $\{A_1^{(i^n)},\ldots,A_K^{(i^n)},A_{\text{fail}}^{(i^n)}\}$ with

$$\begin{aligned} A_k^{(i^n)} &:= \frac{D}{\sqrt{K(1+\epsilon)\operatorname{Tr} M^\dagger_{i^n}M_{i^n}}}M_{i^n}U_k^* \\ A_{\text{fail}}^{(i^n)} &:= \sqrt{\Pi^n_{\mathcal{E}^B,\delta}-\sum_k A_k^\dagger A_k} \end{aligned} \tag{3.39}$$

According to Eq. (3.38), the "fail" outcome has probability $\leq 2\epsilon$ of occurring when Alice applies this POVM to half of $|\Phi_D\rangle$. And since $(U_k^*\otimes\mathbb{1})|\Phi_D\rangle = (\mathbb{1}\otimes U_k^\dagger)|\Phi_D\rangle$, if Alice sends Bob the outcome k and Bob applies U_k then the residual state will be $|\psi''_{i^n}\rangle$.

We now explain how to make the above procedure coherent. First observe that conditioned on not observing the "fail" outcome, the residual state is completely independent of the classical message k. Thus, we can apply Rule I. However, a variant of Rule O is also applicable, in that there is no need to assume the input $|i^n\rangle$ is a classical register. Again conditioning on success, the only record of i^n in the final state is the output state $|\psi''_{i^n}\rangle$. Thus, if Alice performs the POVM

$$A_k := \sum_{i^n}|i^n\rangle\langle i^n|\otimes A_k^{(i^n)}, \tag{3.40}$$

(with A_{fail} defined similarly) and Bob decodes using

$$\sum_k |k\rangle\langle k|\otimes U_k \tag{3.41}$$

then (conditioned on a successful measurement outcome) $\sum_{i^n}\sqrt{p_{i^n}}|i^n\rangle^R|i^n\rangle^{X_A}$ will be coherently mapped to $\sum_{i^n} \sqrt{p_{i^n}} |i^n\rangle^R |i^n\rangle^{X_A} |\psi''_{i^n}\rangle^{AB}$. This achieves a simulation of $\langle U_{\mathcal{E}} : \mathcal{E}^{X_A} \rangle$ using $I(X_A; B)$ cobits and $H(B)$ ebits. According to Rule I, the coherent communication returns $I(X_A; B)$ ebits at the end of the protocol, bringing the net entanglement cost down to $H(B|X_A)$. Thus we have proven Eq. (3.14). □

3.5.4 Proof of Theorem 3.1

For ease of notation, we first consider the $E = 0$ case, so our starting hypothesis is that $\langle U \rangle \geq C_1[c \to c] + C_2[c \leftarrow c]$. At the end of the proof we will return to the $E \neq 0$ case.

The definition of $\mathcal{P}_n$

Formally, Eq. (3.3) indicates the existence of sequences of nonnegative real numbers $\{\epsilon_n\}, \{\delta_n\}$ satisfying $\epsilon_n, \delta_n \to 0$ as $n \to \infty$; a sequence of protocols $\mathcal{P}_n = (V_n \otimes W_n)\, U \cdots U\, (V_1 \otimes W_1)\, U\, (V_0 \otimes W_0)$, where V_j, W_j are local isometries that may also act on extra local ancilla systems, and sequences of integers $C_1^{(n)}, C_2^{(n)}$ satisfying $nC_1 \geq C_1^{(n)} \geq n(C_1 - \delta_n)$, $nC_2 \geq C_2^{(n)} \geq n(C_2 - \delta_n)$, such that the following success criterion holds.

Let $a \in \{0,1\}^{C_1^{(n)}}$ and $b \in \{0,1\}^{C_2^{(n)}}$ be the respective messages of Alice and Bob. Let $|\varphi_{ab}\rangle := \mathcal{P}_n(|a\rangle_{\mathrm{A}_1} |b\rangle_{\mathrm{B}_1})$. Note that $|\varphi_{ab}\rangle$ generally occupies a space of larger dimension than $\mathrm{A}_1 \otimes \mathrm{B}_1$ since $\mathcal{P}_n$ may add local ancillas. To say that $\mathcal{P}_n$ can transmit classical messages, we require that local measurements on $|\varphi_{ab}\rangle$ can generate messages b' for Alice and a' for Bob according to a distribution $\Pr(a'b'|ab)$ such that

$$\forall_{a,b} \quad \sum_{a',b'} \tfrac{1}{2}\, |\Pr(a'b'|ab) - \delta_{a,a'} \delta_{b,b'}| \leq \epsilon_n \tag{3.42}$$

where a', b' are summed over $\{0,1\}^{C_1^{(n)}}$ and $\{0,1\}^{C_2^{(n)}}$ respectively. Eq. (3.42) follows from applying our definition of a protocol to classical communication, taking the final state to be the distribution of the output classical messages. Since any measurement can be implemented as a joint unitary on the system and an added ancilla, up to a redefinition of V_n, W_n, we can assume

$$|\varphi_{ab}\rangle := \mathcal{P}_n(|a\rangle_{\mathrm{A}_1} |b\rangle_{\mathrm{B}_1}) = \sum_{a',b'} |b'\rangle_{\mathrm{A}_1} |a'\rangle_{\mathrm{B}_1} |\gamma^{a,b}_{a',b'}\rangle_{\mathrm{A}_2\,\mathrm{B}_2} \tag{3.43}$$

where the dimensions of A_1 and B_1 are interchanged by $\mathcal{P}_n$, and $|\gamma^{a,b}_{a',b'}\rangle$ are subnormalized states with $\Pr(a'b'|ab) := \langle \gamma^{a,b}_{a',b'} | \gamma^{a,b}_{a',b'} \rangle$ satisfying Eq. (3.42). Thus, for each a, b most of the weight of $|\varphi_{ab}\rangle$ is contained in the $|\gamma^{a,b}_{a,b}\rangle$ term, corresponding to error-free transmission of the messages. See Fig. I(a).

The three main ideas for turning classical communication into coherent classical communication

We first give an informal overview of the construction and the intuition behind it. For simplicity, consider the error-free term with $|\gamma^{a,b}_{a,b}\rangle$ in $\mathrm{A}_2\,\mathrm{B}_2$. To see why classical communication via unitary means should be equivalent to coherent classical communication, consider the special case when $|\gamma^{a,b}_{a,b}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}$ is independent of a, b. In this case, copying a, b to local ancilla systems $\mathrm{A}_0, \mathrm{B}_0$ before $\mathcal{P}_n$ and discarding $\mathrm{A}_2\,\mathrm{B}_2$ after $\mathcal{P}_n$ leaves a state within trace distance ϵ_n of $|b\rangle_{\mathrm{A}_1} |a\rangle_{\mathrm{A}_0} |a\rangle_{\mathrm{B}_1} |b\rangle_{\mathrm{B}_0}$—the desired coherent classical communication. See Fig. I(b). In general $|\gamma^{a,b}_{a,b}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}$ will carry information about a, b, so tracing $\mathrm{A}_2\,\mathrm{B}_2$ will break the coherence of the classical communication. Moreover, if the Schmidt coefficients of $|\gamma^{a,b}_{a,b}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}$ depend on a, b, then knowing a, b is not sufficient to coherently eliminate $|\gamma^{a,b}_{a,b}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}$ without some additional communication. The remainder of our proof is built around the need to coherently eliminate this ancilla.

Our first strategy is to *encrypt* the classical messages a, b by a shared key, in a manner that preserves coherence (similar to that in [Leu02]). The coherent version of a shared key is a maximally entangled state. Thus Alice and Bob (1) again copy their messages to $\mathrm{A}_0, \mathrm{B}_0$, then (2) encrypt, (3) apply $\mathcal{P}_n$, and (4) decrypt. Encrypting the message makes it possible to (5) almost decouple the message from the combined "key-and-ancilla" system, which is approximately in a state $|\Gamma_{00}\rangle$ independent of a, b (exact definitions will follow later). (6) Tracing out $|\Gamma_{00}\rangle$ gives the desired coherent communication. Let $\mathcal{P}'_n$ denote steps (1)-(5) (see Fig. I(c)).

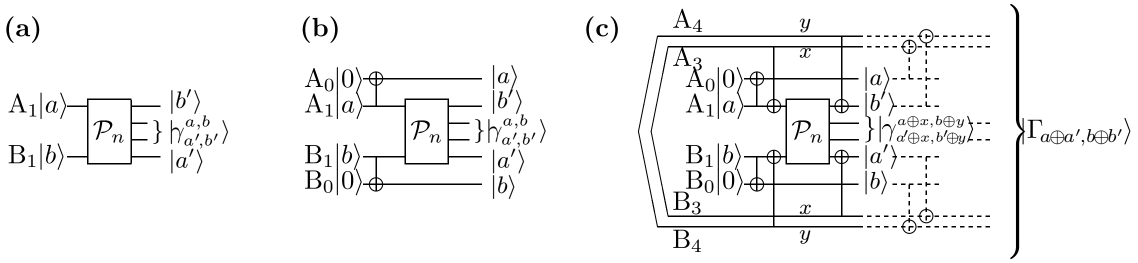


Figure 3-1: Schematic diagrams for $\mathcal{P}_n$ and $\mathcal{P}'_n$. (a) A given protocol $\mathcal{P}_n$ for two-way classical communication. The output is a superposition (over all a', b') of the depicted states, with most of the weight in the $(a', b') = (a, b)$ term. The unlabeled output systems in the state $|\gamma^{a,b}_{a',b'}\rangle$ are $\mathrm{A}_2, \mathrm{B}_2$. (b) The same protocol with the inputs copied to local ancillas $\mathrm{A}_0, \mathrm{B}_0$ before $\mathcal{P}_n$. If $|\gamma^{a,b}_{a,b}\rangle$ is independent of a, b, two-way coherent classical communication is achieved. (c) The five steps of $\mathcal{P}'_n$. Steps (1)-(4) are shown in solid lines. Again, the inputs are copied to local ancillas, but $\mathcal{P}_n$ is used on messages encrypted by a coherent one-time-pad (the input $|a\rangle_{\mathrm{A}_1}$ is encrypted by the coherent version of the key $|x\rangle_{\mathrm{A}_3}$ and the output $|a' \oplus x\rangle_{\mathrm{B}_1}$ is decrypted by $|x\rangle_{\mathrm{B}_3}$; similarly, $|b\rangle_{\mathrm{B}_1}$ is encrypted by $|y\rangle_{\mathrm{B}_4}$ and $|b' \oplus y\rangle_{\mathrm{A}_1}$ decrypted by $|y\rangle_{\mathrm{A}_4}$. The intermediate state is shown in the diagram. Step (5), shown in dotted lines, decouples the messages in $\mathrm{A}_{0,1}, \mathrm{B}_{0,1}$ from $\mathrm{A}_{2,3,4}, \mathrm{B}_{2,3,4}$, which is in the joint state very close to $|\Gamma_{00}\rangle$.

If entanglement were free, then our proof of Theorem 3.1 would be finished. However, we have borrowed $C_1^{(n)}+C_2^{(n)}$ ebits as the encryption key and replaced it with $|\Gamma_{00}\rangle$. Though the entropy of entanglement has not decreased (by any significant amount), $|\Gamma_{00}\rangle$ is not directly usable in subsequent runs of $\mathcal{P}'_n$. To address this problem, we use a second strategy of running k copies of $\mathcal{P}'_n$ in parallel and performing entanglement concentration of $|\Gamma_{00}\rangle^{\otimes k}$. For sufficiently large k, with high probability, we recover most of the starting ebits. The regenerated ebits can be used for more iterations of $\mathcal{P}'^{\otimes k}_n$ to offset the cost of making the initial $k\left(C_1^{(n)}+C_2^{(n)}\right)$ ebits, without the need of borrowing from anywhere.

However, a technical problem arises with simple repetition of $\mathcal{P}'_n$, which is that errors accumulate. In particular, a naïve application of the triangle inequality gives an error $k\epsilon_n$ but k, n are not independent. In fact, the entanglement concentration procedure of [BBPS96] requires $k \gg \mathrm{Sch}(|\Gamma_{00}\rangle) = \exp(O(n))$ and we cannot guarantee that $k\epsilon_n \to 0$ as $k, n \to \infty$. Our third strategy is to treat the k uses of $\mathcal{P}'_n$ as k uses of a slightly noisy channel, and encode only l messages (each having $C_1^{(n)}, C_2^{(n)}$ bits in the two directions) using classical error correcting codes. The error rate then vanishes with a negligible reduction in the communication rate and now making no assumption about how quickly ϵ_n approaches zero. We will see how related errors in decoupling and entanglement concentration are suppressed.

We now describe the construction and analyze the error in detail.

The definition of $\mathcal{P}'_n$

0. Alice and Bob begin with inputs $|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1}$ and the entangled states $|\Phi\rangle^{\otimes C_1^{(n)}}_{\mathrm{A}_3\,\mathrm{B}_3}$ and $|\Phi\rangle^{\otimes C_2^{(n)}}_{\mathrm{A}_4\,\mathrm{B}_4}$. (Systems 3 and 4 hold the two separate keys for the two messages a and b.) The initial state can then be written as
$$\frac{1}{\sqrt{N}}\sum_x |xx\rangle_{\mathrm{A}_3\,\mathrm{B}_3}\sum_y |yy\rangle_{\mathrm{A}_4\,\mathrm{B}_4}\ |a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1} \tag{3.44}$$
where x and y are summed over $\{0,1\}^{C_1^{(n)}}$ and $\{0,1\}^{C_2^{(n)}}$, and $N = \exp\left(C_1^{(n)}+C_2^{(n)}\right)$.

1. They coherently copy the messages to $\mathrm{A}_0, \mathrm{B}_0$.

2. They encrypt the messages using the one-time-pad $|a\rangle_{\mathrm{A}_1}|x\rangle_{\mathrm{A}_3} \rightarrow |a\oplus x\rangle_{\mathrm{A}_1}|x\rangle_{\mathrm{A}_3}$ and $|b\rangle_{\mathrm{B}_1}|y\rangle_{\mathrm{B}_4} \rightarrow |b\oplus y\rangle_{\mathrm{B}_1}|y\rangle_{\mathrm{B}_4}$ coherently to obtain
$$|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\ \frac{1}{\sqrt{N}}\sum_{xy}|x\rangle_{\mathrm{A}_3}|y\rangle_{\mathrm{A}_4}|x\rangle_{\mathrm{B}_3}|y\rangle_{\mathrm{B}_4}\ |a\oplus x\rangle_{\mathrm{A}_1}|b\oplus y\rangle_{\mathrm{B}_1}\,. \tag{3.45}$$

3. Using U n times, they apply $\mathcal{P}_n$ to registers A_1 and B_1, obtaining an output state
$$|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\frac{1}{\sqrt{N}}\sum_{xy}|x\rangle_{\mathrm{A}_3}|y\rangle_{\mathrm{A}_4}|x\rangle_{\mathrm{B}_3}|y\rangle_{\mathrm{B}_4}\sum_{a',b'}|b'\oplus y\rangle_{\mathrm{A}_1}|a'\oplus x\rangle_{\mathrm{B}_1}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{3.46}$$

4. Alice decrypts her message in A_1 using her key A_4 and Bob decrypts B_1 using B_3 coherently as $|b'\oplus y\rangle_{\mathrm{A}_1}|y\rangle_{\mathrm{A}_4} \rightarrow |b'\rangle_{\mathrm{A}_1}|y\rangle_{\mathrm{A}_4}$ and $|a'\oplus x\rangle_{\mathrm{B}_1}|x\rangle_{\mathrm{B}_3} \rightarrow |a'\rangle_{\mathrm{B}_1}|x\rangle_{\mathrm{B}_3}$ producing a state
$$|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\frac{1}{\sqrt{N}}\sum_{xy}|x\rangle_{\mathrm{A}_3}|y\rangle_{\mathrm{A}_4}|x\rangle_{\mathrm{B}_3}|y\rangle_{\mathrm{B}_4}\sum_{a',b'}|b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{3.47}$$

5. Further CNOTs $\mathrm{A}_1 \rightarrow \mathrm{A}_4$, $\mathrm{A}_0 \rightarrow \mathrm{A}_3$, $\mathrm{B}_1 \rightarrow \mathrm{B}_3$ and $\mathrm{B}_0 \rightarrow \mathrm{B}_4$ will leave $\mathrm{A}_{2,3,4}$ and $\mathrm{B}_{2,3,4}$ almost decoupled from the classical messages. To see this, the state has become
$$\begin{aligned}
&|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\sum_{a',b'}|b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}\frac{1}{\sqrt{N}}\sum_{xy}|a\oplus x\rangle_{\mathrm{A}_3}|a'\oplus x\rangle_{\mathrm{B}_3}|b'\oplus y\rangle_{\mathrm{A}_4}|b\oplus y\rangle_{\mathrm{B}_4}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\\
=\ &|a\rangle_{\mathrm{A}_0}|b\rangle_{\mathrm{B}_0}\sum_{a',b'}|b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}\ |\Gamma_{a\oplus a',b\oplus b'}\rangle_{\mathrm{A}_{2,3,4}\,\mathrm{B}_{2,3,4}}\,,
\end{aligned} \tag{3.48}$$
where
$$|\Gamma_{a\oplus a',b\oplus b'}\rangle_{\mathrm{A}_{2,3,4}\,\mathrm{B}_{2,3,4}} := \frac{1}{\sqrt{N}}\sum_{xy}|a\oplus x\rangle_{\mathrm{A}_3}|a'\oplus x\rangle_{\mathrm{B}_3}|b'\oplus y\rangle_{\mathrm{A}_4}|b\oplus y\rangle_{\mathrm{B}_4}|\gamma^{a\oplus x,b\oplus y}_{a'\oplus x,b'\oplus y}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{3.49}$$

The fact $|\Gamma_{a\oplus a',b\oplus b'}\rangle$ depends only on $a\oplus a'$ and $b\oplus b'$, without any other dependence on a and b, can be easily seen by replacing x, y with $a\oplus x, b\oplus y$ in $\sum_{xy}$ in the RHS of the above. Note that $\langle\Gamma_{a\oplus a',b\oplus b'}|\Gamma_{a\oplus a',b\oplus b'}\rangle = \frac{1}{N}\sum_{xy}\Pr(a'\oplus x, b'\oplus y\,|\,a\oplus x, b\oplus y)$, so in particular for the state corresponding to the error-free term, we have $\langle\Gamma_{00}|\Gamma_{00}\rangle = \frac{1}{N}\sum_{xy}\Pr(xy|xy) := 1-\overline{\epsilon}_n \geq 1-\epsilon_n$.*

*Thus it turns out that Eq. (3.42) was more than we needed; the *average* error (over all a, b) would have been sufficient. In general, this argument shows that using shared entanglement (or randomness in the case of classical communication) can convert an average error condition into a maximum error condition, and will be further developed in [DW05b].

Suppose that Alice and Bob could project onto the space where $a' = a$ and $b' = b$, and tell each other they have succeeded (by using a little extra communication); then the resulting ancilla state $\frac{1}{\sqrt{1-\bar{\epsilon}_n}}|\Gamma_{00}\rangle$ has at least $C_1^{(n)}+C_2^{(n)}+\log(1-\epsilon_n)$ ebits, since its largest Schmidt coefficient is $\leq \big[\, \exp(C_1^{(n)}+C_2^{(n)})(1-\bar{\epsilon}_n) \,\big]^{-1/2}$ and $\bar{\epsilon}_n \leq \epsilon_n$ (cf. Proposition 2.10). Furthermore, $|\Gamma_{00}\rangle$ is manifestly independent of a, b. We will see how to improve the probability of successful projection onto the error free subspace by using block codes for error correction, and how correct copies of $|\Gamma_{00}\rangle$ can be identified if Alice and Bob can exchange a small amount of information.

Main idea on how to perform error correction

As discussed before, $|\Gamma_{00}\rangle$ cannot be used directly as an encryption key – our use of entanglement in $\mathcal{P}_n'$ is not catalytic. Entanglement concentration of many copies of $|\Gamma_{00}\rangle$ obtained from many runs of $\mathcal{P}_n'$ will make the entanglement overhead for the one-time-pad negligible, but errors will accumulate. The idea is to suppress the errors in many uses of $\mathcal{P}_n'$ by error correction. This has to be done with care, since we need to simultaneously ensure low enough error rates in both the classical message and the state to be concentrated, as well as sufficient decoupling of the classical messages from other systems.

Our error-corrected scheme will have k parallel uses of $\mathcal{P}_n'$, but the k inputs are chosen to be a valid codeword of an error correcting code. Furthermore, for each use of $\mathcal{P}_n'$, the state in $\mathrm{A}_{2,3,4}\,\mathrm{B}_{2,3,4}$ will only be collected for entanglement concentration if the error syndrome is trivial for that use of $\mathcal{P}_n'$. We use the fact that errors occur rarely (at a rate of ϵ_n, which goes to zero as $n \to \infty$) to show that (1) most states are still used for concentration, and (2) communicating the indices of the states with non trivial error syndrome requires a negligible amount of communication.

Definition of $\mathcal{P}_{nk}''$: error corrected version of $(\mathcal{P}_n')^{\otimes k}$ with entanglement concentration

We construct two codes, one used by Alice to signal to Bob and one from Bob to Alice. We consider high distance codes. The distance of a code is the minimum Hamming distance between any two codewords, i.e. the number of positions in which they are different.

First consider the code used by Alice. Let $N_1 = 2^{C_1^{(n)}}$. Alice is coding for a channel that takes input symbols from $[N_1] := \{1, \ldots, N_1\}$ and has probability $\leq \epsilon_n$ of error on any input (the error rate depends on both a and b). We would like to encode $[N_1]^l$ in $[N_1]^k$ using a code with distance $2k\alpha_n$, where α_n is a parameter that will be chosen later. Such a code can correct up to any $\lfloor k\alpha_n - \frac{1}{2} \rfloor$ errors (without causing much problem, we just say that the code corrects $k\alpha_n$ errors). Using standard arguments*, we can construct such a code with $l \geq k\,\big[\, 1-2\alpha_n - H_2(2\alpha_n)/C_1^{(n)} \,\big]$, where $H_2(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy. The code used by Bob is chosen similarly, with $N_2 = 2^{C_2^{(n)}}$ input symbols to each use of $\mathcal{P}_n'$. For simplicity, Alice's and Bob's codes share the same values of l, k and α_n. We choose $\alpha_n \geq \max(1/C_1^{(n)}, 1/C_2^{(n)})$ so that $l \geq k(1-3\alpha_n)$.

Furthermore, we want the probability of having $\geq k\alpha_n$ errors to be vanishingly small. This probability is $\leq \exp(-kD(\alpha_n \| \epsilon_n)) \leq \exp(k + k\alpha_n \log \epsilon_n)$ (using arguments from [CT91]) $\leq \exp(-k)$ if $\alpha_n \geq -2/\log \epsilon_n$.

Using these codes, Alice and Bob construct $\mathcal{P}_{nk}''$ as follows (with steps 1-3 performed coherently).

0. Let $(a_1^{\circ}, \cdots, a_l^{\circ})$ be a vector of l messages each of $C_1^{(n)}$ bits, and $(b_1^{\circ}, \cdots, b_l^{\circ})$ be l messages each of $C_2^{(n)}$ bits.

*We show the existence of a maximal code by repeatedly adding new codewords that have distance $\geq 2k\alpha_n$ from all other chosen codewords. This gives at least $N^k/\operatorname{Vol}(N, 2k\alpha_n, k)$ codewords, where $\operatorname{Vol}(N, k\delta, k)$ is the number of words in $[N]^k$ within a distance $k\delta$ of a fixed codeword. But $\operatorname{Vol}(N, k\delta, k) \leq \binom{k}{k\delta} N^{k\delta} \leq 2^{kH_2(\delta)} N^{k\delta}$. (See [CT91] or Eq. (6.4) later in this thesis for a derivation of $\binom{k}{k\delta} \leq 2^{kH_2(\delta)}$.) Altogether, the number of codewords $:= N^l \geq N^k/(2^{kH_2(2\alpha_n)} N^{2k\alpha_n})$, thus $l \geq k\,\big[\, 1 - 2\alpha_n - \frac{H_2(2\alpha_n)}{\log N} \,\big]$.

1. Using her error correcting code, Alice encodes $(a_1^\circ, \cdots, a_l^\circ)$ in a valid codeword $\vec{a} = (a_1, \cdots, a_k)$ which is a k-vector. Similarly, Bob generates a valid codeword $\vec{b} = (b_1, \cdots, b_k)$ using his code.

2. Let $\vec{\mathrm{A}}_1 := \mathrm{A}_1^{\otimes k}$ denote a tensor product of k input spaces each of $C_1^{(n)}$ qubits. Similarly, $\vec{\mathrm{B}}_1 := \mathrm{B}_1^{\otimes k}$. (We will also denote k copies of $\mathrm{A}_{0,2,3,4}$, and $\mathrm{B}_{0,2,3,4}$ by adding the vector symbol.) Alice and Bob apply $(\mathcal{P}'_n)^{\otimes k}$ to $|\vec{a}\rangle_{\vec{\mathrm{A}}_1}|\vec{b}\rangle_{\vec{\mathrm{B}}_1}$; that is, in parallel, they apply $\mathcal{P}'_n$ to each pair of inputs (a_j, b_j). The resulting state is a tensor product of states of the form given by Eq. (3.48):
$$\bigotimes_{j=1}^{k} \Big[\ |a_j\rangle_{\mathrm{A}_0}|b_j\rangle_{\mathrm{B}_0} \sum_{a'_j, b'_j} |b'_j\rangle_{\mathrm{A}_1}|a'_j\rangle_{\mathrm{B}_1}\ |\Gamma_{a_j\oplus a'_j, b_j \oplus b'_j}\rangle_{\mathrm{A}_{2,3,4}\,\mathrm{B}_{2,3,4}}\ \Big]. \tag{3.50}$$
Define $|\Gamma_{\vec{a}\oplus\vec{a}', \vec{b}\oplus\vec{b}'}\rangle_{\vec{\mathrm{A}}_{234}\vec{\mathrm{B}}_{234}} := \bigotimes_{j=1}^{k} |\Gamma_{a_j\oplus a'_j, b_j\oplus b'_j}\rangle_{\mathrm{A}_{2,3,4}\,\mathrm{B}_{2,3,4}}$. Then, Eq. (3.50) can be written more succinctly as
$$|\vec{a}\rangle_{\vec{\mathrm{A}}_0}|\vec{b}\rangle_{\vec{\mathrm{B}}_0} \sum_{\vec{a}', \vec{b}'} |\vec{b}'\rangle_{\vec{\mathrm{A}}_1}|\vec{a}'\rangle_{\vec{\mathrm{B}}_1}|\Gamma_{\vec{a}\oplus\vec{a}', \vec{b}\oplus\vec{b}'}\rangle_{\vec{\mathrm{A}}_{234}\vec{\mathrm{B}}_{234}}\,. \tag{3.51}$$

3. Alice performs the error correction step on $\vec{\mathrm{A}}_1$ and Bob does the same on $\vec{\mathrm{B}}_1$. According to our code constructions, this (joint) step fails with probability $p_{\mathrm{fail}} \leq 2 \cdot 2^{-k}$. (We will see below why p_{fail} is independent of $\vec{a}$ and $\vec{b}$.)

   In order to describe the residual state, we now introduce $\mathcal{G}_{\mathrm{A}} = \{\vec{x} \in [N_1]^k : |\vec{x}| \leq k\alpha_n\}$ and $\mathcal{G}_{\mathrm{B}} = \{\vec{x} \in [N_2]^k : |\vec{x}| \leq k\alpha_n\}$, where $|\vec{x}| := |\{j : x_j \neq 0\}|$ denotes the Hamming weight of $\vec{x}$. Thus $\mathcal{G}_{\mathrm{A,B}}$ are sets of correctable (good) errors, in the sense that there exist local decoding isometries $\mathcal{D}_{\mathrm{A}}, \mathcal{D}_{\mathrm{B}}$ such that for any code word $\vec{a} \in [N_1]^k$ we have $\forall \vec{a}' \in \vec{a} \oplus \mathcal{G}_{\mathrm{A}}, \mathcal{D}_{\mathrm{A}}|\vec{a}'\rangle = |\vec{a}\rangle|\vec{a} \oplus \vec{a}'\rangle$ (and similarly, if $\vec{b} \in [N_2]^k$ is a codeword, then $\forall \vec{b}' \in \vec{b} \oplus \mathcal{G}_{\mathrm{B}}, \mathcal{D}_{\mathrm{B}}|\vec{b}'\rangle = |\vec{b}\rangle|\vec{b} \oplus \vec{b}'\rangle$). For concreteness, let the decoding maps take $\vec{\mathrm{A}}_1$ to $\vec{\mathrm{A}}_1\vec{\mathrm{A}}_5$ and $\vec{\mathrm{B}}_1$ to $\vec{\mathrm{B}}_1\vec{\mathrm{B}}_5$.

   Conditioned on success, Alice and Bob are left with
$$\frac{1}{\sqrt{1-p_{\mathrm{fail}}}}\,|\vec{a},\vec{b}\rangle_{\vec{\mathrm{A}}_{0,1}}|\vec{a},\vec{b}\rangle_{\vec{\mathrm{B}}_{0,1}} \sum_{\vec{a}'\in\vec{a}\oplus\mathcal{G}_{\mathrm{A}}}\ \sum_{\vec{b}'\in\vec{b}\oplus\mathcal{G}_{\mathrm{B}}} |\vec{b}\oplus\vec{b}'\rangle_{\vec{\mathrm{A}}_5}|\vec{a}\oplus\vec{a}'\rangle_{\vec{\mathrm{B}}_5}|\Gamma_{\vec{a}\oplus\vec{a}', \vec{b}\oplus\vec{b}'}\rangle_{\vec{\mathrm{A}}_{234}\vec{\mathrm{B}}_{234}} \tag{3.52}$$
$$:= \frac{1}{\sqrt{1-p_{\mathrm{fail}}}}\,|\vec{a},\vec{b}\rangle_{\vec{\mathrm{A}}_{0,1}}|\vec{a},\vec{b}\rangle_{\vec{\mathrm{B}}_{0,1}} \sum_{\vec{a}''\in\mathcal{G}_{\mathrm{A}}}\ \sum_{\vec{b}''\in\mathcal{G}_{\mathrm{B}}} |\vec{b}''\rangle_{\vec{\mathrm{A}}_5}|\vec{a}''\rangle_{\vec{\mathrm{B}}_5}|\Gamma_{\vec{a}'', \vec{b}''}\rangle_{\vec{\mathrm{A}}_{234}\vec{\mathrm{B}}_{234}}, \tag{3.53}$$
   where we have defined $\vec{a}'' := \vec{a} \oplus \vec{a}'$ and $\vec{b}'' := \vec{b} \oplus \vec{b}'$. Note that $2^{-k+1} \geq p_{\mathrm{fail}} = \sum_{(\vec{a}'', \vec{b}'') \notin \mathcal{G}_{\mathrm{A}} \times \mathcal{G}_{\mathrm{B}}} \langle \Gamma_{\vec{a}'', \vec{b}''} | \Gamma_{\vec{a}'', \vec{b}''} \rangle$, which is manifestly independent of $\vec{a}, \vec{b}$. The ancilla is now *completely* decoupled from the message, resulting in coherent classical communication. The only remaining issue is recovering entanglement from the ancilla, so for the remainder of the protocol we ignore the now decoupled states $|\vec{a},\vec{b}\rangle_{\vec{\mathrm{A}}_{0,1}}|\vec{a},\vec{b}\rangle_{\vec{\mathrm{B}}_{0,1}}$.

4. For any $\vec{x}$, define $S(\vec{x}) := \{j : x_j \neq 0\}$ to be set of positions where $\vec{x}$ is nonzero. If $\vec{x} \in \mathcal{G}_{\mathrm{A}}$ (or $\mathcal{G}_{\mathrm{B}}$), then $|S(\vec{x})| \leq k\alpha_n$. Thus, $S(\vec{x})$ can be written using $\leq \log \sum_{j \leq k\alpha_n} \binom{k}{j} \leq \log \binom{k}{k\alpha_n} + \log(k\alpha_n) \leq kH_2(\alpha_n) + \log(k\alpha_n)$ bits.

   The next step is for Alice to compute $|S(\vec{b}'')\rangle$ from $|\vec{b}''\rangle$ and communicate it to Bob using $\big(kH_2(\alpha_n) + \log(k\alpha_n)\big)\,[c \to c]$. Similarly, Bob sends $|S(\vec{a}'')\rangle$ to Alice using $\big(kH_2(\alpha_n) + \log(k\alpha_n)\big)\,[c \leftarrow c]$. Here we need to assume that some (possibly inefficient) protocol to send $O(k)$ bits in either direction with error $\exp(-k-1)$ (chosen for convenience) and with Rk uses of U for some constant R. Such a protocol was given by Proposition 2.5 and the bound on the error can be obtained from the HSW theorem[Hol98, SW97, HN03].

Alice and Bob now have the state

$$\frac{1}{\sqrt{1-p_{\text{fail}}}} \sum_{\vec{a}''\in\mathcal{G}_{\text{A}}} \sum_{\vec{b}''\in\mathcal{G}_{\text{B}}} |S(\vec{a}'')S(\vec{b}'')\rangle_{\vec{\text{A}}_6}\, |\vec{b}''\rangle_{\vec{\text{A}}_5}\, |S(\vec{a}'')S(\vec{b}'')\rangle_{\vec{\text{B}}_6}\, |\vec{a}''\rangle_{\vec{\text{B}}_5}\, |\Gamma_{\vec{a}'',\vec{b}''}\rangle_{\vec{\text{A}}_{234}\vec{\text{B}}_{234}}. \tag{3.54}$$

Conditioning on their knowledge of $S(\vec{a}''), S(\vec{b}'')$, Alice and Bob can now identify $k' \geq k(1-2\alpha_n)$ positions where $a''_j = b''_j = 0$, and extract k' copies of $\frac{1}{\sqrt{1-p_{\text{fail}}}}|\Gamma_{00}\rangle$. Note that leaking $S(\vec{a}''), S(\vec{b}'')$ to the environment will not affect the extraction procedure, therefore, coherent computation and communication of $S(\vec{a}''), S(\vec{b}'')$ is unnecessary. (We have not explicitly included the environment's copy of $|S(\vec{a}'')S(\vec{b}'')\rangle$ in the equations to minimize clutter.) After extracting k' copies of $\frac{1}{\sqrt{1-p_{\text{fail}}}}|\Gamma_{00}\rangle$, we can safely discard the remainder of the state, which is now completely decoupled from both $\left[\frac{1}{\sqrt{1-p_{\text{fail}}}}|\Gamma_{00}\rangle\right]^{\otimes k'}$ and the message $|\vec{a}\rangle_{\text{A}_0}|\vec{b}\rangle_{\text{A}_1}|\vec{b}\rangle_{\text{B}_0}|\vec{a}\rangle_{\text{B}_1}$.

5. Alice and Bob perform entanglement concentration $\mathcal{E}_{\text{conc}}$ (using the techniques of [BBPS96]) on $\left[\frac{1}{\sqrt{1-p_{\text{fail}}}}|\Gamma_{00}\rangle\right]^{\otimes k'}$. Note that since $\frac{1}{\sqrt{1-p_{\text{fail}}}}|\Gamma_{00}\rangle$ can be created using U n times and then using classical communication and postselection, it must have Schmidt rank $\leq \text{Sch}(U)^n$, where $\text{Sch}(U)$ is the Schmidt number of the gate U. Also recall that $E\left[\frac{1}{\sqrt{1-p_{\text{fail}}}}|\Gamma_{00}\rangle\right] \geq C_1^{(n)} + C_2^{(n)} + \log(1-\epsilon_n)$. According to [BBPS96], $\mathcal{E}_{\text{conc}}$ requires no communication and with probability $\geq 1 - \exp\left[-\text{Sch}(U)^n\left(\sqrt{k'} - \log(k'+1)\right)\right]$ produces at least $k'\left[C_1^{(n)}+C_2^{(n)}+\log(1-\epsilon_n)\right] - \text{Sch}(U)^n\left[\sqrt{k'} - \log(k'+1)\right]$ ebits.

Error and resource accounting

$\mathcal{P}''_{nk}$ consumes a total of

(0) nk uses of U (in the k executions of $\mathcal{P}'_n$)
(1) Rk uses of U (for communicating nontrivial syndrome locations)
(2) $k\left[C_1^{(n)}+C_2^{(n)}\right][q\,q]$ (for the encryption of classical messages).

$\mathcal{P}''_{nk}$ produces, with probability and fidelity no less than

$$1-2^{-(k-1)}-2^{-(k-1)} - \exp\left[-\text{Sch}(U)^n\left(\sqrt{k'} - \log(k'+1)\right)\right],$$

at least

(1) $l\,C_1^{(n)}[\![c \to c]\!] + l\,C_2^{(n)}[\![c \leftarrow c]\!]$
(2) $k'\left(C_1^{(n)}+C_2^{(n)}+\log(1-\epsilon_n)\right) - \text{Sch}(U)^n\left(\sqrt{k'} - \log(k'+1)\right)[q\,q]$.

We restate the constraints on the above parameters: $\epsilon_n, \delta_n \to 0$ as $n \to \infty$; $C_1^{(n)} \geq n(C_1-\delta_n)$, $C_2^{(n)} \geq n(C_2-\delta_n)$; $\alpha_n \geq \max(1/C_1^{(n)}, 1/C_2^{(n)}, -2/\log\epsilon_n)$; $k' \geq k(1-2\alpha_n)$; $l \geq k(1-3\alpha_n)$.

We define "error" to include both infidelity and the probability of failure. To leading orders of k, n, this is equal to $2^{-(k-2)} + \exp\left[-\sqrt{k}\ \text{Sch}(U)^n\right]$. We define "inefficiency" to include extra uses of U, net consumption of entanglement, and the amount by which the coherent classical communication rates fall short of the classical capacities. To leading order of k, n, these are respectively Rk, $2\alpha_n k(C_1^{(n)}+C_2^{(n)}) + \sqrt{k}\,\text{Sch}(U)^n \approx 2\alpha_n kn(C_1+C_2) + \sqrt{k}\,\text{Sch}(U)^n$, and $nk(C_1+C_2) - l(C_1^{(n)}+C_2^{(n)}) \leq nk(3\alpha_n(C_1+C_2) + 2\delta_n)$. We would like the error to vanish, as well as the fractional inefficiency, defined as the inefficiency divided by kn, the number of uses of U. Equivalently, we can define $f(k,n)$ to be the *sum* of the error and the fractional inefficiency, and require that $f(k,n) \to 0$ as $nk \to \infty$.

By the above arguments,

$$f(k,n) \leq 2^{-(k-2)} + \exp(-\sqrt{k}\,\mathrm{Sch}(U)^n) + 2\alpha_n(C_1{+}C_2) + \tfrac{1}{n\sqrt{k}}\,\mathrm{Sch}(U)^n + \frac{R}{n} + 3\alpha_n(C_1{+}C_2) + 2\delta_n\,. \tag{3.55}$$

Note that for any fixed value of n, $\lim_{k\to\infty} f(k,n) = 5\alpha_n(C_1{+}C_2) + 2\delta_n + R/n$. (This requires k to be sufficiently large and also $k \gg \mathrm{Sch}(U)^{2n}$.) Now, allowing n to grow, we have

$$\lim_{n\to\infty}\lim_{k\to\infty} f(k,n) = 0. \tag{3.56}$$

The order of limits in this equation is crucial due to the dependence of k on n.

The only remaining problem is our catalytic use of $O(nk)$ ebits. In order to construct a protocol that uses only U, we need to first use U $O(nk)$ times to generate the starting entanglement. Then we repeat $\mathcal{P}''_n$ m times, reusing the same entanglement. The catalyst results in an additional fractional inefficiency of c/m (for some constant c depending only on U) and the errors and inefficiencies of $\mathcal{P}''_n$ add up to no more than $mf(k,n)$. Choosing $m = \lfloor 1/\sqrt{f(k,n)}\rfloor$ will cause all of these errors and inefficiencies to simultaneously vanish. (This technique is essentially equivalent to using Lemmas 1.36 and 1.37 and Theorem 1.22.) The actual error condition is that

$$\lim_{m\to\infty}\lim_{n\to\infty}\lim_{k\to\infty}\; mf(k,n) + \frac{c}{m} \;=\; 0\,. \tag{3.57}$$

This proves the resource inequality

$$U \geq C_1[\![c \to c]\!] + C_2[\![c \leftarrow c]\!]. \tag{3.58}$$

The $E<0$ and $E>0$ cases

If $E<0$ then entanglement is consumed in $\mathcal{P}_n$, so there exists a sequence of integers $E^{(n)} \leq n(E{+}\delta_n)$ such that

$$\mathcal{P}_n\Big(|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1}|\Phi\rangle^{E^{(n)}}_{\mathrm{A}_5\,\mathrm{B}_5}\Big) = \sum_{a',b'} |b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}|\gamma^{a,b}_{a',b'}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{3.59}$$

In this case, the analysis for $E^{(n)}=0$ goes through, only with additional entanglement consumed. Almost all equations are the same, except now the Schmidt rank for $|\Gamma_{00}\rangle$ is upper-bounded by $\left[\mathrm{Sch}(U)2^{E+\delta_n}\right]^n$ instead of $\mathrm{Sch}(U)^n$. This is still $\leq c^n$ for some constant c, so the same proof of correctness applies.

If instead $E>0$, entanglement is created, so for some $E^{(n)} \geq n(E-\delta_n)$ we have

$$\mathcal{P}_n(|a\rangle_{\mathrm{A}_1}|b\rangle_{\mathrm{B}_1}) = \sum_{a',b'} |b'\rangle_{\mathrm{A}_1}|a'\rangle_{\mathrm{B}_1}|\gamma^{a,b}_{a',b'}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}\,. \tag{3.60}$$

for $E(|\gamma^{a,b}_{a,b}\rangle_{\mathrm{A}_2\,\mathrm{B}_2}) \geq E^{(n)}$. Again, the previous construction and analysis go through, with an extra $E^{(n)}$ ebits of entanglement of entropy in $|\Gamma_{00}\rangle$, and thus an extra fractional efficiency of $\leq 2\alpha_n E$ in Eq. (3.55). The Schmidt rank of $|\Gamma_{00}\rangle$ is still upper bounded by $\mathrm{Sch}(U)^n$ in this case. □

Observation 3.10. *If $(C_1,C_2,E) \in \mathrm{CCE}(U)$, but $(C_1,C_2,E+\delta) \notin \mathrm{CCE}(U)$ for any $\delta>0$, then for any $\epsilon,\delta>0$ and for n sufficiently large there is a protocol $\mathcal{P}_n$ and a state $|\varphi\rangle^{AB}$ on $\leq \kappa n\delta$ qubits (for a universal constant κ), such that for any $x\in\{0,1\}^{\lfloor n(C_1-\delta)\rfloor}, y\in\{0,1\}^{\lfloor n(C_2-\delta)\rfloor}$ we have either*

$$\mathcal{P}_n|x\rangle^A|y\rangle^B \approx_\epsilon |xy\rangle^A|xy\rangle^B|\Phi\rangle^{\lfloor n(E-\delta)\rfloor}|\varphi\rangle$$

if $E>0$ or

$$\mathcal{P}_n|x\rangle^A|y\rangle^B|\Phi\rangle^{\lfloor -n(E-\delta)\rfloor} \approx_\epsilon |xy\rangle^A|xy\rangle^B|\varphi\rangle$$

if $E < 0$.

The key point here is that if E *taken to be the maximum possible for a given* C_1, C_2*, then the above proof of Theorem 3.1 in fact produces ancilla systems of a sublinear size.*

3.6 Discussion

Quantum information, like quantum computing, has often been studied under an implicit "quantum co-processor" model, in which quantum resources are used by some controlling classical computer. Thus, we might use quantum computers or quantum channels to perform classical tasks, like solving computational problems, encrypting or authenticating a classical message, demonstrating nonlocal classical correlations, synchronizing classical clocks and so on. On the other hand, since the quantum resources are manipulated by a classical computer, it is natural to think of conditioning quantum logical operations on classical information.

This framework has been quite useful for showing the strengths of quantum information relative to classical information processing techniques; e.g. we find that secure communication is possible, distributed computations require less communication and so on. However, in quantum Shannon theory, it is easy to be misled by the central role of classical information in the quantum co-processor model. While classical communication may still be a useful *goal* of quantum Shannon theory, it is often inappropriate as an intermediate step. Rather, we find in protocol after protocol that coherently decoupled cbits are better thought of as cobits.

Replacing cbits with cobits has significance beyond merely improving the efficiency of quantum protocols. In many cases, cobits give rise to asymptotically reversible protocols, such as coherent teleportation and super-dense coding, or more interestingly, remote state preparation and HSW coding. The resulting resource equalities go a long way towards simplifying the landscape of quantum Shannon theory: (1) The duality of teleportation and super-dense coding resolves a long-standing open question about how the original forms of these protocols could be individually optimal, but wasteful when composed; we now know that all the irreversibility from composing teleportation and super-dense coding is due to the map $[\![c \to c]\!] \geq [c \to c]$. (2) Coherent RSP and HSW coding give a resource equality that allows us to easily derive an expression for unitary gate capacity regions. In the next chapter, we will see more examples of how making classical communication coherent leads to a wide variety of optimal coding theorems.

Although the implications of coherent classical communication are wide-ranging, the fundamental insight is quite simple: when studying quantum Shannon theory, we should set aside our intuition about the central role of classical communication, and instead examine carefully which systems are discarded and when communication can be coherently decoupled.

Chapter 4

Optimal trade-offs in quantum Shannon theory

The main purpose of quantum information theory, or more particularly *quantum Shannon theory*, is to characterize asymptotic resource inter-conversion tasks in terms of quantum information theoretical quantities such as von Neumann entropy, quantum mutual and coherent informations. A particularly important class of problems involves a noisy quantum channel or shared noisy entanglement between two parties which is to be converted into qubits, ebits and/or cbits, possibly assisted by limited use of qubits, ebits or cbits as an auxiliary resource. In this final chapter on quantum Shannon theory, we give a full solution for this class of problems.

In Section 4.1, we will state two dual, purely quantum protocols: for entanglement distillation assisted by quantum communication (the "mother" protocol) and for entanglement assisted quantum communication (the "father" protocol). From these two, we can derive a large class of "children" (including many previously known resource inequalities) by direct application of teleportation or super-dense coding. The key ingredient to deriving the parents, and thus obtaining the entire family, is coherent classical communication. Specifically, we will show how the parents can be obtained by applying Rules I and O to many of the previously known children. In each scenario, we will find that previous proofs of the children already use coherently decoupled cbits (or can be trivially modified to do so), so that the only missing ingredient is coherent classical communication.

Next, we address the question of optimality. Most of the protocols we involve one noisy resource (such as $\langle\mathcal{N}\rangle$) and two noiseless standard ones (such as qubits and ebits), so instead of capacities we need to work with two-dimensional capacity regions whose boundaries determine trade-off curves. We state and prove formulae for each of these capacity regions in Section 4.2.

Finally we give some ideas for improving these results in Section 4.3.

Bibliographical note: Most of the chapter is based on [DHW05], though parts of Section 4.1 appeared before in [DHW04]. Both are joint work with Igor Devetak and Andreas Winter.

4.1 A family of quantum protocols.

In this section, we consider a family of resource inequalities with one noisy resource in the input and two noiseless resources in either the input or the output. The "static" members of the family involve a noisy bipartite state ρ^{AB}, while the "dynamic" members involve a general quantum channel $\mathcal{N} : A' \to B$. In the former case one may define a class of purifications $|\psi\rangle\langle\psi|^{ABE} \supseteq \rho^{AB}$. In the latter case one may define a class of pure states $|\psi\rangle^{RBE}$, which corresponds to the outcome of sending half of some $|\phi\rangle^{RA'}$ through the channel's isometric extension $U_{\mathcal{N}} : A' \to BE$, $U_{\mathcal{N}} \supseteq \mathcal{N}$.

Recall the identities, for a tripartite pure state $|\psi\rangle^{ABE}$,

$$\begin{aligned}\frac{1}{2}I(A;B)_\psi + \frac{1}{2}I(A;E)_\psi &= H(A)_\psi,\\ \frac{1}{2}I(A;B)_\psi - \frac{1}{2}I(A;E)_\psi &= I(A\,\rangle B)_\psi.\end{aligned}$$

Henceforth, all entropic quantities will be defined with respect to $|\psi\rangle^{RBE}$ or $|\psi\rangle^{ABE}$, depending on the context, so we shall drop the ψ subscript.

We now introduce the "parent" resource inequalities, deferring their construction until the end of the section. The "mother" RI is a method for distillating entanglement from a noisy state using quantum communication:

$$\langle\rho\rangle + \frac{1}{2}I(A;E)\,[q\to q] \geq \frac{1}{2}I(A;B)\,[q\,q]. \tag{♀}$$

There exists a dual "father" RI for entanglement-assisted quantum communication, which is related to the mother by interchanging dynamic and static resources, and the A and R systems:

$$\frac{1}{2}I(R;E)\,[q\,q] + \langle\mathcal{N}\rangle \geq \frac{1}{2}I(R;B)\,[q\to q]. \tag{♂}$$

We shall combine these parent RIs with the unit RIs corresponding to teleportation, super-dense coding and entanglement distribution ($[q\to q]\geq[q\,q]$) to recover several previously known "children" protocols.

Each parent has her or his own children (like the Brady Bunch*).

Let us consider the mother first; she has three children. The first is a variation of the hashing inequality Eq. (1.49), which follows from the mother and teleportation.

$$\begin{aligned}\langle\rho\rangle + I(A;E)\,[c\to c] + \frac{1}{2}I(A;E)[q\,q] &\geq \langle\rho\rangle + \frac{1}{2}I(A;E)[q\to q]\\ &\geq \frac{1}{2}I(A;B)[q\,q]\\ &= I(A\,\rangle B)\,[q\,q] + \frac{1}{2}I(A;E)[q\,q].\end{aligned}$$

By the Cancellation Lemma (1.37),

$$\langle\rho\rangle + I(A;E)\,[c\to c] + o[q\,q] \geq I(A\,\rangle B)\,[q\,q]. \tag{4.1}$$

This is slightly weaker than Eq. (1.49). Further combining with teleportation gives a variation on noisy teleportation Eq. (1.50):

$$\langle\rho\rangle + I(A;B)\,[c\to c] + o[q\,q] \geq I(A\,\rangle B)\,[q\to q]. \tag{4.2}$$

The third child is noisy super-dense coding (Eq. (1.48)), obtained by combining the mother with super-dense coding:

$$\begin{aligned}H(A)\,[q\to q] + \langle\rho\rangle &= \frac{1}{2}I(A;B)\,[q\to q] + \frac{1}{2}I(A;E)\,[q\to q] + \langle\rho\rangle\\ &\geq \frac{1}{2}I(A;B)[q\to q] + \frac{1}{2}I(A;B)[q\,q]\\ &\geq I(A;B)\,[c\to c].\end{aligned}$$

[*] *The Brady Bunch*, running from 26 September 1969 till 8 March 1974, was a popular show of the American Broadcasting Company about a couple with three children each from their previous marriages. For more information, see [Mor95].

The father happens to have only two children. One of them is the entanglement-assisted classical capacity RI (1.46), obtained by combining the father with (SD)

$$\begin{aligned} H(R)\,[q\,q] + \langle \mathcal{N} \rangle &= \frac{1}{2} I(R;B)\,[q\,q] + \frac{1}{2} I(R;E)\,[q\,q] + \langle \mathcal{N} \rangle \\ &\geq \frac{1}{2} I(R;B)[q\,q] + \frac{1}{2} I(R;B)[q \to q] \\ &\geq I(R;B)\,[c \to c]. \end{aligned}$$

The second is a variation on the quantum channel capacity result (Eq. (1.47)). It is obtained by combining the father with entanglement distribution.

$$\begin{aligned} \frac{1}{2} I(R;E)\,[q\,q] + \langle \mathcal{N} \rangle &\geq \frac{1}{2} I(R;B)\,[q \to q] \\ &= \frac{1}{2} I(R;E)\,[q \to q] + \frac{1}{2} I(R\rangle B)\,[q \to q] \\ &= \frac{1}{2} I(R;E)\,[q\,q] + \frac{1}{2} I(R\rangle B)\,[q \to q]. \end{aligned}$$

Hence, by the Cancellation Lemma

$$\langle \mathcal{N} \rangle + o[q\,q] \geq I(R\rangle B)\,[q \to q]. \tag{4.3}$$

Alas, we do not know how to get rid of the o term without invoking further results. For instance, the original proof of the hashing inequality and the HSW theorem allow us to get rid of the o term, by Lemma 1.36. Quite possibly the original proof [Llo96, Sho02, Dev05a] is needed.

Constructing the parent protocols using coherification rules.

Having demonstrated the power of the parent resource inequalities, we now address the question of constructing protocols implementing them.

Corollary 4.1. *The mother RI is obtained from the hashing inequality (Eq. (1.49)) by applying rule I.*

It can be readily checked that the protocol from [DW05a, DW04] implementing Eq. (1.49) indeed satisfies the conditions of rule I. The approximate uniformity condition is in fact exact in this case. □

Corollary 4.2. *The father RI follows from the EAC protocol from [BSST02].*

Proof. The main observation is that the protocol from [BSST02] implementing Eq. (1.46) in fact outputs a *private* classical channel as it is! We shall analyze the protocol in the CP picture. Alice and Bob share a maximally entangled state $|\Phi_D\rangle^{AB'}$. Alice encodes her message m via a unitary U_m:

$$m \mapsto (U_m^A \otimes \mathbb{1}^{B'})|\Phi_D\rangle^{AB'} = (\mathbb{1}^A \otimes (U_m^T)^{B'})|\Phi_D\rangle^{AB'}.$$

Applying the channel $(U_{\mathcal{N}}^{A\to BE})^{\otimes n}$ yields

$$|\Upsilon_m\rangle^{BB'E} = ((U_m^T)^{B'} \otimes \mathbb{1}^{BE})|\Psi\rangle^{BB'E},$$

where $|\Psi\rangle^{BB'E} = (U_{\mathcal{N}}^{A\to BE})^{\otimes n}|\Phi_D\rangle^{AB'}$. Bob's decoding operation consists of adding an ancilla system $\overline{B}$ in the state $|0\rangle^{\overline{B}}$, performing some unitary $U^{BB'\overline{B}}$ and von Neumann measuring the ancilla $\overline{B}$. Before the von Neumann measurement the state of the total system is

$$|\Upsilon''_m\rangle^{BB'\overline{B}E} = U^{BB'\overline{B}}|\Upsilon_m\rangle^{BB'E}|0\rangle^{\overline{B}}.$$

After the measurement, the message m is correctly decoded with probability $1-\epsilon$. By the gentle operator lemma[Win99a], $U^{BB'\overline{B}}$ could have been chosen so that upon correct decoding, the post-measurement state $|\Upsilon'_m\rangle^{BB'E}$ satisfies

$$\|\Upsilon'_m - \Upsilon_m\|_1 \leq \sqrt{8\epsilon}.$$

Assuming Bob correctly decodes m, he then applies U^*_m to B', bringing the system $BB'E$ into the state $|\Psi'_m\rangle = ((U^*_m)^{B'} \otimes \mathbb{1}^{BE})|\Upsilon'_m\rangle$, for which

$$\|\Psi'_m - \Psi\|_1 \leq \sqrt{8\epsilon},$$

for all m. Thus m is coherently decoupled from $BB'E$, and we may apply Rule O. □

Corollary 4.3. *The mother RI follows from the NSD protocol from [HHH⁺01].*

Proof. The proof is almost the same as for the previous Corollary. □

4.2 Two dimensional trade-offs for the family

It is natural to ask about the optimality of our family of resource inequalities. In this section we show that they indeed give rise to optimal two dimensional capacity regions, the boundaries of which are referred to as trade-off curves. To each family member corresponds a theorem identifying the operationally defined capacity region $C(\rho^{AB})$ $(C(\mathcal{N}))$ with a formula $\widetilde{C}(\rho^{AB})$ $(\widetilde{C}(\mathcal{N}))$ given in terms of entropic quantities evaluated on states associated with the given noisy resource ρ^{AB} $(\mathcal{N})$. Each such theorem consists of two parts: the *direct coding theorem* which establishes $\widetilde{C} \subseteq C$ and the *converse* which establishes $C \subseteq \widetilde{C}$.

4.2.1 Grandmother protocol

To prove the trade-offs involving static resources, we will first need to extend the mother protocol (Eq. ♀) to a "grandmother" RI by combining it with instrument compression (Eq. 1.35).

Theorem 4.4 (Grandmother). *Given a static resource ρ^{AB}, for any remote instrument $\mathbf{T} : A \to A'X_B$, the following RI holds*

$$\frac{1}{2}I(A';EE'|X_B)_\sigma\,[q \to q] + I(X_B;BE)_\sigma[c \to c] + \langle\rho^{AB}\rangle \geq \frac{1}{2}I(A';B|X_B)_\sigma\,[q\,q]. \tag{4.4}$$

In the above, the state $\sigma^{X_BA'BEE'}$ is defined by

$$\sigma^{X_BA'BEE'} = \widetilde{\mathbf{T}}^{A\to A'E'X_B}(\psi^{ABE}),$$

where $|\psi\rangle\langle\psi|^{ABE} \supseteq \rho^{AB}$ and $\widetilde{\mathbf{T}} : A \to A'E'X_B$ is a QP extension of $\mathbf{T}$.

Proof. By the instrument compression RI (1.35),

$$\begin{array}{rcl}\langle\rho^{AB}\rangle + I(X_B;BE)_\sigma[c \to c] + H(X|BE)_\sigma[c\,c] & \geq & \langle\rho^{AB}\rangle + \langle\overline{\Delta}^{X_B\to X_AX_B} \circ \mathbf{T} : \rho^A\rangle \\ & \geq & \langle\overline{\Delta}^{X_B\to X_AX_B}(\sigma^{X_BA})\rangle.\end{array}$$

On the other hand, by Theorem 1.41 and the mother inequality (♀),

$$\langle\overline{\Delta}^{X_B\to X_AX_B}(\sigma^{X_BA'})\rangle + \frac{1}{2}I(A';EE'|X_B)_\sigma\,[q \to q] \geq \frac{1}{2}I(A';B|X_B)_\sigma\,[q\,q].$$

The grandmother RI is obtained by adding the above RIs, followed by a derandomization via Corollary 1.39. □

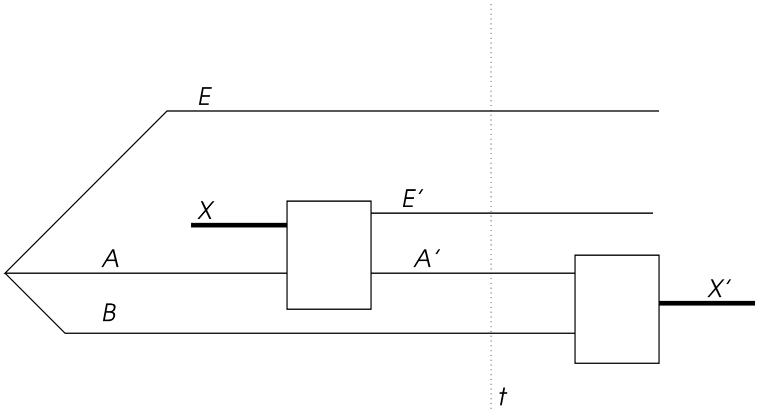


Figure 4-1: A general protocol for noisy super-dense coding.

Corollary 4.5. *In the above theorem, one may consider the special case where* $\mathbf{T} : A \to A'X_B$ *corresponds to some ensemble of operations* $(p_x, \mathcal{E}_x)$, $\mathcal{E}_x : A \to A'$, *via the identification*

$$\mathbf{T} : \rho^A \mapsto \sum_x p_x |x\rangle\langle x|^{X_B} \otimes \mathcal{E}_x(\rho^A).$$

Then the $[c \to c]$ *term from Eq. (4.4) vanishes identically.* □

4.2.2 Trade-off for noisy super-dense coding

Now that we are comfortable with the various formalisms, the formulae will reflect the QP formalism, whereas the language will be more in the CQ spirit.

Given a bipartite state ρ^{AB}, the noisy super-dense coding capacity region $C_{\mathrm{NSD}}(\rho^{AB})$ is the two-dimensional region in the (Q, R) plane with $Q \geq 0$ and $R \geq 0$ satisfying the RI

$$\langle \rho^{AB} \rangle + Q\,[q \to q] \geq R\,[c \to c]. \tag{4.5}$$

Theorem 4.6. *The capacity region* $C_{\mathrm{NSD}}(\rho^{AB})$ *is given by*

$$C_{\mathrm{NSD}}(\rho^{AB}) = \widetilde{C}_{\mathrm{NSD}}(\rho^{AB}) := \overline{\bigcup_{l=1}^{\infty} \frac{1}{l} \widetilde{C}^{(1)}_{\mathrm{NSD}}((\rho^{AB})^{\otimes l})},$$

where the $\overline{S}$ *means the closure of a set* S *and* $\widetilde{C}^{(1)}_{\mathrm{NSD}}(\rho^{AB})$ *is the set of all* $R \geq 0$, $Q \geq 0$ *such that*

$$R \leq Q + \max_{\sigma} \left\{ I(A'\rangle BX)_\sigma : H(A'|X)_\sigma \leq Q \right\}.$$

In the above, σ *is of the form*

$$\sigma^{XA'B} = \sum_x p_x |x\rangle\langle x|^X \otimes \mathcal{E}_x^{A \to A'}(\rho^{AB}). \tag{4.6}$$

for some ensemble of operations $(p_x, \mathcal{E}_x)$, $\mathcal{E}_x : A \to A'$.

Proof. We first prove the converse. Fix $n, R, Q, \delta, \epsilon$, and use the Flattening Lemma (1.17) so that we can assume that $k = 1$. The resources available are

- The state $(\rho^{AB})^{\otimes n}$ shared between Alice and Bob. Let it be contained in the system $A^n B^n$, of total dimension d^n, which we shall call AB for short.

- A perfect quantum channel $\mathrm{id} : A' \to A'$, $\dim A' = 2^{nQ}$, from Alice to Bob (after which A' belongs to Bob despite the notation!).

The resource to be simulated is the perfect classical channel of size $D = 2^{n(R-\delta)}$ on any source, in particular on the random variable X corresponding to the uniform distribution π_D.

In the protocol (see Fig. 4-1), Alice performs a $\{cq \to q\}$ encoding $(\mathcal{E}_x : A \to A')_x$, depending on the source random variable, and then sends the A' system through the perfect quantum channel. After time t Bob performs a POVM $\Lambda : A'B \to X'$, on the system $A'B$, yielding the random variable X'. The protocol ends at time t_f. Unless otherwise stated, the entropic quantities below refer to the state of the system at time t.

Since at time t_f the state of the system XX' is supposed to be ϵ-close to $\overline{\Phi}_D$, Lemma 1.2 implies

$$I(X; X')_{t_f} \geq n(R - \delta) - \eta'(\epsilon) - K\epsilon n R.$$

By the Holevo bound [Hol73],

$$I(X; X')_{t_f} \leq I(X; A'B).$$

Recall from Eq. (1.1) the identity

$$I(X; A'B) = H(A') + I(A' \rangle BX) - I(A'; B) + I(X; B).$$

Since $I(A'; B) \geq 0$, and in our protocol $I(X; B) = 0$, this becomes

$$I(X; A'B) \leq H(A') + I(A' \rangle BX).$$

Observing that

$$nQ \geq H(A') \geq H(A'|X),$$

these all add up to

$$R \leq Q + \frac{1}{n} I(A' \rangle BX) + \delta + KR\epsilon + \frac{\eta'(\epsilon)}{n}.$$

As these are true for any $\epsilon, \delta > 0$ and sufficiently large n, the converse holds.

Regarding the direct coding theorem, it suffices to demonstrate the RI

$$\langle \rho^{AB} \rangle + H(A'|X)_\sigma \, [q \to q] \geq I(A'; B|X)_\sigma \, [c \to c].$$

This, in turn, follows from linearly combining Corollary 4.5 with super-dense coding (Eq. 1.40) much in the same way the noisy super-dense coding RI (Eq. 1.48) follows from the mother (Eq. ♀). □

4.2.3 Trade-off for quantum communication assisted entanglement distillation

Given a bipartite state ρ^{AB}, the quantum communication assisted entanglement distillation capacity region (or "mother" capacity region for short) $C_{\mathrm{M}}(\rho^{AB})$ is the set of (Q, E) with $Q \geq 0$ and $E \geq 0$ satisfying the RI

$$\langle \rho^{AB} \rangle + Q \, [q \to q] \geq E \, [q \, q]. \tag{4.7}$$

(This RI is trivially false for $Q < 0$ and trivially true for $Q \geq 0$ and $E \geq 0$.)

Theorem 4.7. *The capacity region $C_{\mathrm{M}}(\rho^{AB})$ is given by*

$$C_{\mathrm{M}}(\rho^{AB}) = \widetilde{C}_{\mathrm{M}}(\rho^{AB}) := \overline{\bigcup_{l=1}^{\infty} \frac{1}{l} \widetilde{C}_{\mathrm{M}}^{(1)}((\rho^{AB})^{\otimes l})},$$

where $\widetilde{C}_{\mathrm{M}}^{(1)}(\rho^{AB})$ is the set of all $Q \geq 0$, $E \geq 0$ such that

$$E \leq Q + \max_{\sigma} \left\{ I(A'\rangle BX)_\sigma : \frac{1}{2} I(A'; EE'|X)_\sigma \leq Q \right\}. \tag{4.8}$$

In the above, σ is the QP version of Eq. (4.6), namely

$$\sigma^{XA'BEE'} = \sum_x p_x |x\rangle\langle x|^X \otimes U_x^{A\to A'E'}(\psi^{ABE}). \tag{4.9}$$

for some ensemble of isometries (p_x, U_x), $U_x : A \to A'E'$, and purification $|\psi\rangle\langle\psi|^{ABE} \supseteq \rho^{AB}$.

Proof. We first prove the converse, which in this case follows from the converse for the noisy super-dense coding trade-off. The main observation is that super-dense coding (Eq. (1.40)) induces an invertible linear map f between the (Q, E) and (Q, R) planes corresponding to the mother capacity region and that of noisy super-dense coding, respectively, defined by

$$f : (Q, E) \mapsto (Q + E, 2E).$$

By adding superdense coding (i.e. $E[q\,q] + E[q \to q] \geq 2E[c \to c]$) to the mother (Eq. 4.7), we find

$$f(C_{\mathrm{M}}) \subseteq C_{\mathrm{NSD}}. \tag{4.10}$$

On the other hand, by inspecting the definitions of $\widetilde{C}_{\mathrm{NSD}}$ and $\widetilde{C}_{\mathrm{M}}$, we can verify

$$\widetilde{C}_{\mathrm{NSD}} = f(\widetilde{C}_{\mathrm{M}}). \tag{4.11}$$

The converse for the noisy super-dense coding trade-off is written as $C_{\mathrm{NSD}} \subseteq \widetilde{C}_{\mathrm{NSD}}$. As f is a bijection, putting everything together we have

$$C_{\mathrm{M}} \subseteq f^{-1}(C_{\mathrm{NSD}}) \subseteq f^{-1}(\widetilde{C}_{\mathrm{NSD}}) = \widetilde{C}_{\mathrm{M}},$$

which is the converse for the mother trade-off.

The direct coding theorem follows immediately from Corollary 4.5. □

4.2.4 Trade-off for noisy teleportation

Given a bipartite state ρ^{AB}, the noisy super-dense coding capacity region $C_{\mathrm{NTP}}(\rho^{AB})$ is a two-dimensional region in the (R, Q) plane with $R \geq 0$ and $Q \geq 0$ satisfying the RI

$$\langle \rho^{AB} \rangle + R\,[c \to c] \geq Q\,[q \to q]. \tag{4.12}$$

Theorem 4.8. *The capacity region $C_{\mathrm{NTP}}(\rho^{AB})$ is given by*

$$C_{\mathrm{NTP}}(\rho^{AB}) = \widetilde{C}_{\mathrm{NTP}}(\rho^{AB}) := \overline{\bigcup_{l=1}^{\infty} \frac{1}{l} \widetilde{C}_{\mathrm{NTP}}^{(1)}((\rho^{AB})^{\otimes l})},$$

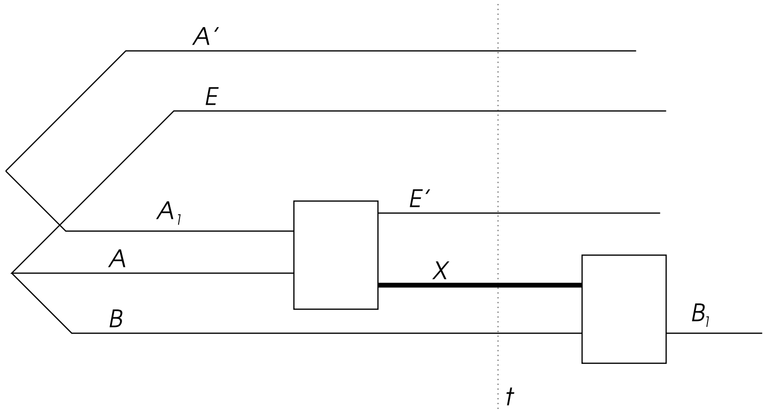


Figure 4-2: A general protocol for noisy teleportation.

where $\widetilde{C}^{(1)}_{\rm NTP}(\rho^{AB})$ is the set of all $R \geq 0$, $Q \geq 0$ such that

$$Q \leq \max_{\sigma} \left\{ I(A' \rangle BX)_\sigma : I(A'; B|X)_\sigma + I(X; BE)_\sigma \leq R \right\}. \tag{4.13}$$

In the above, σ is of the form

$$\sigma^{XA'BE} = \mathbf{T}(\psi^{ABE}), \tag{4.14}$$

for some instrument $\mathbf{T} : A \rightarrow A'X$ and purification $|\psi\rangle\langle\psi|^{ABE} \supseteq \rho^{AB}$.

Proof. We first prove the converse. Fix $n, Q, R, \delta, \epsilon$, and use the Flattening Lemma so we can assume that the depth is one. The resources available are

- The state $(\rho^{AB})^{\otimes n}$ shared between Alice and Bob. Let it be contained in the system $A^n B^n$, which we shall call AB for short.
- A perfect classical channel of size 2^{nR}.

The resource to be simulated is the perfect quantum channel ${\rm id}_D : A_1 \rightarrow B_1$, $D = \dim A_1 = 2^{n(Q-\delta)}$, from Alice to Bob, on any source, in particular on the maximally entangled state $\Phi^{A'A_1}$.

In the protocol (see Fig. 4-2), Alice performs a POVM $\Lambda : AA_1 \rightarrow X$ on the system AA_1, and sends the outcome random variable X through the classical channel. After time t Bob performs a $\{cq \rightarrow q\}$ decoding quantum operation $\mathcal{D} : XB \rightarrow B_1$. The protocol ends at time t_f. Unless otherwise stated, the entropic quantities below refer to the time t.

Our first observation is that performing the POVM Λ induces an instrument $\mathbf{T} : A \rightarrow A'X$,[*] so that the state of the system $XA'BE$ at time t is indeed of the form of Eq. (4.14).

Since at time t_f the state of the system $A'B_1$ is supposed to be ϵ-close to Φ_D, Lemma 1.2 implies

$$I(A'\rangle B_1)_{t_f} \geq n(Q - \delta) - \eta'(\epsilon) - K\epsilon n Q.$$

By the data processing inequality,

$$I(A'\rangle B_1)_{t_f} \leq I(A'\rangle BX).$$

Thus

$$Q \leq \frac{1}{n} I(A' \rangle BX) + \delta + KQ\epsilon + \frac{\eta'(\epsilon)}{n}. \tag{4.15}$$

[*] Indeed, first a pure ancilla $A'A_1$ was appended, then another pure ancilla X was appended, the system $AA'A_1X$ was rotated to $A'E'X$, and finally X was measured and E' was traced out.

To bound R, start with the identity

$$I(X;A'BE) = H(A') + I(A'\,\rangle BEX) - I(A';BE) + I(X;BE).$$

Since $I(A';BE) = 0$, $H(A') \geq H(A'|X)$ and $I(A'\,\rangle BEX) \geq I(A'\,\rangle BX)$, this becomes

$$I(X;A'BE) \geq I(A';B|X) + I(X;BE).$$

Combining this with

$$nR \geq H(X) \geq I(X;A'BE)$$

gives the desired

$$R \geq \frac{1}{n}[I(A';B|X) + I(X;BE)]. \tag{4.16}$$

As Eqns. (4.15) and (4.16) are true for any $\epsilon, \delta > 0$ and sufficiently large n, the converse holds.

Regarding the direct coding theorem, it suffices to demonstrate the RI

$$\langle \rho^{AB} \rangle + (I(A';B|X)_\sigma + I(X;BE)_\sigma)\,[c \to c] \geq I(A'\,\rangle BX)_\sigma\,[q \to q]. \tag{4.17}$$

Linearly combining the grandmother RI (Eq. (4.4)) with teleportation (Eq. (1.39)), much in the same way the variation on the noisy teleportation RI (Eq. (4.2)) was obtained from the mother (Eq. (♀)), we have

$$\langle \rho^{AB} \rangle + (I(A';B|X)_\sigma + I(X;BE)_\sigma)\,[c \to c] + o[q\,q] \geq I(A'\,\rangle BX)_\sigma\,[q \to q].$$

Equation (4.17) follows by invoking Lemma 1.36 and Eq. (1.49). □

4.2.5 Trade-off for classical communication assisted entanglement distillation

Given a bipartite state ρ^{AB}, the classical communication assisted entanglement distillation capacity region (or "entanglement distillation" capacity region for short) $C_{\mathrm{ED}}(\rho^{AB})$ is the two-dimensional region in the (R, E) plane with $R \geq 0$ and $E \geq 0$ satisfying the RI

$$\langle \rho^{AB} \rangle + R\,[c \to c] \geq E\,[q\,q]. \tag{4.18}$$

Theorem 4.9. *The capacity region $C_{\mathrm{ED}}(\rho^{AB})$ is given by*

$$C_{\mathrm{ED}}(\rho^{AB}) = \widetilde{C}_{\mathrm{ED}}(\rho^{AB}) := \overline{\bigcup_{l=1}^{\infty} \frac{1}{l} \widetilde{C}^{(1)}_{\mathrm{ED}}((\rho^{AB})^{\otimes l})},$$

where $\widetilde{C}^{(1)}_{\mathrm{ED}}(\rho^{AB})$ is the set of all $R \geq 0$, $E \geq 0$ such that

$$E \leq \max_{\sigma} \left\{ I(A'\,\rangle BX)_\sigma : I(A';EE'|X)_\sigma + I(X;BE)_\sigma \leq R \right\}, \tag{4.19}$$

In the above, σ is the fully QP version of Eq. (4.14), namely

$$\sigma^{XA'BEE'} = \mathbf{T}'(\psi^{ABE}), \tag{4.20}$$

for some instrument $\mathbf{T} : A \to A'E'X$ with pure quantum output and purification $|\psi\rangle\langle\psi|^{ABE} \supseteq \rho^{AB}$.

Proof. We first prove the converse, which in this case follows from the converse for the noisy teleportation trade-off. The argument very much parallels that of the converse for the mother trade-off. The main observation is that teleportation (Eq. (1.39)) induces an invertible linear map g between the (R, E) and (R, Q) planes corresponding to the entanglement distillation capacity region and that of noisy teleportation, respectively, defined by

$$g : (R, E) \mapsto (R + 2E, E).$$

By applying TP to Eq. (4.18), we find

$$g(C_{\rm ED}) \subseteq C_{\rm NTP}. \tag{4.21}$$

On the other hand, from the definitions of $\widetilde{C}_{\rm ED}$ and $\widetilde{C}_{\rm NTP}$ (Eqns. (4.19) and (4.13)), we have

$$\widetilde{C}_{\rm ED} = g(\widetilde{C}_{\rm NTP}). \tag{4.22}$$

The converse for the noisy teleportation trade-off is written as $C_{\rm NTP} \subseteq \widetilde{C}_{\rm NTP}$. As g is a bijection, putting everything together we have

$$C_{\rm ED} \subseteq g^{-1}(C_{\rm NTP}) \subseteq g^{-1}(\widetilde{C}_{\rm NTP}) = \widetilde{C}_{\rm ED},$$

which is the converse for the entanglement distillation trade-off.

Regarding the direct coding theorem, it suffices to demonstrate the RI

$$\langle \rho^{AB} \rangle + (I(A'; EE'|X)_\sigma + I(X; BE)_\sigma)\,[c \to c] \geq I(A' \rangle BX)_\sigma\,[q\,q]. \tag{4.23}$$

Linearly combining the grandmother RI (Eq. (4.4)) with teleportation (1.39), much in the same way the variation on the hashing RI (Eq. (4.1)) was obtained from the mother (Eq. (♀)), we have

$$\langle \rho^{AB} \rangle + (I(A'; EE'|X)_\sigma + I(X; BE)_\sigma)\,[c \to c] + o[q\,q] \geq I(A' \rangle BX)_\sigma\,[q \to q].$$

Eq. (4.23) follows by invoking Lemma 1.36 and Eq. (1.49). □

4.2.6 Trade-off for entanglement assisted quantum communication

Given a noisy quantum channel $\mathcal{N} : A' \to B$, the entanglement assisted quantum communication capacity region (or "father" capacity region for short) $C_{\rm F}(\mathcal{N})$ is the region of (E, Q) plane with $E \geq 0$ and $Q \geq 0$ satisfying the RI

$$\langle \mathcal{N} \rangle + E\,[q\,q] \geq Q\,[q \to q]. \tag{4.24}$$

Theorem 4.10. *The capacity region $C_{\rm F}(\mathcal{N})$ is given by*

$$C_{\rm F}(\mathcal{N}) = \widetilde{C}_{\rm F}(\mathcal{N}) := \overline{\bigcup_{l=1}^{\infty} \frac{1}{l} \widetilde{C}^{(1)}_{\rm F}(\mathcal{N}^{\otimes l})},$$

where $\widetilde{C}^{(1)}_{\rm F}(\mathcal{N})$ is the set of all $E \geq 0$, $Q \geq 0$ such that

$$\begin{aligned} Q &\leq E + I(A \rangle B)_\sigma \\ Q &\leq \frac{1}{2} I(A; B)_\sigma. \end{aligned}$$

In the above, σ is of the form

$$\sigma^{ABE} = U_{\mathcal{N}} \circ \mathcal{E}(\phi^{AA''}),$$

for some pure input state $|\phi^{AA''}\rangle$, encoding operation $\mathcal{E} : A'' \to A'$, and where $U_{\mathcal{N}} : A' \to BE$ is an isometric extension of $\mathcal{N}$.

This tradeoff region includes two well-known limit points. When $E = 0$, the quantum capacity of

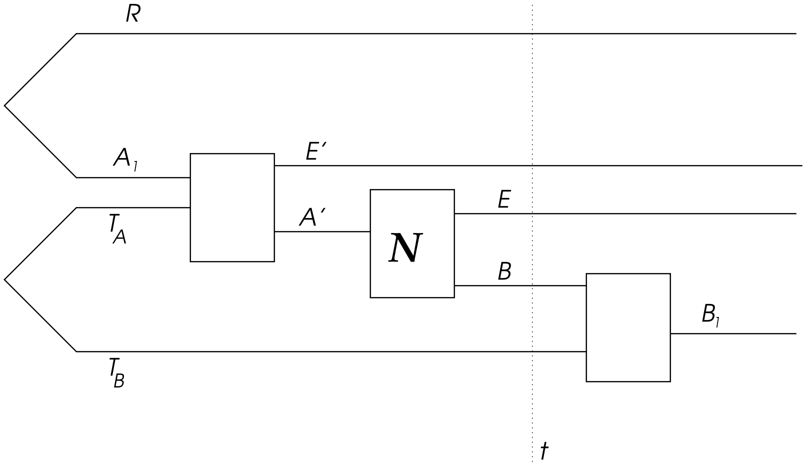


Figure 4-3: A general protocol for entanglement assisted quantum communication.

$\mathcal{N}$ is $I(A\rangle B)$[Llo96, Sho02, Dev05a], and for $E > 0$, entanglement distribution ($[q \to q] \geq [q\,q]$) means it should still be bounded by $I(A\rangle B)+E$. On the other hand, when given unlimited entanglement, the classical capacity is $I(A;B)$[BSST02] and thus the quantum capacity is never greater than $\frac{1}{2}I(A;B)$ no matter how much entanglement is available. These bounds meet when $E = \frac{1}{2}I(A;E)$ and $Q = \frac{1}{2}I(A;E)$, the point corresponding to the father protocol. Thus, the goal of our proof is to show that the father protocol is optimal.

Proof. We first prove the converse. Fix $n, E, Q, \delta, \epsilon$, and use the Flattening Lemma to reduce the depth to one. The resources available are

- The channel $\mathcal{N}^{\otimes n} : A'^n \to B^n$ from Alice to Bob. We shall shorten A'^n to A' and B^n to B.
- The maximally entangled state $\Phi^{T_A T_B}$, $\dim T_A = \dim T_B = 2^{nE}$, shared between Alice and Bob.

The resource to be simulated is the perfect quantum channel $\mathrm{id}_D : A_1 \to B_1$, $D = \dim A_1 = 2^{n(Q-\delta)}$, from Alice to Bob, on any source, in particular on the maximally entangled state Φ^{RA_1}.

In the protocol (see Fig. 4-3), Alice performs a general encoding map $\mathcal{E} : A_1 T_A \to A'E'$ and sends the system A' through the noisy channel $\mathcal{N} : A' \to B$. After time t Bob performs a decoding operation $\mathcal{D} : BT_B \to B_1$. The protocol ends at time t_f. Unless otherwise stated, the entropic quantities below refer to the time t.

Define $A := RT_B$ and $A'' := A_1 T_A$. Since at time t_f the state of the system RB_1 is supposed to be ϵ-close to Φ_D, Lemma 1.2 implies

$$I(R\rangle B_1)_{t_f} \geq n(Q-\delta) - \eta'(\epsilon) - K\epsilon n Q.$$

By the data processing inequality,

$$I(R\rangle B_1)_{t_f} \leq I(R\rangle BT_B).$$

Together with the inequality

$$I(R\rangle BT_B) \leq I(RT_B\rangle B) + H(T_B),$$

since $E = H(T_B)$, the above implies

$$Q \leq E + \frac{1}{n} I(A\rangle B) + \delta + KQ\epsilon + \frac{\eta'(\epsilon)}{n}.$$

Combining this with

$$H(A) = H(R) + H(T_B) = nQ + nE.$$

gives

$$Q \leq \frac{1}{2n} I(A;B) + \delta/2 + KQ\epsilon/2 + \frac{\eta'(\epsilon)}{2n}.$$

As these are true for any $\epsilon, \delta > 0$ and sufficiently large n, the converse holds.

Regarding the direct coding theorem, it follows directly form the father RI

$$\langle \mathcal{N} \rangle + \tfrac{1}{2} I(A;E)_\sigma \, [q\,q] \geq \tfrac{1}{2} I(A;B)_\sigma \, [q \to q].$$

□

4.2.7 Trade-off for entanglement assisted classical communication

The result of this subsection was first proved by Shor in [Sho04b]. Here we state it for completeness, and give an independent proof of the converse. An alternative proof of the direct coding theorem was sketched in [DS03] and is pursued in [DHLS05] to unify this result with the father trade-off.

Given a noisy quantum channel $\mathcal{N} : A' \to B$, the entanglement assisted classical communication capacity region (or "entanglement assisted" capacity region for short) $C_{\mathrm{EA}}(\mathcal{N})$ is the set of all points (E, R) with $E \geq 0$ and $R \geq 0$ satisfying the RI

$$\langle \mathcal{N} \rangle + E \, [q\,q] \geq R \, [c \to c]. \tag{4.25}$$

Theorem 4.11. *The capacity region $C_{\mathrm{EA}}(\mathcal{N})$ is given by*

$$C_{\mathrm{EA}}(\mathcal{N}) = \widetilde{C}_{\mathrm{EA}}(\mathcal{N}) := \overline{\bigcup_{l=1}^{\infty} \frac{1}{l} \widetilde{C}^{(1)}_{\mathrm{EA}}(\mathcal{N}^{\otimes l})},$$

where $\widetilde{C}^{(1)}_{\mathrm{EA}}(\mathcal{N})$ is the set of all $E \geq 0$, $R \geq 0$ such that

$$R \leq \max_{\sigma} \left\{ I(AX;B)_\sigma : E \geq H(A|X)_\sigma \right\}. \tag{4.26}$$

In the above, σ is of the form

$$\sigma^{XAB} = \sum_x p_x |x\rangle\langle x|^X \otimes \mathcal{N}(\phi_x^{AA'}), \tag{4.27}$$

for some pure input ensemble $(p_x, |\phi_x\rangle^{AA'})_x$.

Proof. We first prove the converse. Fix $n, E, Q, \delta, \epsilon$, and again use the flattening lemma to reduce depth to one. The resources available are

- The channel $\mathcal{N}^{\otimes n} : A'^n \to B^n$ from Alice to Bob. We shall shorten A'^n to A' and B^n to B.
- The maximally entangled state $\Phi^{T_A T_B}$, $\dim T_A = \dim T_B = 2^{nE}$, shared between Alice and Bob.

The resource to be simulated is the perfect classical channel of size $D = 2^{n(R-\delta)}$ on any source, in particular on the random variable X corresponding to the uniform distribution π_D.

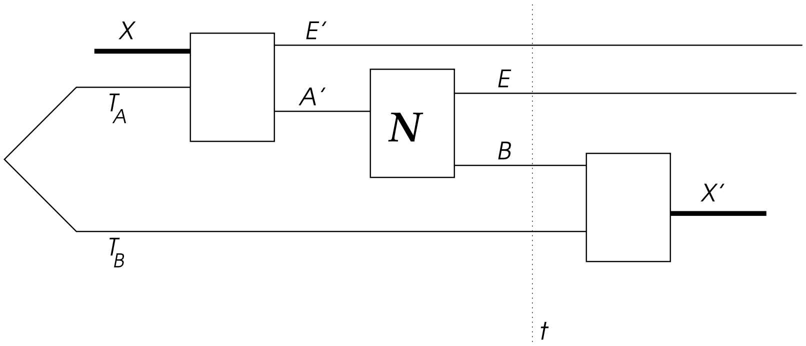


Figure 4-4: A general protocol for entanglement assisted classical communication.

In the protocol (see Fig. 4-4), Alice performs a $\{cq \to q\}$ encoding $(\mathcal{E}_x : T_A \to A')_x$, depending on the source random variable, and then sends the T_A system through the noisy channel $\mathcal{N} : A' \to BE$. After time t Bob performs a POVM $\Lambda : T_B B \to X'$, on the system $T_B B$, yielding the random variable X'. The protocol ends at time t_f. Unless otherwise stated, the entropic quantities below refer to the state of the system at time t.

Since at time t_f the state of the system XX' is supposed to be ϵ-close to $\overline{\Phi}_D$, Lemma 1.2 implies

$$I(X;X')_{t_f} \geq n(R-\delta) - \eta'(\epsilon) - K\epsilon n R.$$

By the Holevo bound

$$I(X;X')_{t_f} \leq I(X;T_B B).$$

Using the chain rule twice, we find

$$\begin{aligned} I(X;T_B B) &= I(X;B|T_B) + I(X;T_B) \\ &= I(XT_B;B) + I(X;T_B) - I(T_B;B) \end{aligned}$$

Since $I(T_B;B) \geq 0$ and in this protocol $I(X;T_B) = 0$, this becomes

$$I(X;T_B B) \geq I(XT_B;B).$$

These all add up to

$$R \leq \frac{1}{n} I(XT_B;B) + \delta + Kd\epsilon + \frac{\eta'\epsilon}{n},$$

while on the other hand,

$$nE \geq H(T_B|X).$$

As these are true for any $\epsilon, \delta > 0$ and sufficiently large n, we have thus shown a variation on the converse with the state σ from (4.27) replaced by $\widetilde{\sigma}$,

$$\widetilde{\sigma}^{XABE'} = \sum_x p_x |x\rangle\langle x|^X \otimes \mathcal{N} \circ U_x^{A'' \to A'E'}(\phi^{AA''}),$$

defining $A := T_B$ and letting $U_x : T_A \to A'E'$ be the isometric extension of $\mathcal{E}_x$.

However, this is a weaker result than we would like; the converse we have proved allows arbitrary noisy encodings and we would like to show that isometric encodings are optimal, or equivalently that the E' register is unnecessary. We will accomplish this, following Shor [Sho04a], by using a standard trick of measuring E' and showing that the protocol can only improve. If we apply the dephasing map $\overline{\mathrm{id}} : E' \to Y$ to $\widetilde{\sigma}^{ABE'}$, we obtain a state of the form

$$\sigma^{XYAB} = \sum_{xy} p_{xy} |x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \otimes \mathcal{N}(\psi_{xy}^{AA'}).$$

The converse now follows from

$$\begin{aligned} I(B;AX)_{\widetilde{\sigma}} &\leq I(B;AXY)_\sigma \\ H(A|X)_{\widetilde{\sigma}} &\geq H(A|XY)_\sigma. \end{aligned}$$

□

4.3 Conclusion

The goal of quantum Shannon theory is to give information-theoretic formulae for the rates at which noisy quantum resources can be converted into noiseless ones. This chapter has taken a major step towards that goal by finding the trade-off curves for most one-way communication scenarios involving a noisy state or channel and two of the three basic noiseless resources (cbits, ebits and qubits). The main tools required for this were the resource formalism of Chapter 1, coherent classical communication (from Chapter 3), derandomization and basic protocols like HSW coding.

However, our expressions for trade-off curves also should be seen more as first steps rather than final answers. For one thing, we would ultimately like to have formulae for the capacity that can be efficiently computed, which will probably require replacing our current regularized expressions with single-letter ones. This is related to the additivity conjectures, which are equivalent for some channel capacities[Sho03], but are false for others[DSS98].

A more reasonable first goal is to strengthen some of the converse theorems, so that they do not require maximizing over as many different quantum operations. As inspiration, note that [BKN00] showed that isometric encodings suffice to achieve the optimal rate of quantum communication through a quantum channel. However, the analogous result for entanglement-assisted quantum communication is not known. Specifically, in Fig. 4-3, I suspect that the E' register (used to discard some of the inputs) is only necessary when Alice and Bob share more entanglement than the protocol can use. Similarly, it seems plausible to assume that the optimal form of protocols for noisy teleportation (Fig. 4-2) is to perform a general TPCP preprocessing operation on the shared entanglement, followed by a unitary interaction between the quantum data and Alice's part of the entangled state. These are only two of the more obvious examples and there ought to be many possible ways of improving our formulae.

Chapter 5

The Schur transform

5.1 Overview

The final four chapters will explore the uses of Schur duality in quantum computing and information theory. Schur duality is a natural way to decompose $(\mathbb{C}^d)^{\otimes n}$ in terms of representations of the symmetric group $\mathcal{S}_n$ and the unitary group $\mathcal{U}_d$. In this chapter, we will describe Schur duality and develop its representation-theoretic background within the framework of quantum information. The primary connection between these fields is that a vector space can be interpreted either as a representation of a group or as state-space of a quantum system. Thus, Schur duality can be interpreted both as a mathematical fact about representations and operationally as a fact about the transformations possible on a quantum system.

Chapter 6 will describe how Schur duality is useful in quantum information theory. We will see that Schur duality is a quantum analogue of the classical method of types, in which strings are described in terms of their empirical distributions. This has a number of applications in information theory, which we will survey while highlighting the role of Schur duality. The chapter concludes with new work describing how i.i.d. quantum channels can be decomposed in the Schur basis.

We then turn to computational issues in Chapters 7 and 8. The unitary transform that relates the Schur basis to the computational basis is known as the *Schur transform* and presenting efficient circuits for the Schur transform is the main goal of Chapter 7. These circuits mean that the information-theoretic tasks described in Chapter 6 can now all be implemented efficiently on a quantum computer; even though computational efficiency is not often considered in quantum information theory, it will be necessary if we ever expect to implement many of the coding schemes that exist.

Finally, Chapter 8 discusses algorithmic connections between the Schur transform and related efficient representation-theoretic transforms, such as the quantum Fourier transform on $\mathcal{S}_n$. Ultimately the goal of this work is to find quantum speedups that use either the Schur transform or the $\mathcal{S}_n$ Fourier transform.

Most of the original work in this chapter has not yet been published. The next two chapters are mostly review, although there are several places where the material is assembled and presented in ways that have not seen before in the literature. The exception is the last section of Chapter 6 on decomposing i.i.d. quantum channels, which is a new contribution. The last two chapters are joint work with Dave Bacon and Isaac Chuang. Parts of Chapter 7 appeared in [BCH04] and the rest of the chapter will be presented in [BCH05a]. Chapter 8 will become [BCH05b].

5.2 Representation theory and quantum computing

5.2.1 Basics of representation theory

In this section, we review aspects of representation theory that will be used in the second half of the thesis. For a more detailed description of representation theory, the reader should consult [Art95] for general facts about group theory and representation theory or [GW98] for representations of Lie groups. See also [FH91] for a more introductory and informal approach to Lie groups and their representations.

Representations: For a complex vector space V, define $\text{End}(V)$ to be set of linear maps from V to itself (endomorphisms). A representation of a group G is a vector space V together with a homomorphism from G to $\text{End}(V)$, i.e. a function $\mathbf{R} : G \to \text{End}(V)$ such that $\mathbf{R}(g_1)\mathbf{R}(g_2) = \mathbf{R}(g_1 g_2)$. If $\mathbf{R}(g)$ is a unitary operator for all g, then we say $\mathbf{R}$ is a unitary representation. Furthermore, we say a representation $(\mathbf{R}, V)$ is finite dimensional if V is a finite dimensional vector space. In this thesis, we will always consider complex finite dimensional, unitary representations and use the generic term 'representation' to refer to complex, finite dimensional, unitary representations. Also, when clear from the context, we will denote a representation $(\mathbf{R}, V)$ simply by the representation space V.

The reason we consider only complex, finite dimensional, unitary representations is so that we can use them in quantum computing. If $d = \dim V$, then a d-dimensional quantum system can hold a unit vector in a representation V. A group element $g \in G$ corresponds to a unitary rotation $\mathbf{R}(g)$, which can in principle be performed by a quantum computer.

Homomorphisms: For any two vector spaces V_1 and V_2, define $\text{Hom}(V_1, V_2)$ to be the set of linear transformations from V_1 to V_2. If G acts on V_1 and V_2 with representation matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ then the canonical action of G on $\text{Hom}(V_1, V_2)$ is given by the map from M to $\mathbf{R}_2(g)M\mathbf{R}_1(g)^{-1}$ for any $M \in \text{Hom}(V_1, V_2)$. For any representation $(\mathbf{R}, V)$ define V^G to be the space of G-invariant vectors of V: i.e. $V^G := \{|v\rangle \in V : \mathbf{R}(g)|v\rangle = |v\rangle \, \forall g \in G\}$. Of particular interest is the space $\text{Hom}(V_1, V_2)^G$, which can be thought of as the linear maps from V_1 to V_2 which commute with the action of G. If $\text{Hom}(V_1, V_2)^G$ contains any invertible maps (or equivalently, any unitary maps) then we say that $(\mathbf{R}_1, V_1)$ and $(\mathbf{R}_2, V_2)$ are *equivalent* representations and write

$$V_1 \overset{G}{\cong} V_2.$$

This means that there exists a unitary change of basis $U : V_1 \to V_2$ such that for any $g \in G$, $U\mathbf{R}_1(g)U^\dagger = \mathbf{R}_2(g)$.

Dual representations: Recall that the *dual* of a vector space V is the set of linear maps from V to $\mathbb{C}$ and is denoted V^*. Usually if vectors in V are denoted by kets (e.g. $|v\rangle$) then vectors in V^* are denoted by bras (e.g. $\langle v|$). If we fix a basis $\{|v_1\rangle, |v_2\rangle, \ldots\}$ for V then the transpose is a linear map from V to V^* given by $|v_i\rangle \to \langle v_i|$. Now, for a representation $(\mathbf{R}, V)$ we can define the *dual representation* $(\mathbf{R}^*, V^*)$ by $\mathbf{R}^*(g)\langle v^*| := \langle v^*|\mathbf{R}(g^{-1})$. If we think of $\mathbf{R}^*$ as a representation on V (using the transpose map to relate V and V^*), then it is given by $\mathbf{R}^*(g) = (\mathbf{R}(g^{-1}))^T$. When $\mathbf{R}$ is a unitary representation, this is the same as the *conjugate representation* $\mathbf{R}(g)^*$, where here * denotes the entrywise complex conjugate. One can readily verify that the dual and conjugate representations are indeed representations and that $\text{Hom}(V_1, V_2) \overset{G}{\cong} V_1^* \otimes V_2$.

Irreducible representations: Generically the unitary operators of a representation may be specified (and manipulated on a quantum computer) in an arbitrary orthonormal basis. The added structure of being a representation, however, implies that there are particular bases which are more fundamental to expressing the action of the group. We say a representation $(\mathbf{R}, V)$ is irreducible (and call it an irreducible representaiton, or *irrep*) if the only subspaces of V which are invariant under $\mathbf{R}$ are the empty subspace $\{0\}$ and the entire space V. For finite groups, any finite-dimensional complex representation is reducible; meaning it is decomposable into a direct sum of irreps. For Lie groups, we need additional conditions, such as demanding that the representation $\mathbf{R}(g)$ be *rational*; i.e. its matrix elements are polynomial functions of the matrix elements g_{ij} and $(\det g)^{-1}$. We say a representation of a Lie group is *polynomial* if its matrix elements are polynomial functions only of the g_{ij}.

Isotypic decomposition: Let $\hat{G}$ be a complete set of inequivalent irreps of G. Then for any reducible representation $(\mathbf{R}, V)$ there is a basis under which the action of $\mathbf{R}(g)$ can be expressed as

$$\mathbf{R}(g) \cong \bigoplus_{\lambda\in\hat{G}} \bigoplus_{j=1}^{n_\lambda} \mathbf{r}_\lambda(g) = \bigoplus_{\lambda\in\hat{G}} \mathbf{r}_\lambda(g) \otimes I_{n_\lambda} \tag{5.1}$$

where $\lambda \in \hat{G}$ labels an irrep $(\mathbf{r}_\lambda, V_\lambda)$ and n_λ is the multiplicity of the irrep λ in the representation V. Here we use $\cong$ to indicate that there exists a unitary change of basis relating the left-hand size to the right-hand side.[*] Under this change of basis we obtain a similar decomposition of the representation space V (known as the *isotypic decomposition*):

$$V \overset{G}{\cong} \bigoplus_{\lambda\in\hat{G}} V_\lambda \otimes \mathbb{C}^{n_\lambda}. \tag{5.2}$$

Thus while generically we may be given a representation in some arbitrary basis, the structure of being a representation picks out a particular basis under which the action of the representation is not just block diagonal but also maximally block diagonal: a direct sum of irreps.

Moreover, the multiplicity space $\mathbb{C}^{n_\lambda}$ in Eq. (5.2) has the structure of $\mathrm{Hom}(V_\lambda, V)^G$. This means that for any representation $(\mathbf{R}, V)$, Eq. (5.2) can be restated as

$$V \overset{G}{\cong} \bigoplus_{\lambda\in\hat{G}} V_\lambda \otimes \mathrm{Hom}(V_\lambda, V)^G. \tag{5.3}$$

Since G acts trivially on $\mathrm{Hom}(V_\lambda, V)^G$, Eq. (5.1) remains the same. As with the other results in this chapter, a proof of Eq. (5.3) can be found in [GW98], or other standard texts on representation theory.

The value of Eq. (5.3) is that the unitary mapping from the right-hand side (RHS) to the left-hand side (LHS) has a simple explicit expression: it corresponds to the canonical map $\varphi : A\otimes\mathrm{Hom}(A, B) \to B$ given by $\varphi(a \otimes f) = f(a)$. Of course, this doesn't tell us how to describe $\mathrm{Hom}(V_\lambda, V)^G$, or how to specify an orthonormal basis for the space, but we will later find this form of the decomposition useful.

5.2.2 The Clebsch-Gordan transform

If $(\mathbf{R}_\mu, V_\mu)$ and $(\mathbf{R}_\nu, V_\nu)$ are representations of G, their tensor product $(\mathbf{R}_\mu \otimes \mathbf{R}_\nu, V_\mu \otimes V_\nu)$ is another representation of G. In general if V_μ and V_ν are irreducible, their tensor product will not necessarily be. According to Eq. (5.3), the tensor product decomposes as

$$V_\mu \otimes V_\nu \overset{G}{\cong} \bigoplus_{\lambda\in\hat{G}} V_\lambda \otimes \mathrm{Hom}(V_\lambda, V_\mu \otimes V_\nu)^G \overset{G}{\cong} \bigoplus_{\lambda\in\hat{G}} V_\lambda \otimes \mathbb{C}^{M_{\mu\nu}^\lambda}, \tag{5.4}$$

where we have defined the multiplicity $M_{\mu,\nu}^\lambda := \dim \mathrm{Hom}(V_\lambda, V_\mu \otimes V_\nu)^G$. When $G = \mathcal{U}_d$, the $M_{\mu\nu}^\lambda$ are known as *Littlewood-Richardson* coefficients.

The decomposition in Eq. (5.4) is known as the Clebsch-Gordan (CG) decomposition and the corresponding unitary map $U_{\mathrm{CG}}^{\mu,\nu}$ is called the CG transform. On a quantum computer, we can think of $U_{\mathrm{CG}}^{\mu,\nu}$ as a map from states of the form $|v_\mu\rangle|v_\nu\rangle$ to superpositions of states $|\lambda\rangle|v_\lambda\rangle|\alpha\rangle$, where $\lambda \in \hat{G}$

[*] We only need to use $\overset{G}{\cong}$ when relating representation spaces. In Eq. (5.1) and other similar isomorphisms, we instead explicitly specify the dependence of both sides on $g \in G$.

labels an irrep, $|v_\lambda\rangle$ is a basis state for V_λ and $\alpha \in \text{Hom}(V_\lambda, V_\mu \otimes V_\nu)^G$. Using the isomorphism $\text{Hom}(A,B) \overset{G}{\cong} A^* \otimes B$ we could also write that $|\alpha\rangle \in (V_\lambda^* \otimes V_\mu \otimes V_\nu)^G$; an interpretation which makes it more obvious how to normalize α.

There are a few issues that arise when implementing the map $U_{\text{CG}}^{\mu,\nu}$. For example, since different V_λ (and different multiplicity spaces) have different dimensions, the register for $|v_\lambda\rangle$ will need to be padded to at least $\lceil \log \max_\lambda \dim V_\lambda \rceil$ qubits. This means that the overall transformation will be an isometry that slightly enlarges the Hilbert space, or equivalently, will be a unitary that requires the input of a small number of ancilla qubits initialized to $|0\rangle$. Also, when G has an infinite number of inequivalent irreps (e.g. when G is a Lie group) then in order to store λ, we need to consider only some finite subset of $\hat{G}$. Fortunately, there is usually a natural way to perform this restriction.

Returning to Eq. (5.4) for a moment, note that all the complexity of the CG transform is pushed into the multiplicity space $\text{Hom}(V_\lambda, V_\mu \otimes V_\nu)^G$. For example, the fact that some values of λ don't appear on the RHS means that some of the multiplicity spaces may be zero. Also, the inverse transform $(U_{\text{CG}}^{\mu,\nu})^\dagger$ is given simply by the map

$$(U_{\text{CG}}^{\mu,\nu})^\dagger |\lambda\rangle|v_\lambda\rangle|\alpha\rangle = \alpha|v_\lambda\rangle. \tag{5.5}$$

We will use these properties of the CG transform when decomposing i.i.d. channels in Section 6.4 and in giving an efficient construction of the CG transform in Section 7.3.

5.2.3 The quantum Fourier transform

Let G be a finite group (we will return to Lie groups later). A useful representation is given by letting each $g \in G$ define an orthonormal basis vector $|g\rangle$. The resulting space $\text{Span}\{|g\rangle : g \in G\}$ is denoted $\mathbb{C}[G]$ and is called the *regular representation.* G can act on $\mathbb{C}[G]$ in two different ways: left multiplication $\mathbf{L}(g)|h\rangle := |gh\rangle$, and right multiplication $\mathbf{R}(g)|h\rangle := |hg^{-1}\rangle$. This means that there are really two different regular representations: the left regular representation $(\mathbf{L}, \mathbb{C}[G])$ and the right regular representation $(\mathbf{R}, \mathbb{C}[G])$. Since these representations commute, we could think of $\mathbf{L}(g_1)\mathbf{R}(g_2)$ as a representation of $G \times G$. Under this action, it can be shown that $\mathbb{C}[G]$ decomposes as

$$\mathbb{C}[G] \overset{G\times G}{\cong} \bigoplus_{\lambda \in \hat{G}} V_\lambda \hat{\otimes} V_\lambda^*. \tag{5.6}$$

Here the V_λ correspond to $\mathbf{L}$ and V_λ^* corresponds to $\mathbf{R}$, and $\hat{\otimes}$ is used to emphasize that we are not considering the tensor product action of a single group, but rather are taking the tensor product of two irreps from two different copies of the group G. This means that if we decompose only one of the regular representations, e.g. $(\mathbf{L}, \mathbb{C}[G])$, the V_λ^* in Eq. (5.6) becomes the multiplicity space for V_λ as follows:

$$\mathbb{C}[G] \overset{G}{\cong} \bigoplus_{\lambda \in \hat{G}} V_\lambda \otimes \mathbb{C}^{\dim V_\lambda}. \tag{5.7}$$

A similar expression holds for $(\mathbf{R}, \mathbb{C}[G])$ with V_λ^* appearing instead of V_λ.

The unitary matrix corresponding to the isomorphism in Eq. (5.6) is called the Fourier transform, or when it acts on quantum registers, the quantum Fourier transform (QFT). Denote this matrix by U_{QFT}. For any $g_1, g_2 \in G$ we have

$$\hat{\mathbf{L}}(g_1)\hat{\mathbf{R}}(g_2) := U_{\text{QFT}} \mathbf{L}(g_1)\mathbf{R}(g_2) U_{\text{QFT}}^\dagger = \sum_{\lambda \in \hat{G}} |\lambda\rangle\langle\lambda| \otimes \mathbf{r}_\lambda(g_1) \otimes \mathbf{r}_\lambda(g_2)^*, \tag{5.8}$$

where $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$ are the Fourier transformed versions of $\mathbf{L}$ and $\mathbf{R}$; $\hat{\mathbf{L}}(g) := U_{\text{QFT}}\mathbf{L}(g)U_{\text{QFT}}^\dagger$ and $\hat{\mathbf{R}}(g) := U_{\text{QFT}}\mathbf{R}(g)U_{\text{QFT}}^\dagger$.

Unlike the CG transform, the Fourier transform has a simple explicit expression.

$$U_{\mathrm{QFT}} = \sum_{g\in G} \sum_{\lambda\in\hat{G}} \sum_{i,j=1}^{\dim V_\lambda} \sqrt{\frac{\dim V_\lambda}{|G|}} \mathbf{r}_\lambda(g)_{ij} |\lambda, i, j\rangle\langle g| \tag{5.9}$$

The best-known quantum Fourier transform is over the cyclic group $G = \mathbb{Z}_N$. Here the form is particularly simple, since all irreps are one-dimensional and the set of irreps $\hat{G}$ is equivalent to $\mathbb{Z}_N$. Thus the $|i, j\rangle$ register can be neglected and we obtain the familiar expression $\sum_{x,y\in\mathbb{Z}_N} N^{-1/2} e^{2\pi ixy/N} |y\rangle\langle x|$. The ability of a quantum computer to efficiently implement this Fourier transform is at the heart of quantum computing's most famous advantages over classical computation[Sho94].

Quantum Fourier transforms can also be efficiently implemented for many other groups. Beals[Bea97] has shown how to implement the $\mathcal{S}_n$ QFT on a quantum computer in poly(n) time, Püschel, Rötteler and Beth[PRB99] have given efficient QFTs for other nonabelian groups and Moore, Rockmore and Russell[MRR04] have generalized these approaches to many other finite groups. Fourier transforms on Lie groups are also possible, though the infinite-dimensional spaces involved lead to additional complications that we will not discuss here. Later (Section 7.3) we will give an efficient algorithm for a $\mathcal{U}_d$ CG transform. However, if some sort of $\mathcal{U}_d$ QFT could be efficiently constructed on a quantum computer, then it would yield an alternate algorithm for the $\mathcal{U}_d$ CG transform. We will discuss this possibility further in Section 8.1.3 (see also Prop 9.1 of [Kup03]) and will discuss the $\mathcal{S}_n$ QFT more broadly in Chapter 8.

5.3 Schur duality

We now turn to the two representations relevant to the Schur transform. Recall that the symmetric group of degree n, $\mathcal{S}_n$, is the group of all permutations of n objects. Then we have the following natural representation of the symmetric group on the space $(\mathbb{C}^d)^{\otimes n}$:

$$\mathbf{P}(s)|i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_n\rangle = |i_{s^{-1}(1)}\rangle \otimes |i_{s^{-1}(2)}\rangle \otimes \cdots \otimes |i_{s^{-1}(n)}\rangle \tag{5.10}$$

where $s \in \mathcal{S}_n$ is a permutation and $s(i)$ is the label describing the action of s on label i. For example, consider the transposition $s = (12)$ belonging to the group $\mathcal{S}_3$. Then $\mathbf{P}(s)|i_1, i_2, i_3\rangle = |i_2, i_1, i_3\rangle$. $(\mathbf{P}, (\mathbb{C}^d)^{\otimes n})$ is the representation of the symmetric group which will be relevant to the Schur transform. Note that $\mathbf{P}$ obviously depends on n, but also has an implicit dependence on d.

Now we turn to the representation of the unitary group. Let $\mathcal{U}_d$ denote the group of $d \times d$ unitary operators. Then there is a representation of $\mathcal{U}_d$ given by the n-fold product action as

$$\mathbf{Q}(U)|i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_n\rangle = U|i_1\rangle \otimes U|i_2\rangle \otimes \cdots \otimes U|i_n\rangle \tag{5.11}$$

for any $U \in \mathcal{U}_d$. More compactly, we could write that $\mathbf{Q}(U) = U^{\otimes n}$. $(\mathbf{Q}, (\mathbb{C}^d)^{\otimes n})$ is the representation of the unitary group which will be relevant to the Schur transform.

Since both $\mathbf{P}(s)$ and $\mathbf{Q}(U)$ meet our above criteria for reducibility, they can each be decomposed into a direct sum of irreps as in Eq. (5.1),

$$\begin{aligned} \mathbf{P}(s) &\overset{\mathcal{S}_n}{\cong} \bigoplus_\alpha I_{n_\alpha} \otimes \mathbf{p}_\alpha(s) \\ \mathbf{Q}(U) &\overset{\mathcal{U}_d}{\cong} \bigoplus_\beta I_{m_\beta} \otimes \mathbf{q}_\beta(U) \end{aligned} \tag{5.12}$$

where n_α (m_β) is the multiplicity of the αth (βth) irrep $\mathbf{p}_\alpha(s)$ ($\mathbf{q}_\beta(U)$) in the representation $\mathbf{P}(s)$ ($\mathbf{Q}(U)$). At this point there is not necessarily any relation between the two different unitary transforms implementing the isomorphisms in Eq. (5.12). However, further structure in this decomposition follows from the fact that $\mathbf{P}(s)$ commutes with $\mathbf{Q}(U)$: $\mathbf{P}(s)\mathbf{Q}(U) = \mathbf{Q}(U)\mathbf{P}(s)$. This implies, via Schur's Lemma, that the action of the irreps of $\mathbf{P}(s)$ must act on the multiplicity labels of the irreps $\mathbf{Q}(U)$ and vice versa. Thus, the simultaneous action of $\mathbf{P}$ and $\mathbf{Q}$ on $(\mathbb{C}^d)^{\otimes n}$ decomposes as

$$\mathbf{Q}(U)\mathbf{P}(s) \overset{\mathcal{U}_d\times\mathcal{S}_n}{\cong} \bigoplus_\alpha \bigoplus_\beta I_{m_{\alpha,\beta}} \otimes \mathbf{q}_\beta(U) \otimes \mathbf{p}_\alpha(s) \tag{5.13}$$

where $m_{\alpha,\beta}$ can be thought of as the multiplicity of the irrep $\mathbf{q}_\beta(U)\hat{\otimes}\mathbf{p}_\alpha(s)$ of the group $\mathcal{U}_d \times \mathcal{S}_n$.

Not only do $\mathbf{P}$ and $\mathbf{Q}$ commute, but the algebras they generate (i.e. $\mathcal{A} := \mathbf{P}(\mathbb{C}[\mathcal{S}_n]) = \operatorname{Span}\{\mathbf{P}(s) : s \in \mathcal{S}_n\}$ and $\mathcal{B} := \mathbf{Q}(\mathbb{C}[\mathcal{U}_d]) = \operatorname{Span}\{\mathbf{Q}(U) : U \in \mathcal{U}_d\}$) *centralize* each other[GW98], meaning that $\mathcal{B}$ is the set of operators in $\operatorname{End}((\mathbb{C}^d)^{\otimes n})$ commuting with $\mathcal{A}$ and vice versa, $\mathcal{A}$ is the set of operators in $\operatorname{End}((\mathbb{C}^d)^{\otimes n})$ commuting with $\mathcal{B}$. This means that the multiplicities $m_{\alpha,\beta}$ are either zero or one, and that each α and β appears at most once. Thus Eq. (5.13) can be further simplified to

$$\mathbf{Q}(U)\mathbf{P}(s) \overset{\mathcal{S}_n\times\mathcal{U}_d}{\cong} \bigoplus_\lambda \mathbf{q}_\lambda(U) \otimes \mathbf{p}_\lambda(s) \tag{5.14}$$

where λ runs over some unspecified set.

Finally, Schur duality (or Schur-Weyl duality)[GW98] provides a simple characterization of the range of λ in Eq. (5.14) and shows how the decompositions are related for different values of n and d. To define Schur duality, we will need to somehow specify the irreps of $\mathcal{S}_n$ and $\mathcal{U}_d$.

Let $\mathcal{I}_{d,n} = \{\lambda = (\lambda_1, \lambda_2, \dots, \lambda_d) | \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0 \text{ and } \sum_{i=1}^d \lambda_i = n\}$ denote partitions of n into $\leq d$ parts. We consider two partitions $(\lambda_1, \dots, \lambda_d)$ and $(\lambda_1, \dots, \lambda_d, 0, \dots, 0)$ equivalent if they differ only by trailing zeroes; according to this principle, $\mathcal{I}_n := \mathcal{I}_{n,n}$ contains all the partitions of n. Partitions label irreps of $\mathcal{S}_n$ and $\mathcal{U}_d$ as follows: if we let d vary, then $\mathcal{I}_{d,n}$ labels irreps of $\mathcal{S}_n$, and if we let n vary, then $\mathcal{I}_{d,n}$ labels polynomial irreps of $\mathcal{U}_d$. Call these $(\mathbf{p}_\lambda, \mathcal{P}_\lambda)$ and $(\mathbf{q}^d_\lambda, \mathcal{Q}^d_\lambda)$ respectively, for $\lambda \in \mathcal{I}_{d,n}$. We need the superscript d because the same partition λ can label different irreps for different $\mathcal{U}_d$; on the other hand the $\mathcal{S}_n$-irrep $\mathcal{P}_\lambda$ is uniquely labeled by λ since $n = \sum_i \lambda_i$.

For the case of n qudits, Schur duality states that there exists a basis (which we label $|\lambda\rangle|q_\lambda\rangle|p_\lambda\rangle_{\text{Sch}}$ and call the *Schur basis*) which simultaneously decomposes the action of $\mathbf{P}(s)$ and $\mathbf{Q}(U)$ into irreps:

$$\begin{aligned} \mathbf{Q}(U)|\lambda\rangle|q_\lambda\rangle|p_\lambda\rangle_{\text{Sch}} &= |\lambda\rangle(\mathbf{q}^d_\lambda(U)|q_\lambda\rangle)|p_\lambda\rangle_{\text{Sch}} \\ \mathbf{P}(s)|\lambda\rangle|q_\lambda\rangle|p_\lambda\rangle_{\text{Sch}} &= |\lambda\rangle|q_\lambda\rangle(\mathbf{p}_\lambda(s)|p_\lambda\rangle)_{\text{Sch}} \end{aligned} \tag{5.15}$$

and that the common representation space $(\mathbb{C}^d)^{\otimes n}$ decomposes as

$$(\mathbb{C}^d)^{\otimes n} \overset{\mathcal{U}_d\times\mathcal{S}_n}{\cong} \bigoplus_{\lambda\in\mathcal{I}_{d,n}} \mathcal{Q}^d_\lambda \hat{\otimes} \mathcal{P}_\lambda. \tag{5.16}$$

The Schur basis can be expressed as superpositions over the standard computational basis states $|i_1, i_2, \dots, i_n\rangle$ as

$$|\lambda, q_\lambda, p_\lambda\rangle_{\text{Sch}} = \sum_{i_1,i_2,\dots,i_n} [\mathbf{U}_{\text{Sch}}]^{\lambda,q_\lambda,p_\lambda}_{i_1,i_2,\dots,i_n} |i_1 i_2 \dots i_n\rangle, \tag{5.17}$$

where $\mathbf{U}_{\text{Sch}}$ is the unitary transformation implementing the isomorphism in Eq. (5.16). Thus, for any $U \in \mathcal{U}_d$ and any $s \in \mathcal{S}_n$,

$$U_{\text{Sch}}\mathbf{Q}(U)\mathbf{P}(s)U^\dagger_{\text{Sch}} = \sum_{\lambda\in\mathcal{I}_{d,n}} |\lambda\rangle\langle\lambda| \otimes \mathbf{q}^d_\lambda(U) \otimes \mathbf{p}_\lambda(s). \tag{5.18}$$

If we now think of U_{Sch} as a quantum circuit, it will map the Schur basis state $|\lambda, q_\lambda, p_\lambda\rangle_{\text{Sch}}$ to the computational basis state $|\lambda, q_\lambda, p_\lambda\rangle$ with λ, q_λ, and p_λ expressed as bit strings. The dimensions of the irreps $\mathbf{p}_\lambda$ and $\mathbf{q}^d_\lambda$ vary with λ, so we will need to pad the $|q_\lambda, p_\lambda\rangle$ registers when they are expressed as bitstrings. We will label the padded basis as $|\lambda\rangle|q\rangle|p\rangle$, explicitly dropping the λ dependence. In Chapter 7 we will show how to do this padding efficiently with only a logarithmic spatial overhead. We will refer to the transform from the computational basis $|i_1, i_2, \ldots, i_n\rangle$ to the basis of three bitstrings $|\lambda\rangle|q\rangle|p\rangle$ as the Schur transform. The Schur transform is shown schematically in Fig. 5-1. Notice that just as the standard computational basis $|i\rangle$ is arbitrary up to a unitary transform, the bases for $\mathcal{Q}^d_\lambda$ and $\mathcal{P}_\lambda$ are also both arbitrary up to a unitary transform, though we will later choose particular bases for $\mathcal{Q}^d_\lambda$ and $\mathcal{P}_\lambda$.

*Example of the Schur transform:*Let $d = 2$. Then for $n = 2$ there are two valid partitions, $\lambda_1 = 2, \lambda_2 = 0$ and $\lambda_1 = \lambda_2 = 1$. Here the Schur transform corresponds to the change of basis from the standard basis to the singlet and triplet basis: $|\lambda = (1,1), q_\lambda = 0, p_\lambda = 0\rangle_{\mathrm{Sch}} = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, $|\lambda = (2,0), q_\lambda = +1, p_\lambda = 0\rangle_{\mathrm{Sch}} = |00\rangle$, $|\lambda = (2,0), q_\lambda = 0, p_\lambda = 0\rangle_{\mathrm{Sch}} = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, and $|\lambda = (2,0), q_\lambda = -1, p_\lambda = 0\rangle_{\mathrm{Sch}} = |11\rangle$. Abstractly, then, the Schur transform then corresponds to a transformation

$$\mathbf{U}_{\mathrm{Sch}} = \begin{array}{c|cccc} & |00\rangle & |01\rangle & |10\rangle & |11\rangle \\ |\lambda=(1,1), q_\lambda=0, p_\lambda=0\rangle_{\mathrm{Sch}} & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ |\lambda=(2,0), q_\lambda=+1, p_\lambda=0\rangle_{\mathrm{Sch}} & 1 & 0 & 0 & 0 \\ |\lambda=(2,0), q_\lambda=0, p_\lambda=0\rangle_{\mathrm{Sch}} & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ |\lambda=(2,0), q_\lambda=-1, p_\lambda=0\rangle_{\mathrm{Sch}} & 0 & 0 & 0 & 1 \end{array} \tag{5.19}$$

It is easy to verify that the $\lambda = (1,1)$ subspace transforms as a one dimensional irrep of $\mathcal{U}_2$ and as the alternating sign irrep of $\mathcal{S}_2$ while the $\lambda = (2,0)$ subspace transforms as a three dimensional irrep of $\mathcal{U}_2$ and as the trivial irrep of $\mathcal{S}_2$. Notice that the labeling scheme for the standard computational basis uses 2 qubits while the labeling scheme for the Schur basis uses more qubits (one such labeling assigns one qubit to $|\lambda\rangle$, none to $|p\rangle$ and two qubits to $|q\rangle$). Thus we see how padding will be necessary to directly implement the Schur transform.

To see a more complicated example of the Schur basis, let $d = 2$ and $n = 3$. There are again two valid partitions, $\lambda = (3,0)$ and $\lambda = (2,1)$. The first of these partitions labels to the trivial irrep of $\mathcal{S}_3$ and a 4 dimensional irrep of $\mathcal{U}_3$. The corresponding Schur basis vectors can be expressed as

$$\begin{aligned} |\lambda = (3,0), q_\lambda = +3/2, p_\lambda = 0\rangle_{\mathrm{Sch}} &= |000\rangle \\ |\lambda = (3,0), q_\lambda = +1/2, p_\lambda = 0\rangle_{\mathrm{Sch}} &= \frac{1}{\sqrt{3}}\left(|001\rangle + |010\rangle + |100\rangle\right) \\ |\lambda = (3,0), q_\lambda = -1/2, p_\lambda = 0\rangle_{\mathrm{Sch}} &= \frac{1}{\sqrt{3}}\left(|011\rangle + |101\rangle + |110\rangle\right) \\ |\lambda = (3,0), q_\lambda = -3/2, p_\lambda = 0\rangle_{\mathrm{Sch}} &= |111\rangle. \end{aligned} \tag{5.20}$$

The second of these partitions labels a two dimensional irrep of $\mathcal{S}_3$ and a two dimensional irrep of $\mathcal{U}_2$. Its Schur basis states can be expressed as

$$\begin{aligned} |\lambda = (2,1), q_\lambda = +1/2, p_\lambda = 0\rangle_{\mathrm{Sch}} &= \frac{1}{\sqrt{2}}\left(|100\rangle - |010\rangle\right) \\ |\lambda = (2,1), q_\lambda = -1/2, p_\lambda = 0\rangle_{\mathrm{Sch}} &= \frac{1}{\sqrt{2}}\left(|101\rangle - |011\rangle\right) \\ |\lambda = (2,1), q_\lambda = +1/2, p_\lambda = 1\rangle_{\mathrm{Sch}} &= \sqrt{\frac{2}{3}}|001\rangle - \frac{|010\rangle + |100\rangle}{\sqrt{6}} \\ |\lambda = (2,1), q_\lambda = -1/2, p_\lambda = 1\rangle_{\mathrm{Sch}} &= \sqrt{\frac{2}{3}}|110\rangle - \frac{|101\rangle + |011\rangle}{\sqrt{6}}. \end{aligned} \tag{5.21}$$

We can easily verify that Eqns. (5.20) and (5.21) indeed transform under $\mathcal{U}_2$ and $\mathcal{S}_3$ the way we expect; not so easy however is coming up with a circuit that relates this basis to the computational basis and generalizes naturally to other values of n and d. However, note that p_λ determines whether the first two qubits are in a singlet or a triplet state. This gives a hint of a recursive structure that we will exploit in Chapter 7 to construct an efficient general algorithm for the Schur transform.

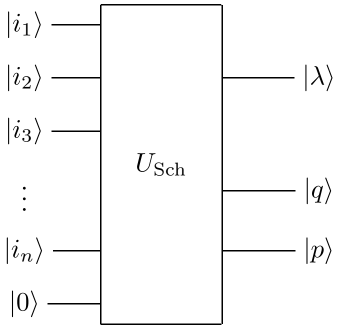


Figure 5-1: The Schur transform. Notice how the direct sum over λ in Eq. (5.16) becomes a tensor product between the $|\lambda\rangle$ register and the $|q\rangle$ and $|p\rangle$ registers. Since the number of qubits needed for $|q\rangle$ and $|p\rangle$ vary with λ, we need slightly more spatial resources, which are here denoted by the ancilla input $|0\rangle$.

5.3.1 Constructing $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$ using Schur duality

So far we have said little about the form of $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$, other than that they are indexed by partitions. It turns out that Schur duality gives a straightforward description of the irreps of $\mathcal{U}_d$ and $\mathcal{S}_n$. We will not use this explicit description to construct the Schur transform, but it is still helpful for understanding the irreps $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$. As with the rest of this chapter, proofs and further details can be found in [GW98].

We begin by expressing $\lambda \in \mathcal{I}_{d,n}$ as a Young diagram in which there are up to d rows with λ_i boxes in row i. For example, to the partition $(4, 3, 1, 1)$ we associate the diagram

$$\begin{array}{|c|c|c|c|}\hline & & & \\ \hline & & & \multicolumn{1}{c}{} \\ \cline{1-3} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\ \cline{1-1} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \quad . \tag{5.22}$$

Now we define a Young tableau T of shape λ to be a way of filling the n boxes of λ with the integers $1, \ldots, n$, using each number once and so that integers increase from left to right and from top to bottom. For example, one valid Young tableau with shape $(4, 3, 1, 1)$ is

$$\begin{array}{|c|c|c|c|}\hline 1 & 4 & 6 & 7 \\ \hline 2 & 5 & 8 & \multicolumn{1}{c}{} \\ \cline{1-3} 3 & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\ \cline{1-1} 9 & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \quad .$$

For any Young tableau T, define $\mathrm{Row}(T)$ to be set of permutations obtained by permuting the integers within each row of T; similarly define $\mathrm{Col}(T)$ to be the permutations that leave each integer in the same column of T. Now we define the *Young symmetrizer* $\Pi_{\lambda:T}$ to be an operator acting on $(\mathbb{C}^d)^{\otimes n}$

as follows:

$$\Pi_{\lambda:T} := \frac{\dim \mathcal{P}_\lambda}{n!} \left(\sum_{c \in \mathrm{Col}(T)} \mathrm{sgn}(c) \mathbf{P}(c) \right) \left(\sum_{r \in \mathrm{Row}(T)} \mathbf{P}(r) \right). \tag{5.23}$$

It can be shown that the Young symmetrizer $\Pi_{\lambda:T}$ is a projection operator whose support is a subspace isomorphic to $\mathcal{Q}_\lambda^d$. In particular $U_{\mathrm{Sch}} \Pi_{\lambda:T} U_{\mathrm{Sch}}^\dagger = |\lambda\rangle\langle\lambda| \otimes |y(T)\rangle\langle y(T)| \otimes I_{\mathcal{Q}_\lambda^d}$ for some unit vector $|y(T)\rangle \in \mathcal{P}_\lambda$. Moreover, these vectors $|y(T)\rangle$ form a basis known as Young's natural basis, though the $|y(T)\rangle$ are not orthogonal, so we will usually not work with them in quantum circuits.

Using Young symmetrizers, we can now explore some more general examples of $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$. If $\lambda = (n)$, then the only valid tableau is

$$\boxed{1}\boxed{2}\ldots\boxed{n}.$$

The corresponding $\mathcal{S}_n$-irrep $\mathcal{P}_{(n)}$ is trivial and the $\mathcal{U}_d$-irrep is given by the action of $\mathbf{Q}$ on the totally symmetric subspace of $(\mathbb{C}^d)^{\otimes n}$, i.e. $\{|v\rangle : \mathbf{P}(s)|v\rangle = |v\rangle \forall s \in \mathcal{S}_n\}$. On the other hand, if $\lambda = (1^n)$, meaning $(1, 1, \ldots, 1)$ (n times), then the only valid tableau is

$$\begin{array}{c} \boxed{1} \\ \boxed{2} \\ \vdots \\ \boxed{n} \end{array}.$$

The $\mathcal{S}_n$-irrep $\mathcal{P}_{(1^n)}$ is still one-dimensional, but now corresponds to the sign irrep of $\mathcal{S}_n$, mapping s to $\mathrm{sgn}(s)$. The $\mathcal{U}_d$-irrep $\mathcal{Q}_{(1^n)}^d$ is equivalent to the totally antisymmetric subspace of $(\mathbb{C}^d)^{\otimes n}$, i.e. $\{|v\rangle : \mathbf{P}(s)|v\rangle = \mathrm{sgn}(s)|v\rangle \forall s \in \mathcal{S}_n\}$. Note that if $d > n$, then this subspace is zero-dimensional, corresponding to the restriction that irreps of $\mathcal{U}_d$ are indexed only by partitions with $\leq d$ rows.

Other explicit examples of $\mathcal{U}_d$ and $\mathcal{S}_n$ irreps are presented from a particle physics perspective in [Geo99]. We also give more examples in Section 7.1.2, when we introduce explicit bases for $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$.

5.4 Dual reductive pairs

Schur duality can be generalized to groups other than $\mathcal{U}_d$ and $\mathcal{S}_n$. The groups for which this is possible are known as dual reductive pairs, and in this section we give an overview of their definition and properties (following Sec 9.2 of [GW98]). The next two chapters will focus primarily on Schur duality, but here we give some ideas about how the techniques used in those chapters could be applied to other groups and other representations.

Suppose G and K are groups with irreps $(\rho_\mu, U_\mu)_{\mu \in \hat{G}}$ and $(\sigma_\nu, V_\nu)_{\nu \in \hat{K}}$ respectively. Then the irreps of $G \times K$ are given by $(\rho_\mu \otimes \sigma_\nu, U_\mu \hat{\otimes} V_\nu)$. Now suppose (γ, Y) is a representation of $G \times K$. Its isotypic decomposition (cf. Eq. (5.2)) is of the form

$$Y \overset{G \times K}{\cong} \bigoplus_{\mu \in \hat{G}} \bigoplus_{\nu \in \hat{K}} U_\mu \hat{\otimes} V_\nu \otimes \mathbb{C}^{m_{\mu,\nu}}, \tag{5.24}$$

where the $m_{\mu,\nu}$ are multiplicity factors. Define the algebras $\mathcal{A} = \gamma(\mathbb{C}[G \times \{e\}])$ and $\mathcal{B} = \gamma(\mathbb{C}[\{e\} \times K])$. Then [GW98] proves the following generalization of Schur duality:

Proposition 5.1. *The following are equivalent:*

(1) Each $m_{\mu,\nu}$ is either 0 or 1, and at most one $m_{\mu,\nu}$ is nonzero for each μ and each ν. In other

words, Eq. (5.24) has the form

$$W \overset{G\times K}{\cong} \bigoplus_{\lambda\in S} U_{\varphi_G(\lambda)} \hat{\otimes} V_{\varphi_K(\lambda)} \tag{5.25}$$

where S is some set and $\varphi_G : S \to \hat{G}, \varphi_K : S \to \hat{K}$ are injective maps.

(2) $\mathcal{B}$ is the commutant of $\mathcal{A}$ in $\operatorname{End}(V)$ *(i.e. $\mathcal{B} = \{x \in End(V) : [x, a] = 0\, \forall a \in \mathcal{A}\}$) and $\mathcal{A}$ is the commutant of $\mathcal{B}$. When this holds we say that $\mathcal{A}$ and $\mathcal{B}$ are* double commutants.

When these conditions hold we say that the groups $\gamma(G \times \{e\})$ and $\gamma(\{e\} \times K)$ form a *dual reductive pair*. In this case, Eq. (5.25) gives us a one-to-one correspondence between the subsets of $\hat{G}$ and $\hat{K}$ that appear in W. This redundancy can often be useful. For example, measuring the G-irrep automatically also measures the K-irrep. In fact, the key idea behind the algorithms we will encounter in Chapters 7 and 8 is that the Schur transform can be approached by working only with $\mathcal{U}_d$-irreps or only with $\mathcal{S}_n$ irreps.

Many of the examples of dual reductive pairs that are known relate to the orthogonal and symplectic Lie groups[How89], and so are not immediately applicable to quantum information. However, in this section we will point out one example of a dual reductive pair that could arise naturally when working with quantum states.

Let $G = \mathcal{U}_{d_A}$ and $K = \mathcal{U}_{d_B}$ and define W to be the n^{th} symmetric product of $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$; i.e.

$$W := \left((\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})^{\otimes n}\right)^{\mathcal{S}_n} = \left\{|v\rangle \in \left(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}\right)^{\otimes n} : \mathbf{P}(s)|v\rangle = |v\rangle \forall s \in \mathcal{S}_n\right\}. \tag{5.26}$$

We have seen in Section 5.3.1 that $W \overset{\mathcal{U}_{d_A d_B}}{\cong} \mathcal{Q}^{d_A d_B}_{(1^n)}$. However, here we are interested in the action of $\mathcal{U}_{d_A} \times \mathcal{U}_{d_B}$ on W, which we define in the natural way; i.e. (U_A, U_B) is mapped to $(U_A \otimes U_B)^{\otimes n}$. It is straightforward to show that $\mathcal{U}_{d_A}$ and $\mathcal{U}_{d_B}$ generate algebras that are double commutants. This means that W decomposes under $\mathcal{U}_{d_A} \times \mathcal{U}_{d_B}$ as

$$W \overset{\mathcal{U}_{d_A}\times\mathcal{U}_{d_B}}{\cong} \bigoplus_{\lambda\in\mathcal{I}_{d,n}} \mathcal{Q}^{d_A}_{\lambda} \hat{\otimes} \mathcal{Q}^{d_B}_{\lambda}, \tag{5.27}$$

where $d = \min(d_A, d_B)$. This yields several nontrivial conclusions. For example, if the system were shared between two parties, then this would mean that the states of both parties would have the same Young frame. Also, it turns out that applying the Schur transform circuit in Section 7.2 to either A or B gives an efficient method for performing the isomorphism in Eq. (5.27).

The implications of other dual reductive pairs for quantum information are largely unknown. However, in principle they offer far-ranging generalizations of Schur duality that remain amenable to manipulation by the same sorts of algorithms.

Chapter 6

Applications of the Schur transform to quantum information theory

In physics, the Schur basis is a natural way to study systems with permutation symmetry. In quantum information theory, the Schur basis is well suited to i.i.d. states and channels, such as $\rho^{\otimes n}$ and $\mathcal{N}^{\otimes n}$. For example, if ρ is a $d \times d$ density matrix, then $\rho^{\otimes n}$ decomposes under the Schur transform as

$$U_{\text{Sch}}\rho^{\otimes n}U_{\text{Sch}}^{\dagger} = \sum_{\lambda \in \mathcal{I}_{d,n}} |\lambda\rangle\langle\lambda| \otimes \mathbf{q}_{\lambda}^{d}(\rho) \otimes I_{\mathcal{P}_{\lambda}}. \tag{6.1}$$

To prove this, and to interpret the $\mathbf{q}_{\lambda}^{d}(\rho)$ term, we note that irreps of $\mathcal{U}_d$ can also be interpreted as irreps of GL_d (the group of $d \times d$ complex invertible matrices)*. If ρ is not an invertible matrix then we can still express ρ as a limit of elements of GL_d and can use the continuity of $\mathbf{q}_{\lambda}^{d}$ to define $\mathbf{q}_{\lambda}^{d}(\rho)$. Then Eq. (6.1) follows from Eq. (5.18).

The rest of the chapter will explore the implications of Eq. (6.1) and related equations. We will see that the first register, $|\lambda\rangle$, corresponds to the spectrum of ρ, and indeed that a good estimate for the spectrum of ρ is given by measuring $|\lambda\rangle$ and guessing $(\lambda_1/n, \ldots, \lambda_d/n)$ for the spectrum. The $\mathcal{Q}_{\lambda}^{d}$ register depends on the spectrum (λ) for its structure, but itself contains information only about the eigenbasis of ρ. Both of these registers are vanishingly small—on the order of $d^2 \log n$ qubits—but contain all the features of ρ. The $\mathcal{P}_{\lambda}$ register contains almost all the entropy, but always carries a uniform distribution that is independent of ρ once we condition on λ.

This situation can be thought of as generalization of the classical method of types, a technique in information theory in which strings drawn from i.i.d. distributions are classified by their empirical distributions. We give a brief review of this method in Section 6.1 so that the reader will be able to appreciate the similarities with the quantum case. In Section 6.2, we show how Eq. (6.1) leads to a quantum method of types, and give quantitative bounds to make the theory useful. We survey known applications of Schur duality to quantum information theory in Section 6.3, using our formulae from Section 6.2 to give concise proofs of the some of the main results from the literature. Finally, we show how Schur duality may be used to decompose i.i.d. quantum channels in Section 6.4.

Only the last section represents completely new work. The idea of Schur duality as a quantum method of types has been known for years, beginning with applications to quantum hypothesis testing[Hay01] and spectrum estimation[KW01], further developed in a series of papers by Hayashi and Matsumoto[HM01, HM02c, HM02a, HM02b, Hay02b, Hay02a], extended to other applications in [Bac01, KBLW01, BRS03, BRS04, vKK04, HHH05], and recently applied to information theory in

*This is because GL_d is the complexification of $\mathcal{U}_d$, meaning that its Lie algebra (the set of all $d \times d$ complex matrices) is equal to the tensor product of $\mathbb{C}$ with the Lie algebra of $\mathcal{U}_d$ (the set of $d \times d$ Hermitian matrices). See [GW98, FH91] for more details. For this reason, mathematicians usually discuss the representation theory of GL_d instead of $\mathcal{U}_d$.

[CM04]. The contribution of the first three sections is to present these results together as applications of the same general method.

6.1 The classical method of types

The method of types is a powerful tool in classical information theory. Here we briefly review the method of types (following [CT91, CK81]) to give an idea of how the Schur basis will later be used for the quantum generalization.

Consider a string $x^n = (x_1, \ldots, x_n) \in [d]^n$, where $[d] := \{1, \ldots, d\}$. Define the type of x^n to be the d-tuple of integers $t(x^n) := \sum_{j=1}^n e_{x_j}$, where $e_i \in \mathbb{Z}^d$ is the unit vector with a one in i$^{\text{th}}$ position. Thus $t(x^n)$ counts the frequency of each symbol $1, \ldots, d$ in x^n. Let $\mathcal{T}_d^n := \{(n_1, \ldots, n_d) : n_1 + \ldots + n_d = n, n_i \geq 0\}$ denote the set of all possible types of strings in $[d]^n$ (also known as the weak d-compositions of n). Since an element of $\mathcal{T}_d^n$ can be written as d numbers ranging from $0, \ldots, n$ we obtain the simple bound $|\mathcal{T}_d^n| \leq (n+1)^d$. In fact, $|\mathcal{T}_d^n| = \binom{n+d-1}{d-1}$, but knowing the exact number is rarely necessary. For a type t, let the normalized probability distribution $\bar{t} := t/n$ denote its empirical distribution.

The set of types $\mathcal{T}_d^n$ is larger than the set of partitions $\mathcal{I}_{d,n}$ because symbol frequencies in types do not have to occur in decreasing order. In principle, we could separate a type $t \in \mathcal{T}_d^n$ into a partition $\lambda \in \mathcal{I}_{d,n}$ (with nonincreasing parts) and a mapping of the parts of λ onto $[d]$, which we call q_λ. The map q_λ corresponds to some $(a_1, \ldots, a_d) \in \mathcal{S}_d$ for which there are λ_i symbols equal to a_i for each $i \in \{1, \ldots, d\}$. However, if not all the λ_i are distinct, then this is more information than we need. In particular, if $\lambda_i = \lambda_{i+1} = \ldots = \lambda_j$, then we don't care about the ordering of $a_i, \ldots, a_j$. Define $m_i(\lambda)$ to be the number of parts of λ equal to i, i.e. $|\{j : \lambda_j = i\}|$. Then the number of distinct q_λ is $d!/m_1! \ldots m_n! =: \binom{d}{m}$. This separation is not usually used for classical information theory, but helps show how the quantum analogue of type is split among the $|\lambda\rangle$ and $|q\rangle$ registers.

For a particular type $t \in \mathcal{T}_d^n$, denote the set of all strings in $[d]^n$ with type t by $T_t = \{x^n \in [d^n] : t(x^n) = t\}$. There are two useful facts about T_t. First, $|T_t| = \binom{n}{t} := n!/t_1! \ldots t_d!$ (or equivalently $|T_t| = \binom{n}{\lambda}$, where λ is a sorted version of t). Second, let P be a probability distribution on $[d]$ and $P^{\otimes n}$ the probability distribution on $[d]^n$ given by n i.i.d. copies of P, i.e. $P^{\otimes n}(x^n) := P(x_1) \cdots P(x_n)$. Then for any $x^n \in T_t$ we have $P^{\otimes n}(x^n) = P(1)^{t_1} \cdots P(d)^{t_d} = \exp(\sum_{j=1}^d t_j \log P(j))$. This has a natural expression in terms of the entropic quantities $H(\bar{t}) := -\sum_j \bar{t}_j \log \bar{t}_j$ and $D(\bar{t}\|P) := \sum_j \bar{t}_j \log \bar{t}_j / P(j)$ as

$$P^{\otimes n}(x^n) = \exp\left(-n\left(H(\bar{t}) + D(\bar{t}\|P)\right)\right). \tag{6.2}$$

These basic facts can be combined with simple probabilistic arguments to prove many results in classical information theory. For example, if we define $P^{\otimes n}(T_t) := \sum_{x^n \in T_t} P^{\otimes n}(x^n)$, then

$$\bar{t}^{\otimes n}(T_t) = \binom{n}{t} \exp(-nH(\bar{t})). \tag{6.3}$$

Since $\bar{t}^{\otimes n}(T_t) \leq 1$, we get the bound $\binom{n}{t} \leq \exp(nH(\bar{t}))$. On the other hand, by doing a bit of algebra[CT91] one can show that $\bar{t}^{\otimes n}(T_t) \geq \bar{t}^{\otimes n}(T_{t'})$ for any $t' \in \mathcal{T}_d^n$; i.e. under the probability distribution $\bar{t}^{\otimes n}$, the most likely type is t. This allows us to lower bound $\binom{n}{t}$ by $\exp(nH(\bar{t}))/|\mathcal{T}_{d,n}|$. Together these bounds are

$$(n+1)^{-d} \exp(nH(\bar{t})) \leq |T_t| = \binom{n}{t} \leq \exp(nH(\bar{t})). \tag{6.4}$$

Combining Eqns. (6.4) and (6.2) for an arbitrary distribution P then gives

$$(n+1)^{-d} \exp\left(-nD(\bar{t}\|P)\right) \leq P^{\otimes n}(T_t) \leq \exp\left(-nD(\bar{t}\|P)\right), \tag{6.5}$$

Thus, as n grows large, we are likely to observe an empirical distribution $\bar{t}$ that is close to the actual distribution P. To formalize this, define the set of *typical sequences* $T^n_{P,\delta}$ by

$$T^n_{P,\delta} := \bigcup_{\substack{t\in\mathcal{T}_{d,n} \\ \|\bar{t}-P\|_1\leq\delta}} T_t. \tag{6.6}$$

To bound $P^{\otimes n}(T^n_{P,\delta})$, we apply Pinsker's inequality[Pin64]:

$$D(Q\|P) \geq \tfrac{1}{2}\|P-Q\|_1^2. \tag{6.7}$$

Denote the complement of $T^n_{P,\delta}$ by $[d]^n - T^n_{P,\delta}$. Then

$$P^{\otimes n}([d]^n - T^n_{P,\delta}) = \sum_{\substack{t\in\mathcal{T}_{d,n} \\ \|\bar{t}-P\|_1>\delta}} P^{\otimes n}(T_t) \leq \sum_{\substack{t\in\mathcal{T}_{d,n} \\ \|\bar{t}-P\|_1>\delta}} \exp\left(-nD(\bar{t}\|P)\right) \leq (n+1)^d \exp\left(-\frac{n\delta^2}{2}\right) \tag{6.8}$$

and therefore

$$P^{\otimes n}(T^n_{P,\delta}) \geq 1-(n+1)^d \exp\left(-\frac{n\delta^2}{2}\right). \tag{6.9}$$

This has several useful consequences:

- *Estimating the probability distribution P:* If the true probability distribution of an i.i.d. process is P and we observe empirical distribution $\bar{t}$ on n samples, the probability that $\|t-P\|_1 > \delta$ is $\leq (n+1)^d \exp\left(-\frac{n\delta^2}{2}\right)$, which decreases exponentially with n for any constant value of δ.

- *Data compression (cf. Eq. (1.28)):* We can compress n letters from an i.i.d. source with distribution P by transmitting only strings in $T^n_{P,\delta}$. Asymptotically, the probability of error is $\leq (n+1)^d \exp\left(-\frac{n\delta^2}{2}\right)$, which goes to zero as $n\to\infty$. The number of bits required is $\lceil \log|T^n_{P,\delta}|\rceil$. To estimate this quantity, use Fannes' inequality (Lemma 1.1) to bound

$$|H(\bar{t}) - H(P)| \leq \eta(\delta) + \delta\log d \tag{6.10}$$

  whenever $\|\bar{t}-P\|_1 \leq \delta$. Thus

$$\log|T^n_{P,\delta}| \leq \log|\mathcal{T}_{d,n}| + n\left[H(P)+\eta(\delta)+\delta\log d\right] \leq n\left[H(P)+\eta(\delta)+\delta\log d + \frac{d}{n}\log(n+1)\right], \tag{6.11}$$

  which asymptotically approaches $H(P)$ bits per symbol.

- *Randomness concentration (cf. Eq. (1.29)):* Suppose we are given a random variable x^n distributed according to $P^{\otimes n}$ and wish to produce from it some uniformly distributed random bits. Then since all x^n with the same type have the same probability, conditioning on the type $t = t(x^n)$ is sufficient to give a uniformly distributed random variable. According to Eqns. (6.10) and (6.9), this yields $\geq n(H(P)-\eta(\delta)-\delta\log d) = n(H(P)-o(1))$ bits with probability that asymptotically approaches one.

If we have two random variables X and Y with a joint probability distribution $P(X,Y)$, then we can define joint types and jointly typical sequences. These can be used to prove more sophisticated results, such as Shannon's noisy coding theorem[Sha48] and the Classical Reverse Shannon Theorem[BSST02, Win02]. Reviewing classical joint types would take us too far afield, but Section 6.4 will develop a quantum analogue of joint types which can be applied to channels or noisy bipartite states.

Let us now summarize in a manner that shows the parallels with the quantum case. A string $x^n \in [d]^n$ can be expressed as a triple $(\lambda, q_\lambda, p_\lambda)$ where $\lambda \in \mathcal{I}_{d,n}$, $q_\lambda \in Q_\lambda$ and $p_\lambda \in P_\lambda$ for sets Q_λ and P_λ satisfying $|Q_\lambda| \leq \operatorname{poly}(n)$ and $\exp(nH(\overline{\lambda}))/\operatorname{poly}(n) \leq |P_\lambda| \leq \exp(nH(\overline{\lambda}))$, if we think of d as a constant. Furthermore, permuting x^n with an element of $\mathcal{S}_n$ affects only the p_λ register and for $f \in \mathcal{S}_d$, the map $x^n \to (f(x_1), \ldots, f(x_n))$ affects only the q_λ register. This corresponds closely with the quantum situation in Eq. (6.1). We now show how dimension counting in the quantum case resembles the combinatorics of the classical method of types.

6.2 Schur duality as a quantum method of types

In this section, we generalize the classical method of types to quantum states. Our goal is to give asymptotically tight bounds on $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$ and the other quantities appearing in Eq. (6.1)(following [GW98, Hay02a, CM04]).

First recall that $|\mathcal{I}_{d,n}| \leq |\mathcal{T}_{d,n}| = (n+1)^d = \operatorname{poly}(n)$. For $\lambda \in \mathcal{I}_{d,n}$, define $\widetilde{\lambda} := \lambda + (d-1, d-2, \ldots, 1, 0)$. Then the dimensions of $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$ are given by[GW98]

$$\dim \mathcal{Q}_\lambda^d = \frac{\prod_{1\leq i<j\leq d}(\widetilde{\lambda}_i - \widetilde{\lambda}_j)}{\prod_{m=1}^d m!} \tag{6.12}$$

$$\dim \mathcal{P}_\lambda = \frac{n!}{\widetilde{\lambda}_1!\widetilde{\lambda}_2!\cdots\widetilde{\lambda}_d!}\prod_{1\leq i<j\leq d}(\widetilde{\lambda}_i - \widetilde{\lambda}_j) \tag{6.13}$$

It is straightforward to bound these by[Hay02a, CM04]

$$\dim \mathcal{Q}_\lambda^d \leq (n+d)^{d(d-1)/2} \tag{6.14}$$

$$\binom{n}{\lambda}(n+d)^{-d(d-1)/2} \leq \dim \mathcal{P}_\lambda \leq \binom{n}{\lambda}. \tag{6.15}$$

Applying Eq. (6.4) to Eq. (6.15) yields the more useful

$$\exp\left(nH(\overline{\lambda})\right)(n+d)^{-d(d+1)/2} \leq \dim \mathcal{P}_\lambda \leq \exp\left(nH(\overline{\lambda})\right). \tag{6.16}$$

We can use Eq. (6.1) to derive a quantum analogue of Eq. (6.2). To do so, we will need to better describe the structure of $\mathcal{Q}_\lambda^d$. Define the torus $\mathcal{U}_1^{\times d} = \mathcal{U}_1 \times \ldots \mathcal{U}_1 \subset \mathcal{U}_d$ as the subgroup of diagonal matrices (in some fixed basis of $\mathbb{C}^d$). For $x \in \mathbb{C}^d$ let $\operatorname{diag}(x)$ denote the diagonal matrix with entries $x_1, \ldots, x_d$. The (one-dimensional) irreps of $\mathcal{U}_1^{\times d}$ are labeled by $\mu \in \mathbb{Z}^d$ and are given by $x^\mu := x_1^{\mu_1}\cdots x_d^{\mu_d}$. We will be interested only in μ with nonnegative entries, and we write $\mathbb{Z}_+^d$ to denote this set (note that this is different from $\mathbb{Z}_{++}^d$ because the components of μ can be in any order).

If $(\mathbf{q}, \mathcal{Q})$ is a polynomial representation of $\mathcal{U}_d$, then upon restriction to $\mathcal{U}_1^{\times d}$ one can show that it breaks up into orthogonal subspaces labeled by different $\mu \in \mathbb{Z}_+^d$. The subspace corresponding to the $\mathcal{U}_1^{\times d}$-representation μ is called the μ-weight space of $\mathcal{Q}$ and is denoted $\mathcal{Q}(\mu)$. Formally, we can define $\mathcal{Q}(\mu) \subset \mathcal{Q}$ by $\mathcal{Q}(\mu) := \{|q\rangle \in \mathcal{Q} : \mathbf{q}(\operatorname{diag}(x_1, \ldots, x_d))|q\rangle = x_1^{\mu_1}\cdots x_d^{\mu_d}|q\rangle \ \forall x_1, \ldots, x_d \in \mathbb{C}\backslash\{0\}\}$. For example $(\mathbb{C}^d)^{\otimes n}(\mu) = \operatorname{Span}\{|x^n\rangle : x^n \in T_\mu\}$.

To describe the weight spaces of $\mathcal{Q}_\lambda^d$ we define the Kostka coefficient $K_{\lambda\mu} := \dim \mathcal{Q}_\lambda^d(\mu)$ (as can be easily checked, $K_{\lambda\mu}$ depends on d only through λ and μ)[GW98]. While no useful formula is known for $K_{\lambda\mu}$, they do satisfy

- $\sum_\mu K_{\lambda\mu} = \dim \mathcal{Q}_\lambda^d$
- $K_{\lambda\mu} \neq 0$ if and only if $\mu \prec \lambda$, meaning that $|\mu| = |\lambda|$ and $\sum_{i=1}^c \mu_i \leq \sum_{i=1}^c \lambda_i$ for $c = 1, \ldots, d-1$.

- $K_{\lambda\lambda} = 1$

If we order weights according to the majorization relation $\prec$, then there exists a *highest-weight vector* spanning the one-dimensional space $\mathcal{Q}_\lambda^d(\lambda)$. At the risk of some ambiguity, we call this vector $|\lambda\rangle$. We will also define an orthonormal basis for $\mathcal{Q}_\lambda^d$, denoted Q_λ^d, in which each basis vector lies in a single weight space. This is clearly possible in general, and also turns out to be consistent with the basis we will introduce in Section 7.1.2 for use in quantum algorithms. To simplify notation later on, whenever we work with a particular density matrix ρ, we will choose the torus $\mathcal{U}_1^{\times d}$ to be diagonal with respect to the same basis as ρ. This means that $\mathbf{q}_\lambda^d(\rho)$ is diagonalized by Q_λ^d, the induced weight basis for $\mathcal{Q}_\lambda^d$.

We now have all the tools we need to find the spectrum of $\mathbf{q}_\lambda^d(\rho)$. Let the eigenvalues of ρ be given by $r_1 \geq \cdots \geq r_d$ (we sometimes write $r = \operatorname{spec}\rho$). Then for all $\mu \in \mathcal{T}_d^n$, $\mathbf{q}_\lambda^d(\rho)$ has eigenvector $r^\mu = r_1^{\mu_1} \cdots r_d^{\mu_d}$ with multiplicity $K_{\lambda\mu}$. The highest eigenvalue is $r^\lambda = \exp[-n(H(\overline{\lambda}) + D(\overline{\lambda}\|r))]$ (since r is nonincreasing and $\mu \prec \lambda$ for any μ with $K_{\lambda\mu} \neq 0$). Thus we obtain the following bounds on $\operatorname{Tr}\mathbf{q}_\lambda^d(\rho)$:

$$r^\lambda \leq \operatorname{Tr}\mathbf{q}_\lambda^d(\rho) = \sum_\mu K_{\lambda\mu} r^\mu \leq r^\lambda \dim \mathcal{Q}_\lambda^d. \tag{6.17}$$

To relate this to quantum states, let Π_λ denote the projector onto $\mathcal{Q}_\lambda^d \otimes \mathcal{P}_\lambda \subset (\mathbb{C}^d)^{\otimes n}$. Explicitly Π_λ is given by

$$\Pi_\lambda = U_{\text{Sch}}^\dagger \left(|\lambda\rangle\langle\lambda| \otimes I_{\mathcal{Q}_\lambda^d} \otimes I_{\mathcal{P}_\lambda} \right) U_{\text{Sch}}. \tag{6.18}$$

From the bounds on $\dim \mathcal{Q}_\lambda^d$ and $\dim \mathcal{P}_\lambda$ in Eqns. (6.14) and (6.16), we obtain

$$\exp\left(nH(\overline{\lambda})\right)(n+d)^{-d(d+1)/2} \leq \operatorname{Tr}\Pi_\lambda \leq \exp\left(nH(\overline{\lambda})\right)(n+d)^{d(d-1)/2} \tag{6.19}$$

Also $\operatorname{Tr}\Pi_\lambda\rho^{\otimes n}\Pi_\lambda = \operatorname{Tr}\mathbf{q}_\lambda^d(\rho) \cdot \dim \mathcal{P}_\lambda$, which can be bounded by

$$\exp\left(-nD(\overline{\lambda}\|r)\right)(n+d)^{-d(d+1)/2} \leq \operatorname{Tr}\Pi_\lambda\rho^{\otimes n}\Pi_\lambda \leq \exp\left(-nD(\overline{\lambda}\|r)\right)(n+d)^{d(d-1)/2} \tag{6.20}$$

Similarly, we have

$$\Pi_\lambda\rho^{\otimes n} = \rho^{\otimes n}\Pi_\lambda = \Pi_\lambda\rho^{\otimes n}\Pi_\lambda \leq r^\lambda\Pi_\lambda = \exp[-n(H(\overline{\lambda}) + D(\overline{\lambda}\|r))]\Pi_\lambda. \tag{6.21}$$

For some values of μ, r^μ can be much smaller, so we cannot express any useful lower bound on the eigenvalues of $\Pi_\lambda\rho^{\otimes n}\Pi_\lambda$, like we can with classical types. Of course, tracing out $\mathcal{Q}_\lambda^d$ gives us a maximally mixed state in $\mathcal{P}_\lambda$, and this is the quantum analogue of the fact that $P^{\otimes n}(\cdot|t)$ is uniformly distributed over T_t.

We can also define the typical projector

$$\Pi_{r,\delta}^n = \sum_{\lambda:\overline{\lambda}\in\mathcal{B}_\delta(r)} \Pi_\lambda = U_{\text{Sch}}^\dagger \left[\sum_{\lambda:\overline{\lambda}\in\mathcal{B}_\delta(r)} |\lambda\rangle\langle\lambda| \otimes I_{\mathcal{Q}_\lambda^d} \otimes I_{\mathcal{P}_\lambda} \right] U_{\text{Sch}}, \tag{6.22}$$

where $\mathcal{B}_\delta(r) := \{\overline{\lambda} : \|\overline{\lambda} - r\|_1 \leq \delta\}$. Using Pinsker's inequality, we find that

$$\operatorname{Tr}\Pi_{r,\delta}^n\rho^{\otimes n} \geq 1 - \exp\left(-\frac{n\delta^2}{2}\right)(n+d)^{d(d+1)/2}, \tag{6.23}$$

similar to the classical case. The typical subspace is defined to be the support of the typical projector. Its dimension can be bounded (using Eqns. (6.23) and (6.10)) by

$$\operatorname{Tr}\Pi_{r,\delta}^n \leq |\mathcal{I}_{d,n}| \max_{\overline{\lambda}\in\mathcal{B}_\delta(r)} \operatorname{Tr}\Pi_\lambda \leq (n+d)^{d(d+1)/2}\exp(nH(r) + \eta(\delta) + \delta\log d), \tag{6.24}$$

which is sufficient to derive Schumacher compression (cf. Eq. (1.24)).

The bounds described in this section are fairly simple, but are already powerful enough to derive many results in quantum information theory. Before discussing those applications, we will describe a variation of the decomposition of $\rho^{\otimes n}$ given in Eq. (6.1). Suppose we are given n copies of a pure state $|\psi\rangle^{AB}$ where $\rho^A = \operatorname{Tr}_B \psi^{AB}$. This situation also arises when we work in the CP formalism (see Section 1.1.1). Purifying both sides of Eq. (6.1) then gives us the alternate decomposition

$$(U^A_{\text{Sch}} \otimes U^B_{\text{Sch}})(|\psi\rangle^{AB})^{\otimes n} = \sum_{\lambda \in \mathcal{I}_{d,n}} c_\lambda |\lambda\rangle^{A_1} |\lambda\rangle^{B_1} \otimes |q_\lambda\rangle^{A_2 B_2} \otimes |\Phi_{\mathcal{P}_\lambda}\rangle^{A_3 B_3} \tag{6.25}$$

Here c_λ are coefficients satisfying $|c_\lambda|^2 = \operatorname{Tr}\Pi_\lambda \rho^{\otimes n}$, $|q_\lambda\rangle$ are arbitrary states in $\mathcal{Q}^d_\lambda$ and $|\Phi_{\mathcal{P}_\lambda}\rangle$ is a maximally entangled state* on $\mathcal{P}_\lambda \otimes \mathcal{P}_\lambda$.

6.3 Applications of Schur duality

The Schur transform is useful in a surprisingly large number of quantum information protocols. Here we will review these applications using the formulae from the last section to rederive the main results. It is worth noting that an efficient implementation of the Schur transform is the only nontrivial step necessary to perform these protocols. Thus our construction of the Schur transform in the next chapter will simultaneously make all of these tasks computationally efficient.

Spectrum and state estimation

Suppose we are given many copies of an unknown mixed quantum state, $\rho^{\otimes n}$. An important task is to obtain an estimate for the spectrum of ρ from these n copies. An asymptotically good estimate (in the sense of large deviation rate) for the spectrum of ρ can be obtained by applying the Schur transform, measuring λ and taking the spectrum estimate to be $(\lambda_1/n, \ldots, \lambda_d/n)$[KW01, VLPT99]. Indeed the probability that $\|\lambda - \operatorname{spec}\rho\|_1 \leq \delta$ for any $\delta > 0$ is bounded by Eq. (6.23). Thus an efficient implementation of the Schur transform will efficiently implement the spectrum estimating protocol (note that it is efficient in d, not in $\log(d)$).

The more general problem of estimating ρ reduces to measuring $|\lambda\rangle$ and $\mathcal{Q}^d_\lambda$, but optimal estimators have only been explicitly constructed for the case of $d = 2$[GM02]. One natural estimation scheme is given by first measuring λ and then performing a covariant POVM on $\mathcal{Q}^d_\lambda$ with POVM elements

$$\mathbf{q}^d_\lambda(U)\, |\lambda\rangle\langle\lambda|\, \mathbf{q}^d_\lambda(U)^\dagger \dim \mathcal{Q}^d_\lambda \, dU, \tag{6.26}$$

where $|\lambda\rangle$ is the highest weight vector in $\mathcal{Q}^d_\lambda$ and dU is a Haar measure for $\mathcal{U}_d$. The corresponding state estimate is then $\hat{\rho} = U\left(\sum_{i=1}^d \lambda_i \,|i\rangle\langle i|\right) U^\dagger$. In this estimation scheme, as $n \to \infty$ the probability that $\|\rho - \hat{\rho}\|_1 > \delta$ scales as $\exp(-nf(\delta))$ with $f(\delta) > 0$ whenever $\delta > 0$; [Key04] proves this and derives the function $f(\delta)$. However, it is not known whether the $f(\delta)$ obtained for this measurement scheme is the best possible.

A related problem is quantum hypothesis testing (determining whether one has been given the state $\rho^{\otimes n}$ or some other state). An optimal solution to quantum hypothesis testing can be obtained by a similar protocol[Hay02b].

Universal distortion-free entanglement concentration

Let $|\psi\rangle_{AB}$ be a bipartite partially entangled state shared between two parties, A and B. Suppose we are given many copies of $|\psi\rangle_{AB}$ and we want to transform these states into copies of a maximally entangled state using only local operations and classical communication. Further, suppose that we

*In fact, we will see in Section 6.4.1 that $|\Phi_{\mathcal{P}_\lambda}\rangle$ is uniquely determined.

wish this protocol to work when neither A nor B know the state $|\psi\rangle^{AB}$. Such a scheme is called a universal (meaning it works with unknown states $|\psi\rangle^{AB}$) entanglement concentration protocol, as opposed to the original entanglement concentration protocol described by [BBPS96]. Further we also would like the scheme to produce perfect maximally entangled states, i.e. to be distortion free. Universal distortion-free entanglement concentration can be performed[HM02c] by both parties performing Schur transforms on their n halves of $|\psi\rangle^{AB}$, measuring their $|\lambda\rangle$, discarding $\mathcal{Q}_\lambda^d$ and retaining $\mathcal{P}_\lambda$. According to Eq. (6.25), the two parties will now share a maximally entangled state of dimension $\dim \mathcal{P}_\lambda$, where λ is observed with probability $\dim \mathcal{P}_\lambda \cdot \operatorname{Tr} \mathbf{q}_\lambda^d(\operatorname{Tr}_B |\psi\rangle\langle\psi|)$.

According to Eqns. (6.23), (6.10) and (6.19), this produces at least $n(S(\rho) - \eta(\delta) - \delta \log d) - \frac{1}{2}d(d+1)\log(n+d)$ ebits with probability $\geq 1 - \exp\left(-\frac{n\delta^2}{2}\right)(n+d)^{d(d+1)/2}$. The rate at which this error probability vanishes for any fixed δ can be shown to be optimal among protocols of this form[HM02c].

Universal Compression with Optimal Overflow Exponent

Measuring $|\lambda\rangle$ weakly so as to cause little disturbance, together with appropriate relabeling, comprises a universal compression algorithm with optimal overflow exponent (rate of decrease of the probability that the algorithm will output a state that is much too large)[HM02a, HM02b].

Alternatively, suppose we are given R s.t. $H(\rho) < R$ and we want to compress $\rho^{\otimes n}$ into nR qubits. Define the projector Π_R^n by

$$\Pi_R^n := \sum_{\substack{\lambda \in \mathcal{I}_{d,n} \\ H(\bar{\lambda}) \leq R_n}} \Pi_\lambda, \tag{6.27}$$

where $R_n := R - \frac{1}{2}d(d+1)\log(n+d)$. Since $\operatorname{Tr} \Pi_R^n \leq \exp(nR)$, projecting onto Π_R^n allows the residual state to be compressed to nR qubits. The error can be shown to be bounded by

$$\leq (n+d)^{d(d+1)/2} \exp\left[-n \min_{P: H(P) > R_n} D(P \| \operatorname{spec} \rho)\right], \tag{6.28}$$

which decreases exponentially with n as long as $R > H(\rho)$.

Encoding and decoding into decoherence-free subsystems

Further applications of the Schur transform include encoding into decoherence-free subsystems[ZR97, KLV00, KBLW01, Bac01]. Decoherence-free subsystems are subspaces of a system's Hilbert space which are immune to decoherence due to a symmetry of the system-environment interaction. For the case where the environment couples identically to all systems, information can be protected from decoherence by encoding into the $|p_\lambda\rangle$ basis. We can use the inverse Schur transform (which, as a circuit can be implemented by reversing the order of all gate elements and replacing them with their inverses) to perform this encoding: simply feed in the appropriate $|\lambda\rangle$ with the state to be encoded into the $\mathcal{P}_\lambda$ register and any state into the $\mathcal{Q}_\lambda^d$ register into the inverse Schur transform. Decoding can similarly be performed using the Schur transform.

This encoding has no error and asymptotically unit efficiency, since $\log \max_\lambda \dim \mathcal{P}_\lambda$ qubits can be sent and $\max_\lambda \dim \mathcal{P}_\lambda \geq d^n / (|\mathcal{I}_{d,n}| \max_\lambda \dim \mathcal{Q}_\lambda^d) \geq d^n (n+d)^{-d(d+1)/2}$.

Communication without a shared reference frame

An application of the concepts of decoherence-free subsystems comes about when two parties wish to communicate (in either a classical or quantum manner) but do not share a reference frame. The effect of not sharing a reference frame is the same as the effect of collective decoherence: the same random unitary rotation is applied to each subsystem. Thus encoding information into the $\mathcal{P}_\lambda$ register will allow this information to be communicated in spite of the fact that the two parties do not share a reference frame[BRS03]. Just as with decoherence-free subsystems, this encoding and decoding can be done with the Schur transform.

6.4 Normal form of memoryless channels

So far we have has only discussed the decomposition of $\rho^{\otimes n}$, or equivalently, of pure bipartite entangled states. However many interesting problems in quantum information theory involve what are effectively tripartite states. Not only are tripartite states $|\psi\rangle^{ABC}$ interesting in themselves[Tha99], they also appear when a noisy bipartite state ρ^{AB} is replaced by its purification $|\psi\rangle^{ABE}$ and when a noisy quantum channel $\mathcal{N}^{A\to B}$ is replaced by its purification $U_{\mathcal{N}}^{A\to BE}$. When considering n copies of these resources, much of their structure can be understood in terms of the vector spaces $(\mathcal{P}_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$. We explain how this follows from the $\mathcal{S}_n$ CG transform in Section 6.4.1 and then apply this to quantum channels in Section 6.4.2. Finally, we generalize the bounds from Section 6.2 to a quantum analogue of joint typicality in Section 6.4.3.

6.4.1 The $\mathcal{S}_n$ Clebsch-Gordan transformation

We begin by describing how the CG transform (cf. Section 5.2.2) specializes to $\mathcal{S}_n$. For $\lambda_A,\lambda_B\in\mathcal{I}_n$, Eq. (5.4) implies

$$\mathcal{P}_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\overset{\mathcal{S}_n}{\cong}\bigoplus_{\lambda_C\in\mathcal{I}_n}\mathcal{P}_{\lambda_C}\otimes\operatorname{Hom}(\mathcal{P}_{\lambda_C},\mathcal{P}_{\lambda_A}\otimes\mathcal{P}_{\lambda_B})^{\mathcal{S}_n}\overset{\mathcal{S}_n}{\cong}\bigoplus_{\lambda_C\in\mathcal{I}_n}\mathcal{P}_{\lambda_C}\otimes\mathbb{C}^{g_{\lambda_A\lambda_B\lambda_C}} \tag{6.29}$$

Here we have defined the Kronecker coefficient $g_{\lambda_A\lambda_B\lambda_C}:=\dim\operatorname{Hom}(\mathcal{P}_{\lambda_C},\mathcal{P}_{\lambda_A}\otimes\mathcal{P}_{\lambda_B})^{\mathcal{S}_n}$.

It can be shown that there is an orthonormal basis for $\mathcal{P}_\lambda$, which we call P_λ, in which $\mathbf{p}_\lambda(s)$ are real and orthogonal.* This means that $\mathcal{P}_\lambda\overset{\mathcal{S}_n}{\cong}\mathcal{P}_\lambda^*$. Since $\operatorname{Hom}(A,B)\cong A^*\otimes B$, it follows that

$$\operatorname{Hom}(\mathcal{P}_{\lambda_C},\mathcal{P}_{\lambda_A}\otimes\mathcal{P}_{\lambda_B})^{\mathcal{S}_n}\overset{\mathcal{S}_n}{\cong}(\mathcal{P}_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_C})^{\mathcal{S}_n}. \tag{6.30}$$

As a corollary, $g_{\lambda_A\lambda_B\lambda_C}$ is unchanged by permuting $\lambda_A,\lambda_B,\lambda_C$. Unfortunately, no efficient method of calculating $g_{\lambda_A\lambda_B\lambda_C}$ is known, though asymptotically they have some connections to the quantum mutual information that will be investigated in future work. The permutation symmetry of $g_{\lambda_A\lambda_B\lambda_C}$ also means that we can consider CG transformations from $AB\to C$, $AC\to B$ or $BC\to A$, with the only difference being a normalization factor which we will explain below.

According to Eq. (6.30), the CG transformation can be understood in terms of tripartite $\mathcal{S}_n$-invariant vectors. Let $|\alpha\rangle$ be a unit vector in $(\mathcal{P}_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_C})^{\mathcal{S}_n}$, with corresponding density matrix $\alpha=|\alpha\rangle\!\langle\alpha|$. Since $\alpha^A:=\operatorname{Tr}_{BC}\alpha$ is invariant under permutations and $\operatorname{Tr}\alpha=1$, Schur's Lemma implies that $\alpha^A=I_{\mathcal{P}_{\lambda_A}}/D_A$, with $D_A:=\dim\mathcal{P}_{\lambda_A}$. This means we can Schmidt decompose $|\alpha\rangle$ as

$$|\alpha\rangle^{ABC}=\frac{1}{\sqrt{D_A}}\sum_{p_A\in P_{\lambda_A}}|p_A\rangle^A W_\alpha^{A'\to BC}|p_A\rangle^{A'} \tag{6.31}$$

where $W_\alpha\in\operatorname{Hom}(\mathcal{P}_{\lambda_A},\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_C})^{\mathcal{S}_n}$ is an isometry. We can express $U_{\mathrm{CG}}^{\lambda_B,\lambda_C}$ in terms of W_α according

*One way to prove this is to consider Young's natural basis, which was introduced in Section 5.3.1. Since the $\Pi_{\lambda:T}$ produce real linear combinations of the states $|i_1,\ldots,i_n\rangle$, the matrices $\mathbf{p}_\lambda(s)$ are also real when written in Young's natural basis. If we generate an orthonormal basis by applying Gram-Schmidt to Young's natural basis, the matrices $\mathbf{p}_\lambda(s)$ remain real.

In Section 7.1.2 we will introduce a different orthonormal basis for $\mathcal{S}_n$, known as Young's orthogonal basis, or as the Young-Yamanouchi basis. [JK81] gives an explicit formula in this basis for $\mathbf{p}_\lambda(s)$ in which the matrices are manifestly real.

to

$$(U_{\mathrm{CG}}^{\lambda_B,\lambda_C})^\dagger|\alpha\rangle|p_A\rangle = W_\alpha|p_A\rangle. \tag{6.32}$$

The simple form of Eq. (6.31) suggests that the CG transformation can also be implemented by teleportation. For any $\lambda \in \mathcal{I}_n$, let $D_\lambda := \dim \mathcal{P}_\lambda$ and define $|\Phi_\lambda\rangle = D_\lambda^{-\frac{1}{2}} \sum_{p\in P_\lambda} |p\rangle|p\rangle$. Note that up to a phase $|\Phi_\lambda\rangle$ is the unique invariant vector in $\mathcal{P}_\lambda \otimes \mathcal{P}_\lambda$. To see that it is invariant, use the fact that $(A \otimes I)|\Phi_\lambda\rangle = (I \otimes A^T)|\Phi_\lambda\rangle$ for any operator A and the fact that $\mathbf{p}_\lambda$ are orthogonal matrices, so $\mathbf{p}_\lambda(s)^T = \mathbf{p}_\lambda(s)^{-1}$. Uniqueness follows from

$$\dim(\mathcal{P}_\lambda \otimes \mathcal{P}_\lambda)^{\mathcal{S}_n} = \dim \mathrm{Hom}(\mathcal{P}_\lambda, \mathcal{P}_\lambda)^{\mathcal{S}_n} = 1, \tag{6.33}$$

where the first equality is because $\mathcal{P}_\lambda \overset{\mathcal{S}_n}{\cong} \mathcal{P}_\lambda^*$ and the second equality is due to Schur's Lemma.

We use $|\Phi_\lambda\rangle$ for teleportation as follows. If $|p_A\rangle \in \mathcal{P}_{\lambda_A}$ is a basis vector (i.e. $|p_A\rangle \in P_{\lambda_A}$), then $\langle\Phi_\lambda|^{AA'}|p_A\rangle^A = \frac{1}{\sqrt{D_A}}\langle p_A|$. Also Eq. (6.31) can be written more simply as

$$|\alpha\rangle^{ABC} = \left(I^A \otimes W_\alpha^{A'\to BC}\right)|\Phi_{\lambda_A}\rangle^{AA'}. \tag{6.34}$$

Combining Eqns. (6.34) and (6.32) then yields

$$\langle\Phi_{\lambda_A}|^{AA'}|\alpha\rangle^{A'BE}|p_A\rangle^A = \frac{1}{D_A}W_\alpha^{A\to BE}|p_A\rangle^A = \frac{1}{D_A}(U_{\mathrm{CG}}^{\lambda_B,\lambda_C})^\dagger|\alpha\rangle|p_A\rangle \tag{6.35}$$

This connection between the quantum CG transform and $\mathcal{S}_n$-invariant tripartite states will now be used to decompose i.i.d. quantum channels.

6.4.2 Decomposition of memoryless quantum channels

Let $\mathcal{N} : A' \to B$ be a quantum channel and $U_\mathcal{N} : A' \to BE$ its isometric extension. Let $d_A = \dim A', d_B = \dim B, d_E = \dim E$ and $d := \max(d_A, d_B, d_E)$. We want to consider n uses of $U_\mathcal{N}$ in the Schur basis. In general this has the form

$$U_{\mathrm{Sch}}U_\mathcal{N}^{\otimes n}U_{\mathrm{Sch}}^\dagger = \sum_{\lambda_A,\lambda_B,\lambda_E\in\mathcal{I}_N} |\lambda_B\lambda_E\rangle\langle\lambda_A| \sum_{\substack{q_A\in Q_{\lambda_A},q_B\in Q_{\lambda_B}\\ q_E\in Q_{\lambda_E}}} |q_Bq_E\rangle\langle q_A| \sum_{\substack{p_A\in P_{\lambda_A},p_B\in P_{\lambda_B}\\ p_E\in P_{\lambda_E}}} |p_Bp_E\rangle\langle p_A|\, C^{\lambda_Aq_Ap_A}_{\lambda_B\lambda_Eq_Bq_Ep_Bp_E}, \tag{6.36}$$

for some coefficients $C^{\lambda_Aq_Ap_A}_{\lambda_B\lambda_Eq_Bq_Ep_Bp_E}$. So far this tells us nothing at all! But we know that $U_\mathcal{N}^{\otimes n}$ is invariant under permutations; i.e. $\left[\mathbf{P}(s^{-1})^B \otimes \mathbf{P}(s^{-1})^E\right] U_\mathcal{N}^{\otimes n}\mathbf{P}(s)^A = U_\mathcal{N}^{\otimes n}$ for all $s \in \mathcal{S}_n$. Thus

$$U_\mathcal{N}^{\otimes n} \in \mathrm{Hom}(\mathbb{C}^{d_A^n}, \mathbb{C}^{d_B^n} \otimes \mathbb{C}^{d_E^n})^{\mathcal{S}_n} \overset{\mathcal{U}_d\times\mathcal{S}_n}{\cong} \bigoplus_{\lambda_A,\lambda_B,\lambda_E\in\mathcal{I}_n} \mathrm{Hom}(\mathcal{Q}_{\lambda_A}^{d_A}, \mathcal{Q}_{\lambda_B}^{d_B} \otimes \mathcal{Q}_{\lambda_E}^{d_E}) \hat{\otimes}\, \mathrm{Hom}(\mathcal{P}_{\lambda_A}, \mathcal{P}_{\lambda_B} \otimes \mathcal{P}_{\lambda_E})^{\mathcal{S}_n} \tag{6.37}$$

Let $P[\lambda_A; \lambda_B, \lambda_E]$ be an orthonormal basis for $\mathrm{Hom}(\mathcal{P}_{\lambda_A}, \mathcal{P}_{\lambda_B} \otimes \mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$. Then we can expand $U_{\mathrm{Sch}}U_\mathcal{N}^{\otimes n}U_{\mathrm{Sch}}^\dagger$ as

$$U_{\mathrm{Sch}}U_\mathcal{N}^{\otimes n}U_{\mathrm{Sch}}^\dagger = \sum_{\substack{\lambda_A,\lambda_B,\lambda_E\in\mathcal{I}_n,\alpha\in P[\lambda_A;\lambda_B,\lambda_E]\\ q_A\in Q_{\lambda_A}^{d_A},q_B\in Q_{\lambda_B}^{d_B},q_E\in Q_{\lambda_E}^{d_E}}} [V_\mathcal{N}^n]^{\lambda_Aq_A}_{\lambda_B\lambda_Eq_Bq_E\alpha}|\lambda_B\lambda_E\rangle\langle\lambda_A| \otimes |q_Bq_E\rangle\langle q_A| \otimes W_\alpha, \tag{6.38}$$

where the coefficients $[V^n_{\mathcal{N}}]^{\lambda_A q_A}_{\lambda_B \lambda_E q_B q_E \alpha}$ correspond to an isometry; i.e.

$$\sum_{\lambda_B,\lambda_E,q_B,q_E,\alpha} ([V^n_{\mathcal{N}}]^{\lambda_A q_A}_{\lambda_B \lambda_E q_B q_E \alpha})^* [V^n_{\mathcal{N}}]^{\lambda'_A q'_A}_{\lambda_B \lambda_E q_B q_E \alpha} = \delta_{\lambda_A,\lambda'_A} \delta_{q_A,q'_A}. \tag{6.39}$$

This is depicted as a quantum circuit in Fig. 6-1.

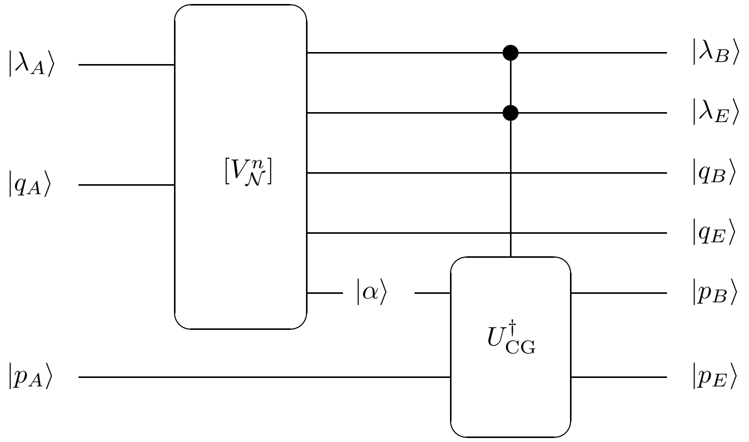


Figure 6-1: The quantum channel $U^{\otimes n}_{\mathcal{N}}$ is decomposed in the Schur basis as in Eq. (6.38). Alice inputs an n qudit state of the form $|\lambda_A\rangle|q_A\rangle|p_A\rangle$ and the channel outputs superpositions of $|\lambda_B\rangle|q_B\rangle|p_B\rangle$ for Bob and $|\lambda_E\rangle|q_E\rangle|p_E\rangle$ for Eve. The intermediate state $|\alpha\rangle$ belongs to $\mathrm{Hom}(\mathcal{P}_{\lambda_A}, \mathcal{P}_{\lambda_B} \otimes \mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$.

Using Eqns. (6.32) and (6.35), we can replace the CG transform in Fig. 6-1 with a teleportation-like circuit. Instead of interpreting α as a member of $\mathrm{Hom}(\mathcal{P}_{\lambda_A}, \mathcal{P}_{\lambda_B} \otimes \mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$, we say that $|\alpha\rangle \in (\mathcal{P}_{\lambda_A} \otimes \mathcal{P}_{\lambda_B} \otimes \mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$. This has the advantage of making its normalization more straightforward and of enhancing the symmetry between A, B and E. The $U^\dagger_{\mathrm{CG}}$ then becomes replaced with a projection onto $|\Phi_{\lambda_A}\rangle$. Since this only succeeds with probability $1/D_A$, the resulting state needs to be normalized by multiplying by $\sqrt{D_A}$. The resulting circuit is given in Fig. 6-2.

6.4.3 Jointly typical projectors in the Schur basis

The channel decomposition in the last section is still extremely general. In particular, the structure of the map is given by the λ_A, λ_B and λ_E which appear in Eq. (6.38), but generically all of the coefficients will be nonzero. However, for large values of n, almost all of the weight will be contained in a small set of *typical* triples of $(\lambda_A, \lambda_B, \lambda_E)$. These triples are the quantum analogue of joint types from classical information theory.

In this section we show the existence of typical sets of $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E)$ onto which a channel's input and output can be projected with little disturbance. In fact, we will define three versions of the typical set $\mathcal{T}^n_{\mathcal{N}}$ and show that they are in a certain sense asymptotically equivalent. For each version, let ρ^A be an arbitrary channel input, and $|\psi\rangle^{ABE} = (I^A \otimes U^{A' \to BE}_{\mathcal{N}})|\Phi_\rho\rangle^{AA'}$ the purified channel output (following the CP formalism). Now define $R(\mathcal{N})$ to be set of ψ^{ABE} that can be generated in this manner.

- Define $\mathcal{T}^*_{\mathcal{N}} := \{(r_A, r_B, r_E) : \exists \psi^{ABE} \in R(\mathcal{N}) \text{ s.t.} \, r_A = \mathrm{spec}(\psi^A), r_B = \mathrm{spec}(\psi^B), r_E = \mathrm{spec}(\psi^E)\}$. This set is simply the set of triples of spectra that can arise from one use of the channel. It has the advantage of being easy to compute and to optimize over, but it doesn't give us direct information about which values of $(\lambda_A, \lambda_B, \lambda_E)$ we need to consider.

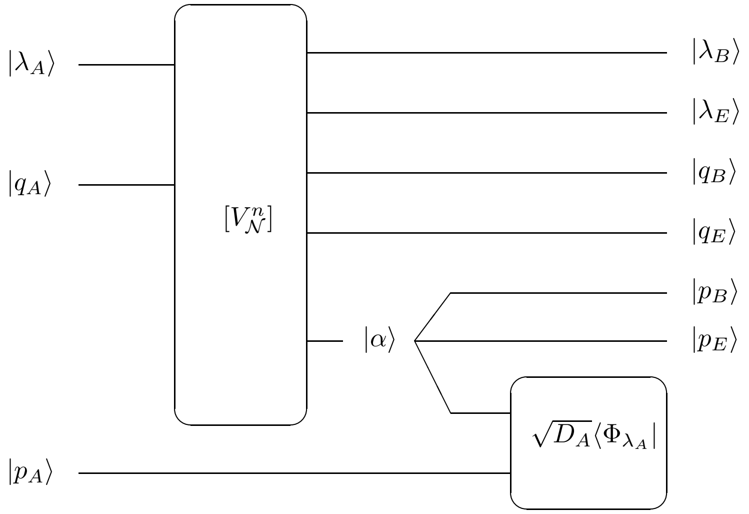


Figure 6-2: The quantum channel $U_{\mathcal{N}}^{\otimes n}$ is decomposed in the Schur basis with teleportation replacing the $\mathcal{S}_n$ CG transform. Here the intermediate state $|\alpha\rangle$ belongs to $(\mathcal{P}_{\lambda_A} \otimes \mathcal{P}_{\lambda_B} \otimes \mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$, the box labeled $\sqrt{D_A}\langle\Phi_{\lambda_A}|$ represents projecting onto the maximally entangled state $|\Phi_{\lambda_A}\rangle$ and normalization requires multiplying the residual state by $\sqrt{D_A}$, where $D_A := \dim \mathcal{P}_{\lambda_A}$.

- Define $\widetilde{\mathcal{T}}_{\mathcal{N}}^n(\epsilon) := \{(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) : \exists \psi^{ABE} \in R(\mathcal{N}) \text{ s.t. } \mathrm{Tr}(\Pi_{\lambda_A}^A \otimes \Pi_{\lambda_B}^B \otimes \Pi_{\lambda_E}^E)\psi^{\otimes n} \geq \epsilon\}$.

  This set tells us which $(\lambda_A, \lambda_B, \lambda_E)$ we need to consider when working with purified outputs of $U_{\mathcal{N}}^{\otimes n}$. To see this note that if $\psi \in R(\mathcal{N})$, then projecting $\psi^{\otimes n}$ onto $\widetilde{\mathcal{T}}_{\mathcal{N}}^n(\epsilon)$ will succeed with probability $\geq 1 - \epsilon(n+1)^{3d}$ since there are $\leq (n+1)^{3d}$ possible triples $(\lambda_A, \lambda_B, \lambda_E)$.

- Define $\mathcal{T}_{\mathcal{N}}^n(\epsilon)$ to be the set of $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E)$ s.t. there exists a subnormalized density matrix ω_A on $\mathcal{Q}_{\lambda_A}^{d_A}$ (i.e. $\mathrm{Tr}\,\omega_A \leq 1$) s.t.

  $$\mathrm{Tr}(|\lambda_B\rangle\langle\lambda_B| \otimes |\lambda_E\rangle\langle\lambda_E| \otimes I_{\mathcal{Q}_{\lambda_B}} \otimes I_{\mathcal{Q}_{\lambda_E}} \otimes I_\alpha) V_{\mathcal{N}}^n(|\lambda_A\rangle\langle\lambda_A| \otimes \omega_A) \geq \epsilon. \tag{6.40}$$

  Since that $V_{\mathcal{N}}^n$ completely determines the map from $\lambda_A \mapsto (\lambda_B, \lambda_E)$, we don't need to consider different values of the $|p_A\rangle$ register.

  This set is useful when considering channel outputs in the CQ formalism. It says that if the input is encoded in $\mathcal{Q}_{\lambda_A}^{d_A} \otimes \mathcal{P}_{\lambda_A}$ then only certain output states need be considered.

All of these sets could also be generalized to include possible $\mathcal{Q}_\lambda^d$ states as well. However, we focus attention on the $(\lambda_A, \lambda_B, \lambda_E)$ since those determine the dimensions of $\mathcal{P}_\lambda$ and hence the possible communication rates.

We claim that the three typical sets described above are close to one another. In other words, for any element in one typical set, the other sets have nearby elements, although we may have to decrease ϵ. Here "nearby" means that the distance goes to zero for any fixed or slowly-decreasing value of ϵ as $n \to \infty$.

In the following proofs we will frequently omit mentioning U_{Sch}, implicitly identifying $\rho^{\otimes n}$ with $U_{\mathrm{Sch}}\rho^{\otimes n}U_{\mathrm{Sch}}^\dagger$ and $U_{\mathcal{N}}^{\otimes n}$ with $U_{\mathrm{Sch}}U_{\mathcal{N}}^{\otimes n}U_{\mathrm{Sch}}^\dagger$.

- $\mathcal{T}_{\mathcal{N}}^* \Rightarrow \widetilde{\mathcal{T}}_{\mathcal{N}}^n(\epsilon)$ (i.e. for any triple in $\mathcal{T}_{\mathcal{N}}^*$ there is a nearby triple in $\widetilde{\mathcal{T}}_{\mathcal{N}}^n(\epsilon)$)

  Proof. Suppose $(r_A, r_B, r_E) \in \mathcal{T}_{\mathcal{N}}^*$ and let ψ^{ABE} be the corresponding state in $R(\mathcal{N})$ whose reduced states have spectra r_A, r_B and r_E. Define the probability distribution $\Pr(\lambda_A, \lambda_B, \lambda_E) :=$

$\operatorname{Tr}(\Pi^A_{\lambda_A} \otimes \Pi^B_{\lambda_B} \otimes \Pi^E_{\lambda_E})\psi^{\otimes n}$. Then by Eq. (6.23), $\Pr(\frac{1}{2}\|r_A - \overline{\lambda}_A\|_1 > \delta) \leq (n+d)^{d(d+1)/2}\exp(-n\delta^2)$ for any $\delta > 0$. Repeating this for $\overline{\lambda}_B$ and $\overline{\lambda}_E$, we find that

$$\Pr\left[\left(\frac{1}{2}\|r_A - \overline{\lambda}_A\|_1 > \delta\right) \vee \left(\frac{1}{2}\|r_B - \overline{\lambda}_B\|_1 > \delta\right) \vee \left(\frac{1}{2}\|r_E - \overline{\lambda}_E\|_1 > \delta\right)\right] \leq 3(n+d)^{\frac{d(d+1)}{2}}\exp(-n\delta^2).$$

Since the number of triples $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E)$ is $\leq (n+1)^{3d}$, this means there exists a triple $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E)$ with $\Pr(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) \geq (n+1)^{-3d}(1 - 3(n+d)^{d(d+1)/2}\exp(-n\delta^2)) =: \epsilon$ (and so $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) \in \widetilde{\mathcal{T}}^n_{\mathcal{N}}(\epsilon)$), satisfying $\frac{1}{2}\|r_A - \overline{\lambda}_A\|_1 \leq \delta$, $\frac{1}{2}\|r_B - \overline{\lambda}_B\|_1 \leq \delta$ and $\frac{1}{2}\|r_E - \overline{\lambda}_E\|_1 \leq \delta$. One natural choice is to take $\delta = (\log n)/\sqrt{n}$ and $\epsilon = 1/\operatorname{poly}(n)$. □

- $\widetilde{\mathcal{T}}^n_{\mathcal{N}}(\epsilon) \subseteq \mathcal{T}^n_{\mathcal{N}}(\epsilon)$

Proof. Suppose $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) \in \widetilde{\mathcal{T}}^n_{\mathcal{N}}(\epsilon)$, meaning that there exists $\psi^{ABE} \in R(\mathcal{N})$ s.t. $\operatorname{Tr}(\Pi^A_{\lambda_A} \otimes \Pi^B_{\lambda_B} \otimes \Pi^E_{\lambda_E})\psi^{\otimes n} \geq \epsilon$. Thus if we set $\rho^A = \operatorname{Tr}_{BE}\psi^{\otimes n}$ then

$$\begin{aligned}
\epsilon &\leq \operatorname{Tr}(\Pi^A_{\lambda_A} \otimes \Pi^B_{\lambda_B} \otimes \Pi^E_{\lambda_E})\left[(I^A \otimes U_{\mathcal{N}}^{A'\to BE})|\Phi_\rho\rangle^{AA'}\right]^{\otimes n} && (6.41)\\
&= \operatorname{Tr}(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E})U_{\mathcal{N}}^{\otimes n}(\Pi_{\lambda_A}\rho^{\otimes n}\Pi_{\lambda_A}) && (6.42)\\
&= \operatorname{Tr}(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E})U_{\mathcal{N}}^{\otimes n}(|\lambda_A\rangle\langle\lambda_A| \otimes \mathbf{q}_{\lambda_A}(\rho) \otimes I_{\mathcal{P}_{\lambda_A}}) && (6.43)\\
&= \operatorname{Tr}(|\lambda_B\rangle\langle\lambda_B| \otimes I_{\mathcal{Q}_{\lambda_B}} \otimes |\lambda_E\rangle\langle\lambda_E| \otimes I_{\mathcal{Q}_{\lambda_E}})V_{\mathcal{N}}^n(|\lambda_A\rangle\langle\lambda_A| \otimes \mathbf{q}_{\lambda_A}(\rho) \cdot \dim\mathcal{P}_{\lambda_A}) && (6.44)\\
&= \operatorname{Tr}\left(|\lambda_B\rangle\langle\lambda_B| \otimes I_{\mathcal{Q}_{\lambda_B}} \otimes |\lambda_E\rangle\langle\lambda_E| \otimes I_{\mathcal{Q}_{\lambda_E}}\right)V_{\mathcal{N}}^n(|\lambda_A\rangle\langle\lambda_A| \otimes \omega_A) && (6.45)
\end{aligned}$$

In the last step we have defined the (subnormalized) density matrix $\omega_A := \mathbf{q}_{\lambda_A}(\rho) \cdot \dim\mathcal{P}_{\lambda_A}$. (It is subnormalized because $\operatorname{Tr}\omega_A = \operatorname{Tr}\Pi_{\lambda_A}\rho^{\otimes n} \leq 1$.) Thus $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) \in \mathcal{T}^n_{\mathcal{N}}(\epsilon)$. □

- $\mathcal{T}^n_{\mathcal{N}}(\epsilon) \subseteq \widetilde{\mathcal{T}}^n_{\mathcal{N}}(\epsilon')$, $\epsilon' = \epsilon(n+d)^{-d^2}$

Proof. If $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) \in \mathcal{T}^n_{\mathcal{N}}(\epsilon)$ then there exists a density matrix ω_A on $\mathcal{Q}^{d_A}_{\lambda_A}$ s.t.

$$\operatorname{Tr}(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E})\mathcal{N}^{\otimes n}\left(|\lambda_A\rangle\langle\lambda_A| \otimes \omega_A \otimes \frac{I_{\mathcal{P}_{\lambda_A}}}{\dim\mathcal{P}_{\lambda_A}}\right) \geq \epsilon. \tag{6.46}$$

In fact, this would remain true if we replaced $I_{\mathcal{P}_{\lambda_A}}/\dim\mathcal{P}_{\lambda_A}$ with any normalized state.

Define $\rho_0 = \sum_{i=1}^{d_A}\overline{\lambda}_{A,i}|i\rangle\langle i|$ and let dU denote a Haar measure on $\mathcal{U}_{d_A}$. By Schur's Lemma, averaging $\mathbf{q}^{d_A}_{\lambda_A}(U\rho_0U^\dagger)$ over dU gives a matrix proportional to the identity. To obtain the proportionality factor, we use Eq. (6.20) to bound

$$\beta := \operatorname{Tr}\Pi_{\lambda_A}\rho_0^{\otimes n}\Pi_{\lambda_A} = \operatorname{Tr}\mathbf{q}^{d_A}_{\lambda_A}(\rho_0) \cdot \dim\mathcal{P}_{\lambda_A} \geq (n+d)^{-d(d+1)/2}. \tag{6.47}$$

Upon averaging, we then find that

$$\beta\frac{I_{\mathcal{Q}^{d_A}_{\lambda_A}}}{\dim\mathcal{Q}^{d_A}_{\lambda_A}} = \int dU\mathbf{q}^{d_A}_{\lambda_A}(U\rho_0U^\dagger) \cdot \dim\mathcal{P}_{\lambda_A} \tag{6.48}$$

Now $\omega_A \leq I_{\mathcal{Q}^{d_A}_{\lambda_A}}$, so $\mathcal{N}^{\otimes n}\left(|\lambda_A\rangle\langle\lambda_A| \otimes \omega_A \otimes \frac{I_{\mathcal{P}_{\lambda_A}}}{\dim \mathcal{P}_{\lambda_A}}\right) \leq \mathcal{N}^{\otimes n}\left(|\lambda_A\rangle\langle\lambda_A| \otimes I_{\mathcal{Q}^{d_A}_{\lambda_A}} \otimes \frac{I_{\mathcal{P}_{\lambda_A}}}{\dim \mathcal{P}_{\lambda_A}}\right)$ and

$$\begin{aligned}
\epsilon &\leq \operatorname{Tr} \mathcal{N}^{\otimes n}\left(|\lambda_A\rangle\langle\lambda_A| \otimes \omega_A \otimes \frac{I_{\mathcal{P}_{\lambda_A}}}{\dim \mathcal{P}_{\lambda_A}}\right)(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) && (6.49)\\
&\leq \operatorname{Tr} \mathcal{N}^{\otimes n}\left(|\lambda_A\rangle\langle\lambda_A| \otimes I_{\mathcal{Q}^{d_A}_{\lambda_A}} \otimes \frac{I_{\mathcal{P}_{\lambda_A}}}{\dim \mathcal{P}_{\lambda_A}}\right)(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) && (6.50)\\
&= \frac{\dim \mathcal{Q}^{d_A}_{\lambda_A}}{\beta} \int dU \operatorname{Tr}(\mathcal{N}^{\otimes n}(\Pi_{\lambda_A}(U\rho_0 U^\dagger)^{\otimes n}\Pi_{\lambda_A}))(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) && (6.51)\\
&\leq \max_U \frac{\dim \mathcal{Q}^{d_A}_{\lambda_A}}{\beta} \operatorname{Tr}(\mathcal{N}^{\otimes n}(\Pi_{\lambda_A}(U\rho_0 U^\dagger)^{\otimes n}\Pi_{\lambda_A}))(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}). && (6.52)
\end{aligned}$$

In the last step we have used the fact that $\int dU = 1$ so that $\int dU f(U) \leq \max_U f(U)$ for any function on $\mathcal{U}_{d_A}$. Therefore $\exists \rho = U\rho_0 U^\dagger$ with $\psi^{ABE} = (I^A \otimes U_{\mathcal{N}}^{A' \to BE})|\Phi_\rho\rangle^{AA'}$ such that $\operatorname{Tr}(\Pi^A_{\lambda_A} \otimes \Pi^B_{\lambda_B} \otimes \Pi^E_{\lambda_E})\psi^{\otimes n} \geq \epsilon\beta / \dim \mathcal{Q}^{d_A}_{\lambda_A} \geq \epsilon(n+d)^{-d^2} =: \epsilon'$.

This means that $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) \in \widetilde{\mathcal{T}}^n_{\mathcal{N}}(\epsilon')$. □

- $\widetilde{\mathcal{T}}^n_{\mathcal{N}}(\epsilon) \Rightarrow \mathcal{T}^*_{\mathcal{N}}$

Proof. Again, we are given $\psi^{ABE} \in R(\mathcal{N})$ and a triple $(\lambda_A, \lambda_B, \lambda_E)$ s.t. $\operatorname{Tr}(\Pi^A_{\lambda_A} \otimes \Pi^B_{\lambda_B} \otimes \Pi^E_{\lambda_E})\psi^{\otimes n} \geq \epsilon$. And again we define $\Pr(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) := \operatorname{Tr}(\Pi^A_{\lambda_A} \otimes \Pi^B_{\lambda_B} \otimes \Pi^E_{\lambda_E})\psi^{\otimes n}$. Now let $\delta := \max_{X \in \{A,B,E\}} \frac{1}{2}\|\overline{\lambda}_X - \operatorname{spec}\psi^X\|_1$ and use Eq. (6.23) to bound

$$\epsilon \leq \Pr(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E) \leq (n+d)^{d(d-1)/2} \exp(-n\delta^2). \tag{6.53}$$

Thus $(r_A, r_B, r_E) = (\operatorname{spec}\psi^A, \operatorname{spec}\psi^B, \operatorname{spec}\psi^E) \in \mathcal{T}^*_{\mathcal{N}}$ and satisfies $\frac{1}{2}\|r_A - \overline{\lambda}_A\|_1 \leq \delta$, $\frac{1}{2}\|r_B - \overline{\lambda}_B\|_1 \leq \delta$ and $\frac{1}{2}\|r_E - \overline{\lambda}_E\|_1 \leq \delta$ for δ s.t.

$$\delta^2 \leq \frac{\binom{d}{2}\log(n+d) + \log 1/\epsilon}{n}. \tag{6.54}$$

□

The preceding set of proofs establishes more than will usually be necessary. The main conclusion to draw from this section is that one can project onto triples $(\overline{\lambda}_A, \overline{\lambda}_B, \overline{\lambda}_E)$ that are all within δ of triples in $\mathcal{T}^*_{\mathcal{N}}$ while disturbing the state by no more than $\operatorname{poly}(n)\exp(-n\delta^2)$.

6.4.4 Conclusions

The results of this chapter should be thought of laying the groundwork for a quantum analogue of joint types. Although many coding theorems have been proved for noisy states and channels without using this formalism, hopefully joint quantum types will give proofs that are simpler, more powerful, or not feasible by other means. One problem for which the technique seems promising is the Quantum Reverse Shannon Theorem[BDH$^+$05], in which it gives a relatively simple method for efficiently simulating a noisy quantum channel on arbitrary sources. It remains to be seen where else the techniques will be useful.

Chapter 7

Efficient circuits for the Schur transform

The previous chapter showed how the Schur transform is a vital ingredient in a wide variety of coding theorems of quantum information theory. However, for these protocols to be of practical value, an efficient (i.e. polynomial time) implementation of the Schur transform will be necessary.

The goal of performing classical coding tasks in polynomial or even linear time has long been studied, but quantum information theory has typically ignored questions of efficiency. For example, random coding results (such as [Hol98, SW97, BHL$^+$05, DW04]) require an exponential number of bits to describe, and like classical random coding techniques, do not yield efficient algorithms. There are a few important exceptions. Some quantum coding tasks, such as Schumacher compression[Sch95, JS94], are essentially equivalent to classical circuits, and as such can be performed efficiently on a quantum computer by carefully modifying an efficient classical algorithm to run reversibly and deal properly with ancilla systems[CD96]. Another example, which illustrates some of the challenges involved, is [KM01]'s efficient implementation of entanglement concentration[BBPS96]. Quantum key distribution[BB84] not only runs efficiently, but can be implemented with entirely, or almost entirely, single-qubit operations and classical computation. Fault tolerance[Sho96] usually seeks to perform error correction with as few gates as possible, although using teleportation-based techniques[GC99, Kni04] computational efficiency may not be quite as critical to the threshold rate. Finally, some randomized quantum code constructions have been given efficient constructions using classical derandomization techniques in [AS04]. Our efficient construction of the Schur transform adds to this list a powerful new tool for finding algorithms that implement quantum communication tasks.

From a broader perspective, the transforms involved in quantum information protocols are important because they show a connection between a quantum problem with structure and transforms of quantum information which exploit this structure. The theory of quantum algorithms has languished relative to the tremendous progress in quantum information theory due in large part to a lack of exactly this type of construction: transforms with interpretations. When we say a quantum algorithm is simply a change of basis, we are doing a disservice to the fact that efficient quantum algorithms must have efficient quantum circuits. In the nonabelian hidden subgroup problem, for example, it is known that there is a transform which solves the problem, but there is no known efficient quantum circuit for this transform[EHK97]. There is great impetus, therefore, to construct efficient quantum circuits for transforms of quantum information where the transform exploits some structure of the problem.

We begin in Section 7.1 by describing explicit bases (known as *subgroup-adapted bases*) for the irreps of the unitary and symmetric groups. In Section 7.2, we show how these bases allow the Schur transform to be decomposed into a series of CG transforms and in Section 7.3 we give an efficient construction of a CG transform. Together these three sections comprise an efficient (i.e. running time polynomial in n, d and $\log 1/\epsilon$ for error ϵ) algorithm for the Schur transform.

7.1 Subgroup-adapted bases for $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$

To construct a quantum circuit for the Schur transform, we will need to explicitly specify the Schur basis. Since we want the Schur basis to be of the form $|\lambda, q, p\rangle$, our task reduces to specifying orthonormal bases for $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$. We will call these bases Q_λ^d and P_λ, respectively.

We will choose Q_λ^d and P_λ to both be a type of basis known as a *subgroup-adapted basis*. In Section 7.1.1 we describe the general theory of subgroup-adapted bases, and in Section 7.1.2, we will describe subgroup-adapted bases for $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$. As we will later see, these bases have a recursive structure that is naturally related to the structure of the algorithms that work with them. Here we will show how the bases can be stored on a quantum computer with a small amount of padding, and later in this chapter we will show how the subgroup-adapted bases described here enable efficient implementations of Clebsch-Gordan and Schur duality transforms.

7.1.1 Subgroup Adapted Bases

First we review the basic idea of a subgroup adapted basis. We assume that all groups we talk about are finite or compact Lie groups. Suppose $(\mathbf{r}, V)$ is an irrep of a group G and H is a proper subgroup of G. We will construct a basis for V via the representations of H.

Begin by restricting the input of $\mathbf{r}$ to H to obtain a representation of H, which we call $(\mathbf{r}|_H, V\!\downarrow_H)$. Unlike V, the H-representation $V\!\downarrow_H$ may be reducible. In fact, if we let $(\mathbf{r}'_\alpha, V'_\alpha)$ denote the irreps of H, then $V\!\downarrow_H$ will decompose under the action of H as

$$V\!\downarrow_H \overset{H}{\cong} \bigoplus_{\alpha \in \hat{H}} V'_\alpha \otimes \mathbb{C}^{n_\alpha} \tag{7.1}$$

or equivalently, $\mathbf{r}|_H$ decomposes as

$$\mathbf{r}(h) = \mathbf{r}|_H(h) \cong \bigoplus_{\alpha \in \hat{H}} \mathbf{r}'_\alpha(h) \otimes I_{n_\alpha} \tag{7.2}$$

where $\hat{H}$ runs over a complete set of inequivalent irreps of H and n_α is the *branching multiplicity* of the irrep labeled by α. Note that since $\mathbf{r}$ is a unitary representation, the subspaces corresponding to different irreps of H are orthogonal. Thus, the problem of finding an orthonormal basis for V now reduces to the problem of (1) finding an orthonormal basis for each irrep of H, V'_α and (2) finding orthonormal bases for the multiplicity spaces $\mathbb{C}^{n_\alpha}$. The case when all the n_α are either 0 or 1 is known as *multiplicity-free branching.* When this occurs, we only need to determine which irreps occur in the decomposition of V, and find bases for them.

Now consider a group G along with a tower of subgroups $G = G_1 \supset G_2 \supset \cdots \supset G_{k-1} \supset G_k = \{e\}$ where $\{e\}$ is the trivial subgroup consisting of only the identity element. For each G_i, denote its irreps by V_α^i, for $\alpha \in \hat{G}_i$. Any irrep $V_{\alpha_1}^1$ of $G = G_1$ decomposes under restriction to G_2 into G_2-irreps: say that $V_{\alpha_2}^2$ appears n_{α_1,α_2} times. We can then look at these irreps of G_2, consider their restriction to G_3 and decompose them into different irreps of G_3. Carrying on in such a manner down this tower of subgroups will yield a labeling for subspaces corresponding to each of these restrictions. Moreover, if we choose orthonormal bases for the multiplicity spaces, this will induce an orthonormal basis for G. This basis is known as a *subgroup-adapted basis* and basis vectors have the form $|\alpha_2, m_2, \alpha_3, m_3, \ldots, \alpha_k, m_k\rangle$, where $|m_i\rangle$ is a basis vector for the ($n_{\alpha_{i-1},\alpha_i}$-dimensional) multiplicity space of $V_{\alpha_i}^i$ in $V_{\alpha_{i-1}}^{i-1}$.

If the branching for each $G_{i+1} \subset G_i$ is multiplicity-free, then we say that the tower of subgroups is *canonical.* In this case, the subgroup adapted basis takes the particularly simple form of $|\alpha_2, \ldots, \alpha_k\rangle$, where each $\alpha_i \in \hat{G}_i$ and α_{i+1} appears in the decomposition of $V_{\alpha_i}\!\downarrow_{G_{i+1}}$. Often we include the original irrep label $\alpha = \alpha_1$ as well: $|\alpha_1, \alpha_2, \ldots, \alpha_k\rangle$. This means that there exists a basis whose vectors are completely determined (up to an arbitrary choice of phase) by which irreps of $G_1, \ldots, G_k$ they transform according to. Notice that a basis for the irrep V_α does not consist of all possible irrep labels α_i, but instead only those which can appear under the restriction which defines the basis.

The simple recursive structure of subgroup adapted bases makes them well-suited to performing explicit computations. Thus, for example, subgroup adapted bases play a major role in efficient quantum circuits for the Fourier transform over many nonabelian groups[MRR04].

7.1.2 Explicit orthonormal bases for $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$

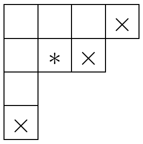
In this section we describe canonical towers of subgroups for $\mathcal{U}_d$ and $\mathcal{S}_n$, which give rise to subgroup-adapted bases for the irreps $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$. These bases go by many names: for $\mathcal{U}_d$ (and other Lie groups) the basis is called the Gel'fand-Zetlin basis (following [GZ50]) and we denote it by Q_λ^d, while for $\mathcal{S}_n$ it is called the Young-Yamanouchi basis, or sometimes Young's orthogonal basis (see [JK81] for a good review of its properties) and is denoted P_λ. The constructions and corresponding branching rules are quite simple, but for proofs we again refer the reader to [GW98].

The Gel'fand-Zetlin basis for $\mathcal{Q}_\lambda^d$: For $\mathcal{U}_d$, it turns out that the chain of subgroups $\{1\} = \mathcal{U}_0 \subset \mathcal{U}_1 \subset \ldots \subset \mathcal{U}_{d-1} \subset \mathcal{U}_d$ is a canonical tower. For $c < d$, the subgroup $\mathcal{U}_c$ is embedded in $\mathcal{U}_d$ by $\mathcal{U}_c := \{U \in \mathcal{U}_d : U|i\rangle = |i\rangle \text{ for } i = c+1, \ldots, d\}$. In other words, it corresponds to matrices of the form

$$U \oplus I_{d-c} := \left(\begin{array}{c|c} U & 0 \\ \hline 0 & I_{d-c} \end{array}\right), \tag{7.3}$$

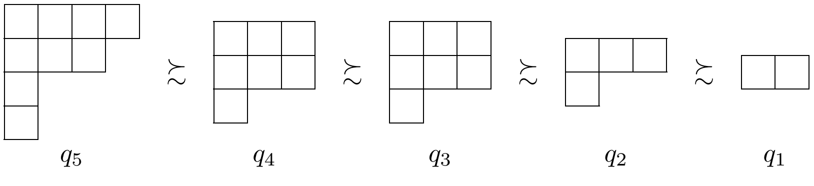
where U is a $c \times c$ unitary matrix.

Since the branching from $\mathcal{U}_d$ to $\mathcal{U}_{d-1}$ is multiplicity-free, we obtain a subgroup-adapted basis Q_λ^d, which is known as the Gel'fand-Zetlin (GZ) basis. Our only free choice in a GZ basis is the initial choice of basis $|1\rangle, \ldots, |d\rangle$ for $\mathbb{C}^d$ which determines the canonical tower of subgroups $\mathcal{U}_1 \subset \ldots \subset \mathcal{U}_d$. Once we have chosen this basis, specifying Q_λ^d reduces to knowing which irreps $\mathcal{Q}_\mu^{d-1}$ appear in the decomposition of $\mathcal{Q}_\lambda^d\!\downarrow_{\mathcal{U}_{d-1}}$. Recall that the irreps of $\mathcal{U}_d$ are labeled by elements of $\mathcal{I}_{d,n}$ with n arbitrary. This set can be denoted by $\mathbb{Z}_{++}^d := \cup_n \mathcal{I}_{d,n} = \{\lambda \in \mathbb{Z}^d : \lambda_1 \geq \ldots \geq \lambda_d \geq 0\}$. For $\mu \in \mathbb{Z}_{++}^{d-1}, \lambda \in \mathbb{Z}_{++}^d$, we say that μ *interlaces* λ and write $\mu \precsim \lambda$ whenever $\lambda_1 \geq \mu_1 \geq \lambda_2 \ldots \geq \lambda_{d-1} \geq \mu_{d-1} \geq \lambda_d$. In terms of Young diagrams, this means that μ is a valid partition (i.e. a nonnegative, nonincreasing sequence) obtained from removing zero or one boxes from each column of λ. For example, if $\lambda = (4,3,1,1)$ (as in Eq. (5.22)), then $\mu \precsim \lambda$ can be obtained by removing any subset of the marked boxes below, although if the box marked $*$ on the second line is removed, then the other marked box on the line must also be removed.



(7.4)

Thus a basis vector in Q_λ^d corresponds to a sequence of partitions $q = (q_d, \ldots, q_1)$ such that $q_d = \lambda$, $q_1 \precsim q_2 \precsim \ldots \precsim q_d$ and $q_j \in \mathbb{Z}_{++}^j$ for $j = 1, \ldots, d$. Again using $\lambda = (4,3,1,1)$ as an example, and choosing $d = 5$ (any $d \geq 4$ is possible), we might have the sequence



(7.5)

Observe that it is possible in some steps not to remove any boxes, as long as q_j has no more than j rows.

In order to work with the Gel'fand-Zetlin basis vectors on a quantum computer, we will need an efficient way to write them down. Typically, we think of d as constant and express our resource use in terms of n. Then an element of $\mathcal{I}_{d,n}$ can be expressed with $d\log(n+1)$ bits, since it consists of d integers between 0 and n. (This is a crude upper bound on $|\mathcal{I}_{d,n}| = \binom{n+d-1}{d-1}$, but for constant d it is good enough for our purposes.) A Gel'fand-Zetlin basis vector then requires no more than $d^2\log(n+1)$ bits, since it can be expressed as d partitions of integers no greater than n into $\leq d$ parts. (Here we assume that all partitions have arisen from a decomposition of $(\mathbb{C}^d)^{\otimes n}$, so that no Young diagram has more than n boxes.) Unless otherwise specified, our algorithms will use this encoding of the GZ basis vectors.

It is also possible to express GZ basis vectors in a more visually appealing way by writing numbers in the boxes of a Young diagram. If $q_1 \precsim \ldots \precsim q_d$ is a chain of partitions, then we write the number j in each box contained in q_j but not q_{j-1} (with $q_0 = (0)$). For example, the sequence in Eq. (7.5) would be denoted

$$\begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 5 \\ \hline 2 & 3 & 3 \\ \cline{1-3} 3 \\ \cline{1-1} 5 \\ \cline{1-1} \end{array} \quad . \tag{7.6}$$

Equivalently, any method of filling a Young diagram with numbers from $1, \ldots, d$ corresponds to a valid chain of irreps as long as the numbers are nondecreasing from left to right and are strictly increasing from top to bottom. This gives another way of encoding a GZ basis vector; this time using $n\log d$ bits. (In fact, we have an exact formula for $\dim \mathcal{Q}_\lambda^d$ (Eq. (6.12)) and later in this section we will give an algorithm for efficiently encoding a GZ basis vector in the optimal $\lceil \log \dim \mathcal{Q}_\lambda^d \rceil$ qubits. However, this is not necessary for most applications.)

Example: irreps of $\mathcal{U}_2$: To ground the above discussion in an example more familiar to physicists, we show how the GZ basis for $\mathcal{U}_2$ irreps corresponds to states of definite angular momentum along one axis. An irrep of $\mathcal{U}_2$ is labeled by two integers (λ_1, λ_2) such that $\lambda_1 + \lambda_2 = n$ and $\lambda_1 \geq \lambda_2 \geq 0$. A GZ basis vector for $\mathcal{Q}_\lambda^2$ has $\lambda_2 + m$ 1's in the first row, followed by $\lambda_1 - (\lambda_2 + m)$ 2's in the first row and λ_2 2's in the second row, where m ranges from 0 to $\lambda_1 - \lambda_2$. This arrangement is necessary to satisfy the constraint that numbers are strictly increasing from top to bottom and are nondecreasing from left to right. Since the GZ basis vectors are completely specified by m, we can label the vector $|(\lambda_1, \lambda_2); (\lambda_2 + m)\rangle \in Q_\lambda^2$ simply by $|m\rangle$. For example, $\lambda = (9, 4)$ and $m = 2$ would look like

$$\begin{array}{|c|c|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 \\ \hline 2 & 2 & 2 & 2 \\ \cline{1-4} \end{array} \quad . \tag{7.7}$$

Now observe that $\dim \mathcal{Q}_\lambda^2 = \lambda_1 - \lambda_2 + 1$, a fact which is consistent with having angular momentum $J = (\lambda_1 - \lambda_2)/2$. We claim that m corresponds to the Z component of angular momentum (specifically, the Z component of angular momentum is $m - J = m - (\lambda_1 - \lambda_2)/2$). To see this, first note that $\mathcal{U}_1$ acts on a GZ basis vector $|m\rangle$ according to the representation $x \rightarrow x^{\lambda_2+m}$, for $x \in \mathcal{U}_1$; equivalently $\mathbf{q}_\lambda^2\left(\left(\begin{smallmatrix} x & 0 \\ 0 & 1 \end{smallmatrix}\right)\right)|m\rangle = x^{\lambda_2+m}|m\rangle$. Since $\mathbf{q}_\lambda^2(yI_2)|m\rangle = y^n|m\rangle = y^{\lambda_1+\lambda_2}|m\rangle$, we can find the action of $e^{i\theta\sigma_z} = \left(\begin{smallmatrix} e^{2i\theta} & 0 \\ 0 & 1 \end{smallmatrix}\right)\begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$ on $|m\rangle$. We do this by combining the above arguments to find that $\mathbf{q}_\lambda^2(e^{i\theta\sigma_z})|m\rangle = e^{2i\theta(\lambda_2+m)}e^{-i\theta(\lambda_1+\lambda_2)}|m\rangle = e^{2i\theta(m-J)}|m\rangle$. Thus we obtain the desired action of a Z rotation on a particle with total angular momentum J and Z-component of angular momentum m.

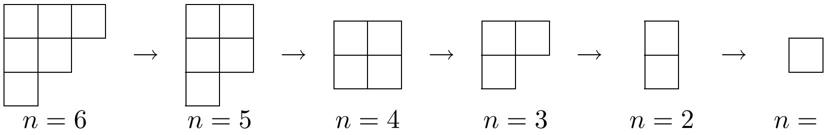
Example: The defining irrep of $\mathcal{U}_d$: The simplest nontrivial irrep of $\mathcal{U}_d$ is its action on $\mathbb{C}^d$. This corresponds to the partition (1), so we say that $(\mathbf{q}_{(1)}^d, \mathcal{Q}_{(1)}^d)$ is the *defining irrep* of $\mathcal{U}_d$ with $\mathcal{Q}_{(1)}^d = \mathbb{C}^d$ and $\mathbf{q}_{(1)}^d(U) = U$. Let $|1\rangle, \ldots, |d\rangle$ be an orthonormal basis for $\mathbb{C}^d$ corresponding to the canonical tower of subgroups $\mathcal{U}_1 \subset \cdots \subset \mathcal{U}_d$. It turns out that this is already a GZ basis. To see this, note that $\mathcal{Q}^d_{(1)}\downarrow_{\mathcal{U}_{d-1}} \overset{\mathcal{U}_{d-1}}{\cong} \mathcal{Q}^{d-1}_{(0)} \oplus \mathcal{Q}^{d-1}_{(1)}$. This is because $|d\rangle$ generates $\mathcal{Q}^{d-1}_{(0)}$, a trivial irrep of $\mathcal{U}_{d-1}$; and $|1\rangle, \ldots, |d-1\rangle$ generate $\mathcal{Q}^{d-1}_{(1)}$, a defining irrep of $\mathcal{U}_{d-1}$. Another way to say this is that $|j\rangle$ is acted on according to the trivial irrep of $\mathcal{U}_1, \ldots, \mathcal{U}_{j-1}$ and according to the defining irrep of $\mathcal{U}_j, \ldots, \mathcal{U}_d$. Thus $|j\rangle$ corresponds to the chain of partitions $\{(0)^{j-1}, (1)^{d-j+1}\}$. We will return to this example several times in the rest of the chapter.

The Young-Yamanouchi basis for $\mathcal{P}_\lambda$: The situation for $\mathcal{S}_n$ is quite similar. Our chain of subgroups is $\{e\} = \mathcal{S}_1 \subset \mathcal{S}_2 \subset \ldots \subset \mathcal{S}_n$, where for $m < n$ we define $\mathcal{S}_m \subset \mathcal{S}_n$ to be the permutations in $\mathcal{S}_n$ which leave the last $n-m$ elements fixed. For example, if $n = 3$, then $\mathcal{S}_3 = \{e, (12), (23), (13), (123), (321)\}$, $\mathcal{S}_2 = \{e, (12)\}$, and $\mathcal{S}_1 = \{e\}$. Recall that the irreps of $\mathcal{S}_n$ can be labeled by $\mathcal{I}_n = \mathcal{I}_{n,n}$: the partitions of n into $\leq n$ parts.

Again, the branching from $\mathcal{S}_n$ to $\mathcal{S}_{n-1}$ is multiplicity-free, so to determine an orthonormal basis P_λ for the space $\mathcal{P}_\lambda$ we need only know which irreps occur in the decomposition of $\mathcal{P}_\lambda\downarrow_{\mathcal{S}_{n-1}}$. It turns out that the branching rule is given by finding all ways to remove one box from λ while leaving a valid partition. Denote the set of such partitions by $\lambda - \square$. Formally, $\lambda - \square := \mathcal{I}_n \cap \{\lambda - e_j : j = 1, \ldots, n\}$, where we recall that e_j is the unit vector in $\mathbb{Z}^n$ with a one in the j^{th} position and zeroes elsewhere. Thus, the general branching rule is

$$\mathcal{P}_\lambda\downarrow_{\mathcal{S}_{n-1}} \overset{\mathcal{S}_{n-1}}{\cong} \bigoplus_{\mu \in \lambda - \square} \mathcal{P}_\mu. \tag{7.8}$$

For example, if $\lambda = (3, 2, 1)$, we might have the chain of partitions:



(7.9)

Again, we can concisely label this chain by writing the number j in the box that is removed when restricting from $\mathcal{S}_j$ to $\mathcal{S}_{j-1}$. The above example would then be

$$\begin{array}{|c|c|c|}\hline 1 & 3 & 6 \\ \hline 2 & 4 & \multicolumn{1}{c}{} \\ \cline{1-2} 5 & \multicolumn{2}{c}{} \\ \cline{1-1}\end{array} \quad . \tag{7.10}$$

Note that the valid methods of filling a Young diagram are slightly different than for the $\mathcal{U}_d$ case. Now we use each integer in $1, \ldots, n$ exactly once such that the numbers are increasing from left to right and from top to bottom. (The same filling scheme appeared in the description of Young's natural representation in Section 5.3.1, but the resulting basis states are of course quite different.)

This gives rise to a straightforward, but inefficient, method of writing an element of P_λ using $\log n!$ bits. However, for applications such as data compression[HM02a, HM02b] we will need an encoding which gives us closer to the optimal $\log P_\lambda$ bits. First recall that Eq. (6.13) gives an exact (and efficiently computable) expression for $|P_\lambda| = \dim \mathcal{P}_\lambda$. Now we would like to efficiently and reversibly map an element of P_λ (thought of as a chain of partitions $p = (p_n = \lambda, \ldots, p_1 = (1)) \in P_\lambda$, with $p_j \in p_{j+1} - \square$) to an integer in $[|P_\lambda|] := \{1, \ldots, |P_\lambda|\}$. We will construct this bijection $f_n : P_\lambda \to [|P_\lambda|]$ by defining an ordering on P_λ and setting $f_n(p) := |\{p' \in P_\lambda : p' \leq p\}|$. First fix an arbitrary, but easily computable, (total) ordering on partitions in $\mathcal{I}_n$ for each n; for example, lexicographical order. This induces an ordering on P_λ if we rank a basis vector $p \in P_\lambda$ first according to p_{n-1}, using the order on partitions we have chosen, then according to p_{n-2} and so on. We skip p_n, since it is always equal to λ. In other words, for $p, p' \in P_\lambda$, $p > p'$ if $p_{n-1} > p'_{n-1}$ *or* $p_{n-1} = p'_{n-1}$ and $p_{n-2} > p'_{n-2}$ *or* $p_{n-1} = p'_{n-1}$, $p_{n-2} = p'_{n-2}$ and $p_{n-3} > p'_{n-3}$, and so on. Thus $f_n : P_\lambda \to [|P_\lambda|]$ can be easily verified to be

$$f_n(p) = f_n(p_1, \ldots, p_n) := 1 + \sum_{k=2}^{n} \sum_{\substack{\mu \in p_k - \square \\ \mu < p_{k-1}}} \dim \mathcal{P}_\mu. \tag{7.11}$$

Thus f_n is an injective map from P_λ to $[|P_\lambda|]$. Moreover, since there are $O(n^2)$ terms in Eq. (7.11) and Eq. (6.13) gives an efficient way to calculate each $|P_\lambda|$, this mapping can be performed in time polynomial in n.

Of course, the same techniques could be used to efficiently write an element of Q^d_λ in $\lceil \log |Q^d_\lambda| \rceil$ bits, but unless d is large this usually is not necessary.

7.2 Constructing the Schur transform from a series of Clebsch-Gordan transforms

In this section, we will show how the Schur transform on $(\mathbb{C}^d)^{\otimes n}$ can be reduced to a series of CG transforms on $\mathcal{U}_d$. The argument is divided into two parts. First, we give the theoretical underpinnings in Section 7.2.1 by using Schur duality to relate the $\mathcal{U}_d$ CG transform to branching in $\mathcal{S}_n$. Then we show how the actual algorithm works in Section 7.2.2.

7.2.1 Branching rules and Clebsch-Gordan series for $\mathcal{U}_d$

Recall that CG transform for $\mathcal{U}_d$ is given by

$$\mathcal{Q}^d_\mu \otimes \mathcal{Q}^d_\nu \overset{\mathcal{U}_d}{\cong} \bigoplus_{\lambda \in \mathbb{Z}^d_{++}} \mathcal{Q}^d_\lambda \otimes \mathbb{C}^{M^\lambda_{\mu\nu}}. \tag{7.12}$$

For now, we will work with Littlewood-Richardson coefficients $M^\lambda_{\mu\nu}$ rather than the more structured space $\operatorname{Hom}(\mathcal{Q}^d_\lambda, \mathcal{Q}^d_\mu \otimes \mathcal{Q}^d_\nu)^{\mathcal{U}_d}$. The partitions λ appearing on the RHS of Eq. (7.12) are sometimes known as the *Clebsch-Gordan series*. In this section, we will show (following [GW98]) how the $\mathcal{U}_d$ Clebsch-Gordan series is related to the behavior of $\mathcal{S}_n$ irreps under restriction.

For integers k, n with $1 \leq k \leq n$, embed $\mathcal{S}_k \times \mathcal{S}_{n-k}$ as a subgroup of $\mathcal{S}_n$ in the natural way; as permutations that leave the sets $\{1, \ldots, k\}$ and $\{k+1, \ldots, n\}$ invariant. The irreps of $\mathcal{S}_k \times \mathcal{S}_{n-k}$ are $\mathcal{P}_\mu \hat{\otimes} \mathcal{P}_\nu$, where $\mu \in \mathcal{I}_k$ and $\nu \in \mathcal{I}_{n-k}$.

Under restriction to $\mathcal{S}_k \times \mathcal{S}_{n-k} \subset \mathcal{S}_n$, the $\mathcal{S}_n$-irrep $\mathcal{P}_\lambda$ decomposes as

$$\mathcal{P}_\lambda \overset{\mathcal{S}_k \times \mathcal{S}_{n-k}}{\cong} \bigoplus_{\mu \in \mathcal{I}_k} \bigoplus_{\nu \in \mathcal{I}_{n-k}} \mathcal{P}_\mu \hat{\otimes} \mathcal{P}_\nu \otimes \mathbb{C}^{N^\lambda_{\mu\nu}}, \tag{7.13}$$

for some multiplicities $N^\lambda_{\mu\nu}$ (possibly zero).

Claim 7.1. $M^\lambda_{\mu\nu} = N^\lambda_{\mu\nu}$.

As a corollary, $M^\lambda_{\mu\nu}$ is only nonzero when $|\lambda| = |\mu| + |\nu|$.

Proof. Consider the action of $\mathcal{S}_k \times \mathcal{S}_{n-k} \times \mathcal{U}_d$ on $(\mathbb{C}^d)^{\otimes n}$. On the one hand, Eq. (7.13) gives

$$(\mathbb{C}^d)^{\otimes n} \overset{\mathcal{S}_n \times \mathcal{U}_d}{\cong} \bigoplus_{\lambda \in \mathcal{I}_{d,n}} \mathcal{P}_\lambda \hat{\otimes} \mathcal{Q}^d_\lambda \overset{\mathcal{S}_k \times \mathcal{S}_{n-k}}{\cong} \bigoplus_{\substack{\mu \in \mathcal{I}_{d,k}, \nu \in \mathcal{I}_{d,n-k} \\ \lambda \in \mathcal{I}_{d,n}}} \mathcal{P}_\mu \hat{\otimes} \mathcal{P}_\nu \hat{\otimes} \mathcal{Q}^d_\lambda \otimes \mathbb{C}^{N^\lambda_{\mu\nu}}. \tag{7.14}$$

On the other hand, we can apply Eq. (7.12) to obtain

$$(\mathbb{C}^d)^{\otimes n} \cong (\mathbb{C}^d)^{\otimes k}\otimes(\mathbb{C}^d)^{\otimes n-k} \overset{\mathcal{S}_k\times\mathcal{S}_{n-k}}{\cong} \bigoplus_{\substack{\mu\in\mathcal{I}_{d,k}\\ \nu\in\mathcal{I}_{d,n-k}}} (\mathcal{P}_\mu\otimes\mathcal{Q}^d_\mu)\hat{\otimes}(\mathcal{P}_\nu\otimes\mathcal{Q}^d_\nu) \overset{\mathcal{U}_d}{\cong} \bigoplus_{\substack{\mu\in\mathcal{I}_{d,k},\nu\in\mathcal{I}_{d,n-k}\\ \lambda\in\mathbb{Z}^d_{++}}} \mathcal{P}_\mu\hat{\otimes}\mathcal{P}_\nu\hat{\otimes}\mathcal{Q}^d_\lambda\otimes\mathbb{C}^{M^\lambda_{\mu\nu}}. \tag{7.15}$$

Equating Eqns. (7.14) and (7.15) proves the desired equality. □

This means that the branching rules of $\mathcal{S}_n$ determine the CG series for $\mathcal{U}_d$.[*] In particular, suppose $k = n-1$. Then $\mathcal{S}_1$ is the trivial group, so restricting to $\mathcal{S}_{n-1}\times\mathcal{S}_1$ is equivalent to simply restricting to $\mathcal{S}_{n-1}$. According to the branching rule stated in Eq. (7.8), this means that $M^{\lambda'}_{\lambda,(1)}$ is one if $\lambda \in \lambda' - \square$ and zero otherwise. In other words, for the case when one irrep is the defining irrep, the CG series is

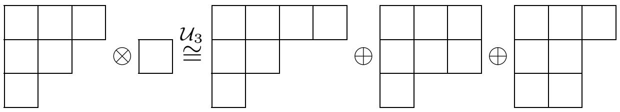
$$\mathcal{Q}^d_\lambda \otimes \mathcal{Q}^d_{(1)} \cong \bigoplus_{\lambda'\in\lambda+\square} \mathcal{Q}^d_{\lambda'}. \tag{7.16}$$

Here $\lambda + \square$ denotes the set of valid Young diagrams obtained by *adding* one box to λ.

For example if $\lambda = (3, 2, 1)$ then

$$\mathcal{Q}^3_{(3,2,1)} \otimes \mathcal{Q}^3_{(1)} \overset{\mathcal{U}_3}{\cong} \mathcal{Q}^3_{(4,2,1)} \oplus \mathcal{Q}^3_{(3,3,1)} \oplus \mathcal{Q}^3_{(3,2,2)} \tag{7.17}$$

or in Young diagram form



(7.18)

Note that if we had $d > 3$, then the partition $(3, 2, 1, 1)$ would also appear.

We now seek to define the CG transform as a quantum circuit. We specialize to the case where one of the input irreps is the defining irrep, but allow the other irrep to be specified by a quantum input. The resulting CG transform is defined as:

$$U_{\text{CG}} = \sum_{\lambda\in\mathbb{Z}^d_{++}} |\lambda\rangle\langle\lambda| \otimes U^{\lambda,(1)}_{\text{CG}}. \tag{7.19}$$

This takes as input a state of the form $|\lambda\rangle|q\rangle|i\rangle$, for $\lambda \in \mathbb{Z}^d_{++}$, $|q\rangle \in Q^d_\lambda$ and $i \in [d]$. The output is a superposition over vectors $|\lambda\rangle|\lambda'\rangle|q'\rangle$, where $\lambda' = \lambda + e_j \in \mathbb{Z}^d_{++}$, $j \in [d]$ and $|q'\rangle \in Q^d_{\lambda'}$. Equivalently, we could output $|\lambda\rangle|j\rangle|q'\rangle$ or $|j\rangle|\lambda'\rangle|q'\rangle$, since (λ, λ'), (λ, j) and (λ', j) are all trivially related via reversible classical circuits.

To better understand the input space of U_{CG}, we introduce the *model representation* $\mathcal{Q}^d_* := \bigoplus_{\lambda\in\mathbb{Z}^d_{++}} \mathcal{Q}^d_\lambda$, with corresponding matrix $\mathbf{q}^d_*(U) = \sum_\lambda |\lambda\rangle\langle\lambda| \otimes \mathbf{q}^d_\lambda(U)$. The model representation (also sometimes called the *Schwinger representation*) is infinite dimensional and contains each irrep once.[†] Its basis vectors are of the form $|\lambda, q\rangle$ for $\lambda \in \mathbb{Z}^d_{++}$ and $|q\rangle \in Q^d_\lambda$. Since $\mathcal{Q}^d_*$ is infinite-dimensional, we cannot store it on a quantum computer and in this thesis work only with representations $\mathcal{Q}^d_\lambda$ with $|\lambda| \le n$; nevertheless $\mathcal{Q}^d_*$ is a useful abstraction.

Thus U_{CG} decomposes $\mathcal{Q}^d_* \otimes \mathcal{Q}^d_{(1)}$ into irreps. There are two important things to notice about this version of the CG transform. First is that it operates simultaneously on different input irreps. Second is that different input irreps must remain orthogonal, so in order to to maintain unitarity U_{CG} needs

[*] We can similarly obtain the CG series for $\mathcal{S}_n$ by studying the branching from $\mathcal{U}_{d_1 d_2}$ to $\mathcal{U}_{d_1} \otimes \mathcal{U}_{d_2}$. This is a useful tool for studying the relation between spectra of a bipartite density matrix ρ^{AB} and of the reduced density matrices ρ^A and ρ^B[CM04, Kly04].

[†] By contrast, $L^2(\mathcal{U}_d)$, which we will not use, contains $\mathcal{Q}^d_\lambda$ with multiplicity dim $\mathcal{Q}^d_\lambda$.

to keep the information of which irrep we started with. However, since $\lambda' = \lambda + e_j$, this information requires only storing some $j \in [d]$. Thus, U_{CG} is a map from $\mathcal{Q}^d_* \otimes \mathbb{C}^d$ to $\mathcal{Q}^d_* \otimes \mathbb{C}^d$, where the $\mathbb{C}^d$ in the input is the defining representation and the $\mathbb{C}^d$ in the output tracks which irrep we started with.

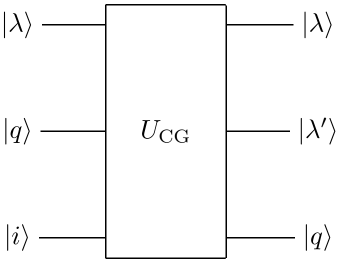


Figure 7-1: Schematic of the Clebsch-Gordan transform. Equivalently, we could replace either the λ output or the λ' output with j.

7.2.2 Constructing the Schur Transform from Clebsch-Gordan Transforms

We now describe how to construct the Schur transform out of a series of Clebsch-Gordan transforms. Suppose we start with an input vector $|i_1, \ldots, i_n\rangle \in (\mathbb{C}^d)^{\otimes n}$, corresponding to the $\mathcal{U}_d$-representation $(\mathcal{Q}^d_{(1)})^{\otimes n}$. According to Schur duality (Eq. (5.16)), to perform the Schur transform it suffices to decompose $(\mathcal{Q}^d_{(1)})^{\otimes n}$ into $\mathcal{U}_d$-irreps. This is because Schur duality means that the multiplicity space of $\mathcal{Q}^d_\lambda$ must be isomorphic to $\mathcal{P}_\lambda$. In other words, if we show that

$$(\mathcal{Q}^d_{(1)})^{\otimes n} \overset{\mathcal{U}_d}{\cong} \bigoplus_{\lambda \in \mathbb{Z}^d_{++}} \mathcal{Q}^d_\lambda \otimes \mathcal{P}'_\lambda, \tag{7.20}$$

then we must have $\mathcal{P}'_\lambda \overset{\mathcal{S}_n}{\cong} \mathcal{P}_\lambda$ when $\lambda \in \mathcal{I}_{d,n}$ and $\mathcal{P}'_\lambda = \{0\}$ otherwise.

To perform the $\mathcal{U}_d$-irrep decomposition of Eq. (7.20), we simply combine each of $|i_1\rangle, \ldots, |i_n\rangle$ using the CG transform, one at a time. We start by inputting $|\lambda^{(1)}\rangle = |(1)\rangle$, $|i_1\rangle$ and $|i_2\rangle$ into U_{CG} which outputs $|\lambda^{(1)}\rangle$ and a superposition of different values of $|\lambda^{(2)}\rangle$ and $|q_2\rangle$. Here $\lambda^{(2)}$ can be either $(2,0)$ or $(1,1)$ and $|q_2\rangle \in Q^d_{\lambda^{(2)}}$. Continuing, we apply U_{CG} to $|\lambda^{(2)}\rangle|q_2\rangle|i_3\rangle$, and output a superposition of vectors of the form $|\lambda^{(2)}\rangle|\lambda^{(3)}\rangle|q_3\rangle$, with $\lambda^{(3)} \in \mathcal{I}_{d,3}$ and $|q_3\rangle \in Q^d_{\lambda^{(3)}}$. Each time we are combining an arbitrary irrep $\lambda^{(k)}$ and an associated basis vector $|q_k\rangle \in Q^d_{\lambda^{(k)}}$, together with a vector from the defining irrep $|i_{k+1}\rangle$. This is repeated for $k = 1, \ldots, n-1$ and the resulting circuit is depicted in Fig. 7-2.

Finally, we are left with a superposition of states of the form $|\lambda^{(1)}, \ldots, \lambda^{(n)}\rangle|q_n\rangle$, where $|q_n\rangle \in Q^d_{\lambda^{(n)}}$, $\lambda^{(k)} \in \mathcal{I}_{d,k}$ and each $\lambda^{(k)}$ is obtained by adding a single box to $\lambda^{(k-1)}$; i.e. $\lambda^{(k)} = \lambda^{(k-1)} + e_{j_k}$ for some $j_k \in [d]$. If we define $\lambda = \lambda^{(n)}$ and $|q\rangle = |q_n\rangle$, then we have the decomposition of Eq. (7.20) with $\mathcal{P}'_\lambda$ spanned by the vectors $|\lambda^{(1)}, \ldots, \lambda^{(n-1)}\rangle$ satisfying the constraints described above. But this is precisely the Young-Yamanouchi basis P_λ that we have defined in Section 7.1! Since the first k qudits transform under $\mathcal{U}_d$ according to $\mathcal{Q}^d_{\lambda^{(k)}}$, Schur duality implies that they also transform under $\mathcal{S}_n$ according to $\mathcal{P}_{\lambda^{(k)}}$. Thus we set $|p\rangle = |\lambda^{(1)}, \ldots, \lambda^{(n-1)}\rangle$ (optionally compressing to $\lceil \log |P_\lambda| \rceil$ qubits using the techniques described in the last section) and obtain the desired $|\lambda\rangle|q\rangle|p\rangle$. As a check on this result, note that each $\lambda^{(k)}$ is invariant under $\mathbf{Q}(\mathcal{U}_d)$ since $U^{\otimes n}$ acts on the first k qubits simply as $U^{\otimes k}$.

If we choose not to perform the poly(n) steps to optimally compress $|\lambda^{(1)}, \ldots, \lambda^{(n-1)}\rangle$, we could instead have our circuit output the equivalent $|j_1, \ldots, j_{n-1}\rangle$, which requires only $n \log d$ qubits and asymptotically no extra running time.

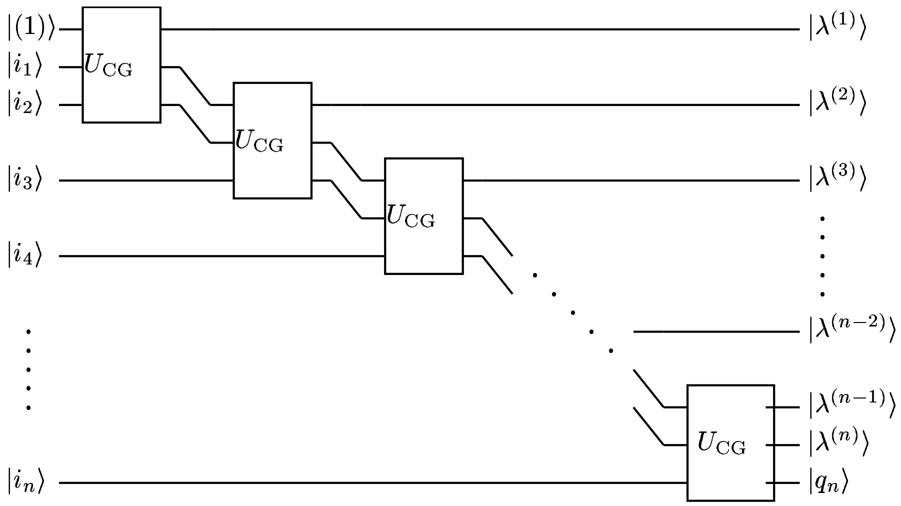


Figure 7-2: Cascading Clebsch-Gordan transforms to produce the Schur transform. Not shown are any ancilla inputs to the Clebsch-Gordan transforms. The structure of inputs and outputs of the Clebsch-Gordan transforms are the same as in Fig. 7-1.

We can now appreciate the similarity between the $\mathcal{U}_d$ CG "add a box" prescription and the $\mathcal{S}_{n-1} \subset \mathcal{S}_n$ branching rule of "remove a box." Schur duality implies that the representations $\mathcal{Q}^d_{\lambda'}$ that are obtained by decomposing $\mathcal{Q}^d_\lambda \otimes \mathcal{Q}^d_{(1)}$ are the same as the $\mathcal{S}_n$-irreps $\mathcal{P}_{\lambda'}$ that include $\mathcal{P}_\lambda$ when restricted to $\mathcal{S}_{n-1}$.

Define $T_{\text{CG}}(n, d, \epsilon)$ to be the time complexity (in terms of number of gates) of performing a single $\mathcal{U}_d$ CG transform to accuracy ϵ on Young diagrams with $\leq n$ boxes. Then the total complexity for the Schur transform is $n \cdot (T_{\text{CG}}(n, d, \epsilon/n) + O(1))$, possibly plus a poly$(n)$ factor for compressing the $\mathcal{P}_\lambda$ register to $\lceil \log \dim \mathcal{P}_\lambda \rceil$ qubits (as is required for applications such as data compression and entanglement concentration, cf. Section 6.3). In the next section we will show that $T_{\text{CG}}(n, d, \epsilon)$ is poly$(\log n, d, \log 1/\epsilon)$, but first we give a step-by-step description of the algorithm for the Schur transform.

Algorithm: Schur transform (plus optional compression)

```
Inputs: (1) Classical registers $d$ and $n$. (2) An $n$ qudit quantum register $|i_1, \ldots, i_n\rangle$.
Outputs: Quantum registers $|\lambda\rangle|q\rangle|p\rangle$, with $\lambda \in \mathcal{I}_{d,n}$, $q \in Q^d_\lambda$ and $p \in P_\lambda$.
Runtime: $n \cdot (T_{\text{CG}}(n, d, \epsilon/n) + O(1))$ to achieve accuracy $\epsilon$.
         (Optionally plus poly$(n)$ to compress the $\mathcal{P}_\lambda$ register to $\lceil \log \dim \mathcal{P}_\lambda \rceil$ qubits.)
Procedure:
1. Initialize $|\lambda^{(1)}\rangle := |(1)\rangle$ and $|q_1\rangle = |i_1\rangle$.
2. For $k = 1, \ldots, n-1$:
3.     Apply $U_{\text{CG}}$ to $|\lambda^{(k)}\rangle|q_k\rangle|i_{k+1}\rangle$ to obtain output $|j_k\rangle|\lambda^{(k+1)}\rangle|q_{k+1}\rangle$, where $\lambda^{(k+1)} = \lambda^{(k)} + e_{j_k}$.
4. Output $|\lambda\rangle := |\lambda^{(n)}\rangle$, $|q\rangle := |q_n\rangle$ and $|p\rangle := |j_1, \ldots, j_{n-1}\rangle$.
5. (Optionally use Eq. (7.11) to reversibly map $|j_1, \ldots, j_{n-1}\rangle$ to an integer $p \in [\dim \mathcal{P}_\lambda]$.)
```

This algorithm will be made efficient in the next section, where we efficiently construct the CG transform for $\mathcal{U}_d$, proving that $T_{\text{CG}}(n, d, \epsilon) = \text{poly}(\log n, d, \log 1/\epsilon)$.

7.3 Efficient circuits for the Clebsch-Gordan transform

We now turn to the actual construction of the circuit for the Clebsch-Gordan transform described in Section 7.2.1. To get a feel for the what will be necessary, we start by giving a circuit for the CG transform that is efficient when d is constant; i.e. it has complexity $n^{O(d^2)}$, which is poly(n) for any constant value of d.

First recall that $\dim \mathcal{Q}_\lambda^d \leq (n+1)^{d^2}$. Thus, controlled on λ, we want to construct a unitary transform on a D-dimensional system for $D = \max_{\lambda\in\mathcal{I}_{d,n}} \dim \mathcal{Q}_\lambda^d = \text{poly}(n)$. There are classical algorithms[Lou70] to compute matrix elements of U_{CG} to an accuracy ϵ_1 in time $\text{poly}(D)\,\text{poly}\log(1/\epsilon_1)$. Once we have calculated all the relevant matrix elements (of which there are only polynomially many), we can (again in time $\text{poly}(D)\,\text{poly}\log(1/\epsilon)$) decompose U_{CG} into $D^2\,\text{poly}\log(D)$ elementary one and two-qubit operations[SBM04, RZBB94, Bar95, NC00]. These can in turn be approximated to accuracy ϵ_2 by products of unitary operators from a fixed finite set (such as Clifford operators and a $\pi/8$ rotation) with a further overhead of $\text{poly}\log(1/\epsilon_2)$[DN05, KSV02]. We can either assume the relevant classical computations (such as decomposing the $D\times D$ matrix into elementary gates) are performed coherently on a quantum computer, or as part of a polynomial-time classical Turing machine which outputs the quantum circuit. In any case, the total complexity is $\text{poly}(n, \log 1/\epsilon)$ if the desired final accuracy is ϵ and d is held constant.

The goal of this section is to reduce this running time to $\text{poly}(n, d, \log(1/\epsilon))$; in fact, we will achieve circuits of size $\text{poly}(d, \log n, \log(1/\epsilon))$. To do so, we will reduce the $\mathcal{U}_d$ CG transform to two components; first, a $\mathcal{U}_{d-1}$ CG transform, and second, a $d\times d$ unitary matrix whose entries can be computed classically in $\text{poly}(d, \log n, 1/\epsilon)$ steps. After computing all d^2 entries, the second component can then be implemented with $\text{poly}(d, \log 1/\epsilon)$ gates according to the above arguments.

This reduction from the $\mathcal{U}_d$ CG transform to the $\mathcal{U}_{d-1}$ CG transform is a special case of the Wigner-Eckart Theorem, which we review in Section 7.3.1. Then, following [BL68, Lou70], we use the Wigner-Eckart Theorem to give an efficient recursive construction for U_{CG} in Section 7.3.2. Putting everything together, we obtain a quantum circuit for the Schur transform that is accurate to within ϵ and runs in time $n\cdot\text{poly}(\log n, d, \log 1/\epsilon)$, optionally plus an additional poly(n) time to compress the $|p\rangle$ register.

7.3.1 The Wigner-Eckart Theorem and Clebsch-Gordan transform

In this section, we introduce the concept of an irreducible tensor operator, which we use to state and prove the Wigner-Eckart Theorem. Here we will find that the CG transform is a key part of the Wigner-Eckart Theorem, while in the next section we will turn this around and use the Wigner-Eckart Theorem to give a recursive decomposition of the CG transform.

Suppose $(\mathbf{r}_1, V_1)$ and $(\mathbf{r}_2, V_2)$ are representations of $\mathcal{U}_d$. Recall that $\text{Hom}(V_1, V_2)$ is a representation of $\mathcal{U}_d$ under the map $T \to \mathbf{r}_2(U)T\mathbf{r}_1(U)^{-1}$ for $T \in \text{Hom}(V_1, V_2)$. If $\boldsymbol{T} = \{T_1, T_2, \ldots\} \subset \text{Hom}(V_1, V_2)$ is a basis for a $\mathcal{U}_d$-invariant subspace of $\text{Hom}(V_1, V_2)$, then we call $\boldsymbol{T}$ a *tensor operator*. Note that a tensor operator $\mathbf{T}$ is a collection of operators $\{T_i\}$ indexed by i, just as a tensor (or vector) is a collection of scalars labeled by some index. For example, the Pauli matrices $\{\sigma_x, \sigma_y, \sigma_z\} \subset \text{Hom}(\mathbb{C}^2, \mathbb{C}^2)$ comprise a tensor operator, since conjugation by $\mathcal{U}_2$ preserves the subspace that they span.

Since $\text{Hom}(V_1, V_2)$ is a representation of $\mathcal{U}_d$, it can be decomposed into irreps. If $\boldsymbol{T}$ is a basis for one of these irreps, then we call it an *irreducible tensor operator*. For example, the Pauli matrices mentioned above comprise an irreducible tensor operator, corresponding to the three-dimensional irrep $\mathcal{Q}_{(2)}^2$. Formally, we say that $\boldsymbol{T}^\nu = \{T_{q_\nu}^\nu\}_{q_\nu\in Q_\nu^d} \subset \text{Hom}(V_1, V_2)$ is an irreducible tensor operator (corresponding to the irrep $\mathcal{Q}_\nu^d$) if for all $U \in \mathcal{U}_d$ we have

$$\mathbf{r}_2(U)T_{q_\nu}^\nu\mathbf{r}_1(U)^{-1} = \sum_{q'_\nu\in Q_\nu^d} \langle q'_\nu|\mathbf{q}_\nu^d(U)|q_\nu\rangle T_{q'_\nu}^\nu. \tag{7.21}$$

Now assume that V_1 and V_2 are irreducible (say $V_1 = \mathcal{Q}_\mu^d$ and $V_2 = \mathcal{Q}_\lambda^d$), since if they are not, we could always decompose $\mathrm{Hom}(V_1, V_2)$ into a direct sum of homomorphisms from an irrep in V_1 to an irrep in V_2. We can decompose $\mathrm{Hom}(\mathcal{Q}_\mu^d, \mathcal{Q}_\lambda^d)$ into irreps using Eq. (5.3) and the identity $\mathrm{Hom}(A, B) \cong A^* \otimes B$ as follows:

$$\begin{aligned}
\mathrm{Hom}(\mathcal{Q}_\mu^d, \mathcal{Q}_\lambda^d) &\overset{\mathcal{U}_d}{\cong} \bigoplus_{\nu\in\mathbb{Z}_{++}^d} \mathcal{Q}_\nu^d \otimes \mathrm{Hom}(\mathcal{Q}_\nu^d, \mathrm{Hom}(\mathcal{Q}_\mu^d, \mathcal{Q}_\lambda^d))^{\mathcal{U}_d} \\
&\overset{\mathcal{U}_d}{\cong} \bigoplus_{\nu\in\mathbb{Z}_{++}^d} \mathcal{Q}_\nu^d \otimes \mathrm{Hom}(\mathcal{Q}_\nu^d, (\mathcal{Q}_\mu^d)^* \otimes \mathcal{Q}_\lambda^d)^{\mathcal{U}_d} \\
&\overset{\mathcal{U}_d}{\cong} \bigoplus_{\nu\in\mathbb{Z}_{++}^d} \mathcal{Q}_\nu^d \otimes \big((\mathcal{Q}_\mu^d)^* \otimes (\mathcal{Q}_\nu^d)^* \otimes \mathcal{Q}_\lambda^d\big)^{\mathcal{U}_d} \\
&\overset{\mathcal{U}_d}{\cong} \bigoplus_{\nu\in\mathbb{Z}_{++}^d} \mathcal{Q}_\nu^d \otimes \mathrm{Hom}(\mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d, \mathcal{Q}_\lambda^d)^{\mathcal{U}_d}
\end{aligned} \tag{7.22}$$

Now consider a particular irreducible tensor operator $\mathbf{T}^\nu \subset \mathrm{Hom}(\mathcal{Q}_\mu^d, \mathcal{Q}_\lambda^d)$ with components $T_{q_\nu}^\nu$ where q_ν ranges over Q_ν^d. We can define a linear operator $\hat{T} : \mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d \to \mathcal{Q}_\lambda^d$ by letting

$$\hat{T}|q_\mu\rangle|q_\nu\rangle := T_{q_\nu}^\nu|q_\mu\rangle \tag{7.23}$$

for all $q_\mu \in Q_\mu^d, q_\nu \in Q_\nu^d$ and extending it to the rest of $\mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d$ by linearity. By construction, $\hat{T} \in \mathrm{Hom}(\mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d, \mathcal{Q}_\lambda^d)$, but we claim that in addition $\hat{T}$ is invariant under the action of $\mathcal{U}_d$; i.e. that it lies in $\mathrm{Hom}(\mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d, \mathcal{Q}_\lambda^d)^{\mathcal{U}_d}$. To see this, apply Eqns. (7.21) and (7.23) to show that for any $U \in \mathcal{U}_d$, $q_\mu \in Q_\mu^d$ and $q_\nu \in Q_\nu^d$, we have

$$\begin{aligned}
\mathbf{q}_\lambda^d(U)\hat{T}\big[\mathbf{q}_\mu^d(U)^{-1} \otimes \mathbf{q}_\nu^d(U)^{-1}\big]|q_\mu\rangle|q_\nu\rangle &= \sum_{q'_\nu\in Q_\nu^d} \langle q'_\nu|\mathbf{q}_\nu^d(U)^{-1}|q_\nu\rangle \mathbf{q}_\lambda^d(U) T_{q'_\nu}^\nu \mathbf{q}_\mu^d(U)^{-1}|q_\mu\rangle \\
&= \sum_{q'_\nu, q''_\nu\in Q_\nu^d} \langle q''_\nu|\mathbf{q}_\nu^d(U)|q'_\nu\rangle\langle q'_\nu|\mathbf{q}_\nu^d(U)^{-1}|q_\nu\rangle T_{q''_\nu}^\nu|q_\mu\rangle \\
&= T_{q_\nu}^\nu|q_\mu\rangle = \hat{T}|q_\mu\rangle|q_\nu\rangle.
\end{aligned} \tag{7.24}$$

Now, fix an orthonormal basis for $\mathrm{Hom}(\mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d, \mathcal{Q}_\lambda^d)^{\mathcal{U}_d}$ and call it $M_{\mu,\nu}^\lambda$. Then we can expand $\hat{T}$ in this basis as

$$\hat{T} = \sum_{\alpha\in M_{\mu,\nu}^\lambda} \hat{T}_\alpha \cdot \alpha, \tag{7.25}$$

where the $\hat{T}_\alpha$ are scalars. Thus

$$\langle q_\lambda|T_{q_\nu}^\nu|q_\mu\rangle = \sum_{\alpha\in M_{\mu,\nu}^\lambda} \hat{T}_\alpha\langle q_\lambda|\alpha|q_\mu, q_\nu\rangle. \tag{7.26}$$

This last expression $\langle q_\lambda|\alpha|q_\mu, q_\nu\rangle$ bears a striking resemblance to the CG transform. Indeed, note that the multiplicity space $\mathrm{Hom}(\mathcal{Q}_\lambda^d, \mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d)^{\mathcal{U}_d}$ from Eq. (5.4) is the dual of $\mathrm{Hom}(\mathcal{Q}_\mu^d \otimes \mathcal{Q}_\nu^d, \mathcal{Q}_\lambda^d)^{\mathcal{U}_d}$ (which contains α), meaning that we can map between the two by taking the transpose. In fact, taking the conjugate transpose of Eq. (5.5) gives $\langle q_\lambda|\alpha = \langle q_\lambda, \alpha^\dagger|U_{\mathrm{CG}}^{\mu,\nu}$. Thus

$$\langle q_\lambda|\alpha|q_\mu, q_\nu\rangle = \langle q_\lambda, \alpha^\dagger|U_{\mathrm{CG}}^{\mu,\nu}|q_\mu, q_\nu\rangle. \tag{7.27}$$

The arguments in the last few paragraphs constitute a proof of the Wigner-Eckart theorem[Mes62], which is stated as follows:

Theorem 7.2 (Wigner-Eckart). *For any irreducible tensor operator* $\mathbf{T}^\nu = \{T^\nu_{q_\nu}\}_{q_\nu \in Q^d_\nu} \subset \operatorname{Hom}(\mathcal{Q}^d_\mu, \mathcal{Q}^d_\lambda)$*, there exist* $\hat{T}_\alpha \in \mathbb{C}$ *for each* $\alpha \in M^\lambda_{\mu,\nu}$ *such that for all* $|q_\mu\rangle \in \mathcal{Q}^d_\mu$*,* $|q_\nu\rangle \in \mathcal{Q}^d_\nu$ *and* $|q_\lambda\rangle \in \mathcal{Q}^d_\lambda$*:*

$$\langle q_\lambda | T^\nu_{q_\nu} | q_\mu \rangle = \sum_{\alpha \in M^\lambda_{\mu,\nu}} \hat{T}_\alpha \langle q_\lambda, \alpha^\dagger | U^{\mu,\nu}_{CG} | q_\mu, q_\nu \rangle. \tag{7.28}$$

Thus, the action of tensor operators can be related to a component $\hat{T}_\alpha$ that is invariant under $\mathcal{U}_d$ and a component that is equivalent to the CG transform. We will use this in the next section to derive an efficient quantum circuit for the CG transform.

7.3.2 A recursive construction of the Clebsch-Gordan transform

In this section we show how the $\mathcal{U}_d$ CG transform (which here we call $U^{[d]}_{\mathrm{CG}}$) can be efficiently reduced to the $\mathcal{U}_{d-1}$ CG transform (which we call $U^{[d-1]}_{\mathrm{CG}}$). Our strategy, following [BL68], will be to express $U^{[d]}_{\mathrm{CG}}$ in terms of $\mathcal{U}_{d-1}$ tensor operators and then use the Wigner-Eckart Theorem to express it in terms of $U^{[d-1]}_{\mathrm{CG}}$. After we have explained this as a relation among operators, we describe a quantum circuit for $U^{[d]}_{\mathrm{CG}}$ that uses $U^{[d-1]}_{\mathrm{CG}}$ as a subroutine.

First, we express $U^{[d]}_{\mathrm{CG}}$ as a $\mathcal{U}_d$ tensor operator. For $\mu \in \mathbb{Z}^d_{++}$, $|q\rangle \in Q^d_\mu$ and $i \in [d]$, we can expand $U^{[d]}_{\mathrm{CG}}|\mu\rangle|q\rangle|i\rangle$ as

$$U^{[d]}_{\mathrm{CG}}|\mu\rangle|q\rangle|i\rangle = |\mu\rangle \sum_{\substack{j\in[d] \text{ s.t.} \\ \mu+e_j \in \mathbb{Z}^d_{++}}} \sum_{q' \in Q^d_{\mu+e_j}} C^{\mu,j}_{q,i,q'} |\mu + e_j\rangle |q'\rangle. \tag{7.29}$$

for some coefficients $C^{\mu,j}_{q,i,q'} \in \mathbb{C}$. Now define operators $T^{\mu,j}_i : \mathcal{Q}^d_\mu \to \mathcal{Q}^d_{\mu+e_j}$ by

$$T^{\mu,j}_i = \sum_{q \in Q^d_\mu} \sum_{q' \in Q^d_{\mu+e_j}} C^{\mu,j}_{q,i,q'} |q'\rangle\langle q|, \tag{7.30}$$

so that $U^{[d]}_{\mathrm{CG}}$ decomposes as

$$U^{[d]}_{\mathrm{CG}}|\mu\rangle|q\rangle|i\rangle = |\mu\rangle \sum_{\substack{j\in[d] \text{ s.t.} \\ \mu+e_j \in \mathbb{Z}^d_{++}}} |\mu + e_j\rangle T^{\mu,j}_i |q\rangle. \tag{7.31}$$

Thus $U^{[d]}_{\mathrm{CG}}$ can be understood in terms of the maps $T^{\mu,j}_i$, which are irreducible tensor operators in $\operatorname{Hom}(\mathcal{Q}^d_\mu, \mathcal{Q}^d_{\mu+e_j})$ corresponding to the irrep $\mathcal{Q}^d_{(1)}$. (This is unlike the notation of the last section in which the superscript denoted the irrep corresponding to the tensor operator.)

The plan for the rest of the section is to decompose the $T^{\mu,j}_i$ operators under the action of $\mathcal{U}_{d-1}$, so that we can apply the Wigner-Eckart theorem. This involves decomposing three different $\mathcal{U}_d$ irreps into $\mathcal{U}_{d-1}$ irreps: the input space $\mathcal{Q}^d_\mu$, the output space $\mathcal{Q}^d_{\mu+e_j}$ and the space $\mathcal{Q}^d_{(1)}$ corresponding to the subscript i. Once we have done so, the Wigner-Eckart Theorem gives an expression for $T^{\mu,j}_i$ (and hence for $U^{[d]}_{\mathrm{CG}}$) in terms of $U^{[d-1]}_{\mathrm{CG}}$ and a small number of coefficients, known as *reduced Wigner coefficients*. These coefficients can be readily calculated, and in the next section we cite a formula from [BL68] for doing so.

First, we examine the decomposition of $\mathcal{Q}^d_{(1)}$, the $\mathcal{U}_d$-irrep according to which the $T^{\mu,j}_i$ transform. Recall that $\mathcal{Q}^d_{(1)} \overset{\mathcal{U}_{d-1}}{\cong} \mathcal{Q}^{d-1}_{(0)} \oplus \mathcal{Q}^{d-1}_{(1)}$. In terms of the tensor operator we have defined, this means that $T^{\mu,j}_d$ is an irreducible $\mathcal{U}_{d-1}$ tensor operator corresponding to the trivial irrep $\mathcal{Q}^{d-1}_{(0)}$ and $\{T^{\mu,j}_1, \ldots, T^{\mu,j}_{d-1}\}$ comprise an irreducible $\mathcal{U}_{d-1}$ tensor operator corresponding to the defining irrep $\mathcal{Q}^{d-1}_{(1)}$.

Next, we would like to decompose $\mathrm{Hom}(\mathcal{Q}^d_\mu, \mathcal{Q}^d_{\mu+e_j})$ into maps between irreps of $\mathcal{U}_{d-1}$. This is slightly more complicated, but can be derived from the $\mathcal{U}_{d-1} \subset \mathcal{U}_d$ branching rule introduced in Section 7.1.2. Recall that $\mathcal{Q}^d_\mu \overset{\mathcal{U}_{d-1}}{\cong} \bigoplus_{\mu' \precsim \mu} \mathcal{Q}^{d-1}_{\mu'}$, and similarly $\mathcal{Q}^d_{\mu+e_j} \overset{\mathcal{U}_{d-1}}{\cong} \bigoplus_{\mu'' \precsim \mu+e_j} \mathcal{Q}^{d-1}_{\mu''}$. This is the moment that we anticipated in Section 7.1.2 when we chose our set of basis vectors Q^d_μ to respect these decompositions. As a result, a vector $|q\rangle \in Q^d_\mu$ can be expanded as $q = (q_{d-1}, q_{d-2}, \ldots, q_1) = (\mu', q_{(d-2)})$ with $q_{d-1} = \mu' \in \mathbb{Z}^{d-1}_{++}$, $\mu' \precsim \mu$ and $|q_{(d-2)}\rangle = |q_{d-2}, \ldots, q_1\rangle \in Q^{d-1}_{\mu'}$. In other words, we will separate vectors in Q^d_μ into a $\mathcal{U}_{d-1}$ irrep label $\mu' \in \mathbb{Z}^{d-1}_{++}$ and a basis vector from $\mathcal{Q}^{d-1}_{\mu'}$.

This describes how to decompose the spaces $\mathcal{Q}^d_\mu$ and $\mathcal{Q}^d_{\mu+e_j}$. To extend this to decomposition of $\mathrm{Hom}(\mathcal{Q}^d_\mu, \mathcal{Q}^d_{\mu+e_j})$, we use the canonical isomorphism $\mathrm{Hom}(\bigoplus_x A_x, \bigoplus_y B_y) \cong \bigoplus_{x,y} \mathrm{Hom}(A_x, B_y)$, which holds for any sets of vector spaces $\{A_x\}$ and $\{B_y\}$. Thus

$$\mathrm{Hom}(\mathcal{Q}^d_\mu, \mathcal{Q}^d_{\mu+e_j}) \overset{\mathcal{U}_{d-1}}{\cong} \bigoplus_{\mu' \precsim \mu} \bigoplus_{\mu'' \precsim \mu+e_j} \mathrm{Hom}(\mathcal{Q}^{d-1}_{\mu'}, \mathcal{Q}^{d-1}_{\mu''}). \tag{7.32a}$$

Sometimes we will find it convenient to denote the $\mathcal{Q}^{d-1}_{\mu'}$ subspace of $\mathcal{Q}^d_\mu$ by $\mathcal{Q}^{d-1}_{\mu'} \subset \mathcal{Q}^d_\mu$, so that Eq. (7.32a) becomes

$$\mathrm{Hom}(\mathcal{Q}^d_\mu, \mathcal{Q}^d_{\mu+e_j}) \overset{\mathcal{U}_{d-1}}{\cong} \bigoplus_{\mu' \precsim \mu} \bigoplus_{\mu'' \precsim \mu+e_j} \mathrm{Hom}(\mathcal{Q}^{d-1}_{\mu'} \subset \mathcal{Q}^d_\mu, \mathcal{Q}^{d-1}_{\mu''} \subset \mathcal{Q}^d_{\mu+e_j}). \tag{7.32b}$$

According to Eq. (7.32) (either version), we can decompose $T^{\mu,j}_i$ as

$$T^{\mu,j}_i = \sum_{\mu' \precsim \mu} \sum_{\mu'' \precsim \mu+e_j} |\mu''\rangle\langle\mu'| \otimes T^{\mu,j,\mu',\mu''}_i. \tag{7.33}$$

Here $T^{\mu,j,\mu',\mu''}_i \in \mathrm{Hom}(\mathcal{Q}^{d-1}_{\mu'} \subset \mathcal{Q}^d_\mu, \mathcal{Q}^{d-1}_{\mu''} \subset \mathcal{Q}^d_{\mu+e_j})$ and we have implicitly decomposed $|q\rangle \in Q^d_\mu$ into $|\mu'\rangle|q_{(d-2)}\rangle$.

The next step is to decompose the representions in Eq. (7.32) into irreducible components. In fact, we are not interested in the entire space $\mathrm{Hom}(\mathcal{Q}^{d-1}_{\mu'}, \mathcal{Q}^{d-1}_{\mu''})$, but only the part that is equivalent to $\mathcal{Q}^{d-1}_{(1)}$ or $\mathcal{Q}^{d-1}_{(0)}$, depending on whether $i \in [d-1]$ or $i = d$ (since $T^{\mu,j,\mu',\mu''}_i$ transforms according to $\mathcal{Q}^{d-1}_{(1)}$ if $i \in \{1, \ldots, d-1\}$ and according to $\mathcal{Q}^{d-1}_{(0)}$ if $i = d$). This knowledge of how $T^{\mu,j,\mu',\mu''}_i$ transforms under $\mathcal{U}_{d-1}$ will give us two crucial simplifications: first, we can greatly reduce the range of μ'' for which $T^{\mu,j,\mu',\mu''}_i$ is nonzero, and second, we can apply the Wigner-Eckart theorem to describe $T^{\mu,j,\mu',\mu''}_i$ in terms of $U^{[d-1]}_{\mathrm{CG}}$.

The simplest case is $\mathcal{Q}^{d-1}_{(0)}$, when $i = d$: according to Schur's Lemma the invariant component of $\mathrm{Hom}(\mathcal{Q}^{d-1}_{\mu'}, \mathcal{Q}^{d-1}_{\mu''})$ is zero if $\mu' \neq \mu''$ and consists of the matrices proportional to $I_{\mathcal{Q}^{d-1}_{\mu'}}$ if $\mu' = \mu''$. In other words $T^{\mu,j,\mu',\mu''}_d = 0$ unless $\mu' = \mu''$, in which case $T^{\mu,j,\mu',\mu'}_d := \hat{T}^{\mu,j,\mu',0} I_{\mathcal{Q}^{d-1}_{\mu'}}$ for some scalar $\hat{T}^{\mu,j,\mu',0}$. (The final superscript 0 will later be convenient when we want a single notation to encompass both the $i = d$ and the $i \in \{1, \ldots, d-1\}$ cases.)

The $\mathcal{Q}^{d-1}_{(1)}$ case, which occurs when $i \in \{1, \ldots, d-1\}$, is more interesting. We will simplify the $T_i^{\mu,j,\mu',\mu''}$ operators (for $i = 1, \dots, d-1$) in two stages: first using the branching rules from Section 7.1.2 to reduce the number of nonzero terms and then by applying the Wigner-Eckart theorem to find an exact expression for them. Begin by recalling from Eq. (7.22) that the multiplicity of $\mathcal{Q}_{(1)}^{d-1}$ in the isotypic decomposition of $\mathrm{Hom}(\mathcal{Q}_{\mu'}^{d-1}, \mathcal{Q}_{\mu''}^{d-1})$ is given by $\dim \mathrm{Hom}(\mathcal{Q}_{\mu'}^{d-1} \otimes \mathcal{Q}_{(1)}^{d-1}, \mathcal{Q}_{\mu''}^{d-1})^{\mathcal{U}_{d-1}}$. According to the $\mathcal{U}_{d-1}$ CG "add a box" prescription (Eq. (7.16)), this is one if $\mu' \in \mu'' - \Box$ and zero otherwise. Thus if $i \in [d-1]$, then $T_i^{\mu,j,\mu',\mu''}$ is zero unless $\mu'' = \mu' + e_{j'}$ for some $j' \in [d-1]$. Since we need not consider all possible μ'', we can define $T_i^{\mu,j,\mu',j'} := T_i^{\mu,j,\mu',\mu'+e_{j'}}$. This notation can be readily extended to cover the case when $i = d$; define $e_0 = 0$, so that the only nonzero operators for $i = d$ are of the form $T_d^{\mu,j,\mu',0} := T_d^{\mu,j,\mu',\mu'} = \hat{T}^{\mu,j,\mu',0} I_{\mathcal{Q}_{\mu'}^{d-1}}$. Thus, we can replace Eq. (7.33) with

$$T_i^{\mu,j} = \sum_{\mu' \precsim \mu} \sum_{j'=0}^{d-1} |\mu' + e_{j'}\rangle\langle\mu'| \otimes T_i^{\mu,j,\mu',\mu'+e_{j'}}. \tag{7.34}$$

Now we show how to apply the Wigner-Eckart theorem to the $i \in [d-1]$ case. The operators $T_i^{\mu,j,\mu',j'}$ map $\mathcal{Q}_{\mu'}^{d-1}$ to $\mathcal{Q}_{\mu'+e_{j'}}^{d-1}$ and comprise an irreducible $\mathcal{U}_{d-1}$ tensor operator corresponding to the irrep $\mathcal{Q}_{(1)}^{d-1}$. This means we can apply the Wigner-Eckart Theorem and since the multiplicity of $\mathcal{Q}_{\mu'+e_{j'}}^{d-1}$ in $\mathcal{Q}_{\mu'}^{d-1} \otimes \mathcal{Q}_{(1)}^{d-1}$ is one, the sum over the multiplicity label α has only a single term. The theorem implies the existence of a set of scalars $\hat{T}^{\mu,j,\mu',j'}$ such that for any $|q\rangle \in Q_{\mu'}^{d-1}$ and $|q'\rangle \in Q_{\mu'+e_{j'}}^{d-1}$,

$$\langle q'|T_i^{\mu,j,\mu',j'}|q\rangle = \hat{T}^{\mu,j,\mu',j'} \langle \mu', \mu' + e_{j'}, q'|U_{\mathrm{CG}}^{[d-1]}|\mu', q, i\rangle. \tag{7.35}$$

Sometimes the matrix elements of U_{CG} or $T_i^{\mu,j,\mu',j'}$ are called *Wigner coefficients* and the $\hat{T}^{\mu,j,\mu',j'}$ are known as *reduced Wigner coefficients.*

Let us now try to interpret these equations operationally. Eq. (7.31) reduces the $\mathcal{U}_d$ CG transform to a $\mathcal{U}_d$ tensor operator, Eq. (7.34) decomposes this tensor operator into d^2 different $\mathcal{U}_{d-1}$ tensor operators (weighted by the $\hat{T}^{\mu,j,\mu',j'}$ coefficients) and Eq. (7.35) turns this into a $\mathcal{U}_{d-1}$ CG transform followed by a $d \times d$ unitary matrix. The coefficients for this matrix are the $\hat{T}^{\mu,j,\mu',j'}$, which we will see in the next section can be efficiently computed by conditioning on μ and μ'.

Now we spell this recursion out in more detail. Suppose we wish to apply $U_{\mathrm{CG}}^{[d]}$ to $|\mu\rangle|q\rangle|i\rangle = |\mu\rangle|\mu'\rangle|q_{(d-2)}\rangle|i\rangle$, for some $i \in \{1, \dots, d-1\}$. Then Eq. (7.35) indicates that we should first apply $U_{\mathrm{CG}}^{[d-1]}$ to $|\mu'\rangle|q_{(d-2)}\rangle|i\rangle$ to obtain output that is a superposition over states $|\mu' + e_{j'}\rangle|j'\rangle|q'_{(d-2)}\rangle$ for $j' \in \{1, \dots, d-1\}$ and $|q'_{(d-2)}\rangle \in Q_{\mu'+e_{j'}}^{d-1}$. Then, controlled by μ and μ', we want to map the $(d-1)$-dimensional $|j'\rangle$ register into the d-dimensional $|j\rangle$ register, which will then tell us the output irrep $\mathcal{Q}_{\mu+e_j}^d$. According to Eq. (7.35), the coefficients of this $d \times (d-1)$ matrix are given by the reduced Wigner coefficients $\hat{T}^{\mu,j,\mu',j'}$, so we will denote the overall matrix $\hat{T}_{\mu,\mu'}^{[d]} := \sum_{j,j'} \hat{T}^{\mu,j,\mu'+e_{j'},j'}|j\rangle\langle j'|$.[*] The resulting circuit is depicted in Fig. 7-3: a $\mathcal{U}_{d-1}$ CG transform is followed by the $\hat{T}^{[d]}$ operator, which is defined to be

$$\hat{T}^{[d]} = \sum_{\mu' \precsim \mu} \sum_{j,j'} \hat{T}^{\mu,j,\mu',j'} \, |\mu\rangle\langle\mu| \otimes |\mu + e_j\rangle\langle\mu'| \otimes |\mu' + e_{j'}\rangle\langle\mu' + e_{j'}| \,. \tag{7.36}$$

Then Fig. 7-4 shows how $\hat{T}^{[d]}$ can be expressed as a $d \times (d-1)$ matrix $\hat{T}_{\mu,\mu'}^{[d]}$ that is controlled by μ and μ'. In fact, once we consider the $i = d$ case in the next paragraph, we will find that $\hat{T}_{\mu,\mu'}^{[d]}$ is actually a $d \times d$ unitary matrix. In the next section, we will then show how the individual reduced

[*] The reason why $\mu' + e_{j'}$ appears in the superscript rather than μ' is that after applying $\hat{T}_{\mu,\mu'}^{[d]}$ we want to keep a record of $\mu' + e_{j'}$ rather than of μ'. This is further illustrated in Fig. 7-4.

Wigner coefficients $\hat{T}^{\mu,j,\mu',j'}$ can be efficiently computed, so that ultimately $\hat{T}^{[d]}_{\mu,\mu'}$ can be implemented in time $\mathrm{poly}(d, \log 1/\epsilon)$.

Now we turn to the case of $i = d$. The circuit is much simpler, but we also need to explain how it works in coherent superposition with the $i \in [d-1]$ case. Since $i = d$ corresponds to the trivial representation of $\mathcal{U}_{d-1}$, the $U_{\mathrm{CG}}^{[d-1]}$ operation is not performed. Instead, $|\mu'\rangle$ and $|q_{(d-2)}\rangle$ are left untouched and the $|i\rangle = |d\rangle$ register is relabeled as a $|j'\rangle = |0\rangle$ register. We can combine this relabeling operation with $U_{\mathrm{CG}}^{[d-1]}$ in the $i \in [d-1]$ case by defining

$$\widetilde{U}_{\mathrm{CG}}^{[d-1]} := \left(|0\rangle\langle d| \otimes \sum_{\mu' \in \mathbb{Z}^{d-1}_{++}} |\mu'\rangle\langle\mu'| \right) \otimes I_{\mathcal{Q}^{d-1}_{\mu'}} + U_{\mathrm{CG}}^{[d-1]}. \tag{7.37}$$

This ends up mapping $i \in \{1, \ldots, d\}$ to $j' \in \{0, \ldots, d-1\}$ while mapping $\mathcal{Q}^{d-1}_{\mu'}$ to $\mathcal{Q}^{d-1}_{\mu'+e_{j'}}$. Now we can interpret the sum on j' in the above definitions of $\hat{T}^{[d]}$ and $\hat{T}^{[d]}_{\mu,\mu'}$ as ranging over $\{0, \ldots, d-1\}$, so that $\hat{T}^{[d]}_{\mu,\mu'}$ is a $d \times d$ unitary matrix. We thus obtain the circuit in Fig. 7-3 with the implementation of $\hat{T}^{[d]}$ depicted in Fig. 7-4.

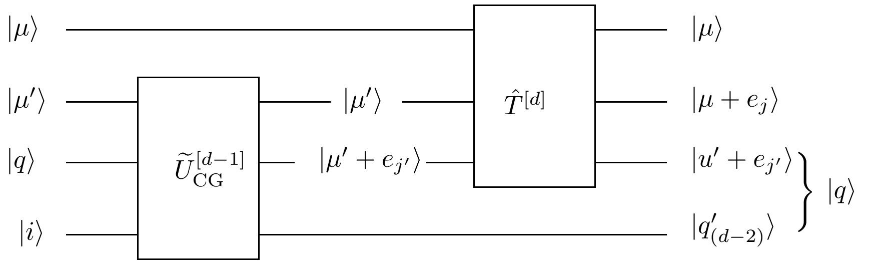


Figure 7-3: The $\mathcal{U}_d$ CG transform, $U_{\mathrm{CG}}^{[d]}$, is decomposed into a $\mathcal{U}_{d-1}$ CG transform $\widetilde{U}_{\mathrm{CG}}^{[d-1]}$ (see Eq. (7.37)) and a reduced Wigner operator $\hat{T}^{[d]}$. In Fig. 7-4 we show how to reduce the reduced Wigner operator to a $d \times d$ matrix conditioned on μ and $\mu' + e_{j'}$.

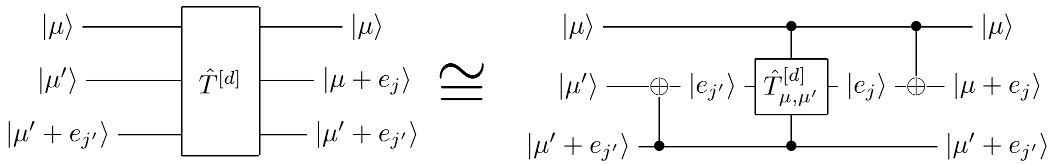


Figure 7-4: The reduced Wigner transform $\hat{T}^{[d]}$ can be expressed as a $d \times d$ rotation whose coefficients are controlled by μ and $\mu' + e_{j'}$.

We have now reduced the problem of performing the CG transform $U_{\mathrm{CG}}^{[d]}$ to the problem of computing reduced Wigner coefficients $\hat{T}^{\mu,j,\mu',j'}$.

7.3.3 Efficient Circuit for the Reduced Wigner Operator

The method of Biedenharn and Louck[BL68] allows us to compute reduced Wigner coefficients for the cases we are interested in. This will allow us to construct an efficient circuit to implement the controlled-$\hat{T}$ operator to accuracy ϵ using an overhead which scales like $\mathrm{poly}(\log n, d, \log(\epsilon^{-1}))$.

To compute $\hat{T}^{\mu,j,\mu',j'}$, we first introduce the vectors $\widetilde{\mu} := \mu + \sum_{j=1}^{d}(d-j)e_j$ and $\widetilde{\mu}' := \mu' + \sum_{j=1}^{d-1}(d-1-j)e_j$. Also define $S(j-j')$ to be 1 if $j \geq j'$ and -1 if $j < j'$. Then according to Eq. (38) in Ref [BL68],

$$\hat{T}^{\mu,j,\mu',j'} = \begin{cases} S(j-j')\left[\frac{\prod_{s\in[d-1]\backslash j}(\widetilde{\mu}_j - \widetilde{\mu}'_s)\prod_{t\in[d]\backslash j'}(\widetilde{\mu}'_{j'} - \widetilde{\mu}_t + 1)}{\prod_{s\in[d]\backslash j}(\widetilde{\mu}'_j - \widetilde{\mu}'_s)\prod_{t\in[d-1]\backslash j'}(\widetilde{\mu}'_{j'} - \widetilde{\mu}'_t + 1)}\right]^{\frac{1}{2}} & \text{if } j' \in \{1,\ldots,d-1\}. \\ S(j-d)\left[\frac{\prod_{s\in[d-1]\backslash j}(\widetilde{\mu}_j - \widetilde{\mu}'_s)}{\prod_{s\in[d]\backslash j}(\widetilde{\mu}'_j - \widetilde{\mu}'_s)}\right]^{\frac{1}{2}} & \text{if } j' = 0. \end{cases} \tag{7.38}$$

The elements of the partitions here are of size $O(n)$, so the total computation necessary is poly$(d, \log n)$. Now how do we implement the $\hat{T}^{[d]}$ transform given this expression?

As in the introduction to this section, note that any unitary gate of dimension d can be implemented using a number of two qubit gates polynomial in d[RZBB94, Bar95, NC00]. The method of this construction is to take a unitary gate of dimension d with *known* matrix elements and then convert this into a series of unitary gates which act non-trivially only on two states. These two state gates can then be constructed using the methods described in [Bar95]. In order to modify this for our work, we calculate, to the specified accuracy ϵ, the elements of the $\hat{T}^{[d]}$ operator, conditional on the μ and $\mu' + e_{j'}$ inputs, perform the decomposition into two qubit gates as described in [RZBB94, Bar95] *online*, and then, conditional on this calculation perform the appropriate controlled two-qubit gates onto the space where $\hat{T}^{[d]}$ will act. Finally this classical computation must be undone to reset any garbage bits created during the classical computation. To produce an accuracy ϵ we need a classical computation of size poly$(\log(1/\epsilon))$ since we can perform the appropriate controlled rotations with bitwise accuracy.

Putting everything together as depicted in figures 7-3 and 7-4 gives a poly$(d, \log n, \log 1/\epsilon)$ algorithm to reduce $U_{\text{CG}}^{[d]}$ to $U_{\text{CG}}^{[d-1]}$. Naturally this can be applied d times to yield a poly$(d, \log n, \log 1/\epsilon)$ algorithm for $U_{\text{CG}}^{[d]}$. (We can end the recursion either at $d = 2$, using the construction in [BCH04], or at $d = 1$, where the CG transform simply consists of the map $\mu \to \mu + 1$ for $\mu \in \mathbb{Z}$, or even at $d = 0$, where the CG transform is completely trivial.) We summarize the CG algorithm as follows.

Algorithm: Clebsch-Gordan transform

Inputs: (1) Classical registers d and n. (2) Quantum registers $|\lambda\rangle$ (in any superposition over different $\lambda \in \mathcal{I}_{d,n}$), $|q\rangle \in \mathcal{Q}_{\lambda}^{d}$ (expressed as a superposition of GZ basis elements) and $|i\rangle \in \mathbb{C}^d$.

Outputs: (1) Quantum registers $|\lambda\rangle$ (equal to the input), $|j\rangle \in \mathbb{C}^d$ (satisfying $\lambda + e_j \in \mathcal{I}_{d,n+1}$) and $|q'\rangle \in \mathcal{Q}_{\lambda+e_j}^{d}$.

Runtime: d^3 poly$(\log n, \log 1/\epsilon)$ to achieve accuracy ϵ.

Procedure:

```
1.  If $d = 1$
2.  Then output $|j\rangle := |i\rangle = |1\rangle$ and $|q'\rangle := |q\rangle = |1\rangle$ (i.e. do nothing).
3.  Else
4.      Unpack $|q\rangle$ into $|\mu'\rangle|q_{(d-2)}\rangle$, such that $\mu' \in \mathcal{I}_{d,m}$, $m \leq n$, $\mu' \precsim \mu$ and $|q_{(d-2)}\rangle \in \mathcal{Q}_{\mu'}^{d-1}$.
5.      If $i < d$
6.      Then perform the CG transform with inputs $(d-1, m, |\mu'\rangle, |q_{(d-2)}\rangle, |i\rangle)$ and outputs
        $(|\mu'\rangle, |j'\rangle, |q'_{(d-2)}\rangle)$.
7.      Else (if $i = d$)
8.          Replace $|i\rangle = |d\rangle$ with $|j'\rangle := |0\rangle$ and set $|q'_{(d-2)}\rangle := |q'_{(d-2)}\rangle$.
9.      End. (Now $i \in \{1,\ldots,d\}$ has been replaced by $j \in \{0,\ldots,d-1\}$.)
10.     Map $|\mu'\rangle|j'\rangle$ to $|\mu' + e_{j'}\rangle|j'\rangle$.
11.     Conditioned on $\mu$ and $\mu' + e'_j$, calculate the gate sequence necessary to implement $\hat{T}^{[d]}$,
        which inputs $|j'\rangle$ and outputs $|j\rangle$.
12.     Execute this gate sequence, implementing $\hat{T}^{[d]}$.
13.     Undo the computation from 11.
```

```
14.    Combine $|\mu' + e_{j'}\rangle$ and $|q'_{(d-2)}\rangle$ to form $|q'\rangle$.
15. End.
```

Finally, in Section 7.2 we described how n CG transforms can be used to perform the Schur transform, so that U_{Sch} can be implemented in time $n \cdot \text{poly}(d, \log n, \log 1/\epsilon)$, optionally plus an additional $\text{poly}(n)$ time to compress the $|p\rangle$ register.

Chapter 8

Relations between the Schur transform and the $\mathcal{S}_n$ QFT

This final chapter is devoted to algorithmic connections between the Schur transform and the quantum Fourier transform on $\mathcal{S}_n$. In Section 8.1 we describe *generalized phase estimation*, which is a reduction from measuring in the Schur basis (a weaker problem than the full Schur transform) to the $\mathcal{S}_n$ QFT. Then in Section 8.2 we show a reduction in the other direction, from the $\mathcal{S}_n$ QFT to the Schur transform. The goal of these reductions is not so much to perform new tasks efficiently, since efficient implementations of the QFT already exist, but to help clarify the position of the Schur transform *vis-a-vis* known algorithms.

8.1 Generalized phase estimation

The last chapter developed the Schur transform based on the $\mathcal{U}_d$ CG transform. Can we instead build the Schur transform out of operations on $\mathcal{S}_n$? This section explores that possibility. We will see that using the $\mathcal{S}_n$ QFT allows us to efficiently measure a state in the Schur basis, a slightly weaker task than performing the full Schur transform. Our algorithm for this measurement generalizes the quantum circuits used to estimate the phase of a black-box unitary transform[Sho94, KSV02] (see also [KR03]) to a nonabelian setting; hence we call it generalized phase estimation (GPE). As we will see, our techniques actually extend to measuring the irrep labels in reducible representations of any group for which we can efficiently perform group operations and a quantum Fourier transform.

The main idea behind GPE is presented in Section 8.1.1, where we show how it can be used to measure $|\lambda\rangle$ (and optionally $|p\rangle$ as well) in the Schur basis. Here the techniques are completely general and we show how similar results hold for any group. We specialize to Schur basis measurements in Section 8.1.2, where we show that GPE can be extended to also measure the $\mathcal{Q}_\lambda^d$ register, thereby making a complete Schur basis measurement possible based only on the $\mathcal{S}_n$ QFT. We conclude in Section 8.1.3 with an alternate interpretation of GPE, which shows its close connection with the $\mathcal{S}_n$ CG transform.

8.1.1 Using GPE to measure λ and $\mathcal{P}_\lambda$

Let G be an arbitrary finite group over which there exists an efficient circuit for the quantum Fourier transform[MRR04], U_{QFT}. Fix a set of inequivalent irreps $\hat{G}$, where $\mu \in \hat{G}$ corresponds to the irrep $(\mathbf{r}_\mu, V_\mu)$. U_{QFT} then maps the group algebra $\mathbb{C}[G]$ to $\bigoplus_{\mu\in\hat{G}} V_\mu \otimes V_\mu^*$, and is explicitly given by Eq. (5.9).

Now suppose (ρ, V) is a representation of G for which we can efficiently perform the controlled-ρ operation, $C_\rho = \sum_{g\in G} |g\rangle\!\langle g| \otimes \rho(g)$. To specialize to the Schur transform we will choose $V = (\mathbb{C}^d)^{\otimes n}$

and $\rho = \mathbf{P}$, but everything in the section can be understood in terms of arbitrary G and (ρ, V). Let the multiplicity of the irrep V_ν in V be given by m_ν, so that V decomposes as

$$V \overset{G}{\cong} \bigoplus_{\nu \in \hat{G}} \mathbb{C}^{m_\nu} \otimes V_\nu. \tag{8.1}$$

This induces a basis for V, analogous to the Schur basis, given by $|\nu, \alpha, k\rangle_{\mathrm{V}}$, where $\nu \in \hat{G}$, $\alpha \in [m_\nu]$ and $k \in [d_\nu]$, where $d_\nu := \dim V_\nu$. For any $\lambda \in \hat{G}$, define the projector onto the V_λ-isotypic subspace in terms of this basis as

$$\Pi_\lambda = |\lambda\rangle\langle\lambda| \otimes I_{m_\lambda} \otimes I_{d_\lambda}. \tag{8.2}$$

Note that this becomes Eq. (6.18) for the special case of $(\rho, V) = (\mathbf{P}, (\mathbb{C}^d)^{\otimes n})$.

The problem is that, as with the Schur basis, there is no immediately obvious way to measure or otherwise access the register labeling the irreps. We are given no information about the isomorphism in Eq. (8.1) or about how to implement it. However, by using the Fourier transform along with the controlled-ρ operator, it is possible to efficiently perform the projective measurement $\{\Pi_\lambda\}_{\lambda \in \hat{G}}$. To do so, we define the operator

$$\hat{C}_\rho = (U_{\mathrm{QFT}} \otimes I_V) C_\rho (U_{\mathrm{QFT}}^\dagger \otimes I_V) \tag{8.3}$$

acting on $\bigoplus_\mu V_\mu \otimes V_\mu^* \otimes V$. This is represented in Fig. 8-1.

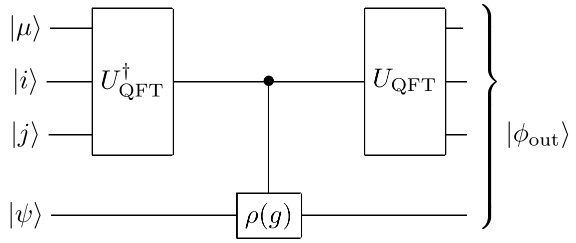


Figure 8-1: Quantum circuit $\hat{C}_\rho$ used in generalized phase estimation.

The procedure for performing the projective measurement $\{\Pi_\lambda\}_{\lambda \in \hat{G}}$ is as follows:

Algorithm: Generalized Phase Estimation

Inputs: A state $|\psi\rangle \in V$.

Outputs: (1) Classical variable λ with probability $p_\lambda := \langle\psi|\Pi_\lambda|\psi\rangle$
(2) The state $\Pi_\lambda|\psi\rangle/\sqrt{p_\lambda}$.

Runtime: $2T_{\mathrm{QFT}} + T_{C_\rho}$ where T_{QFT} (resp. T_{C_ρ}) is the running time for the QFT on G (resp. the controlled-ρ operation).

Procedure:

1. Create registers $|\mu\rangle|i\rangle|j\rangle$ (see Fig. 8-1) with μ corresponding to the trivial representation V_0 and $|i\rangle = |j\rangle = |1\rangle \in V_0$.

2. Apply $\hat{C}_\rho$. This involves three steps.
a) Apply $U_{\mathrm{QFT}}^\dagger$ to $|\mu\rangle|i\rangle|j\rangle$, obtaining the uniform superposition $|G|^{-1/2} \sum_{g \in G} |g\rangle$.
b) Perform $C_\rho = \sum_g |g\rangle\langle g| \otimes \rho(g)$.
c) Apply U_{QFT} to the first register.
The output $|\psi_{\mathrm{out}}\rangle$ is a superposition of $|\lambda\rangle|i'\rangle|j'\rangle|v\rangle$ with $\lambda \in \hat{G}$, $|i'\rangle \in V_\lambda$, $|j'\rangle \in V_\lambda^*$ and $|v\rangle \in V$.

3. Measure λ.

4. Optionally perform $\hat{C}_\rho^\dagger$. This is only necessary if we need the residual state $\Pi_\lambda|\psi\rangle$.

To analyze this circuit, expand $|\psi\rangle$ in the $|\mu\rangle|\alpha\rangle|k\rangle_{\mathrm{V}}$ basis as

$$|\psi\rangle = \sum_{\mu\in\hat{G}} \sum_{\alpha=1}^{m_\mu} \sum_{k=1}^{d_\mu} c_{\mu,\alpha,k} |\mu,\alpha,k\rangle_{\mathrm{V}}. \tag{8.4}$$

Eq. (8.1) means that $\rho(g)$ acts on $|\psi\rangle$ according to

$$\rho(g)|\psi\rangle = \sum_{\mu\in\hat{G}} \sum_{\alpha=1}^{m_\mu} \sum_{k=1}^{d_\mu} c_{\mu,\alpha,k} |\mu,\alpha\rangle \mathbf{r}_\mu(g)|k\rangle_{\mathrm{V}}. \tag{8.5}$$

Now examine the $\mathbb{C}[G]$ register. The initial $U_{\mathrm{QFT}}^\dagger$ in $\hat{C}_\rho$ maps the trivial irrep to the uniform superposition of group elements $\frac{1}{|G|}\sum_{g\in G}|g\rangle$. This is analogous to the initialization step of phase estimation on abelian groups[Sho94, KSV02]. Thus the output of the circuit in Fig. 8-1 is

$$|\phi_{\mathrm{out}}\rangle = \sum_{g\in G} \sum_{\nu\in\hat{G}} \sum_{\lambda\in\hat{G}} \sum_{i,j=1}^{d_\lambda} \frac{\sqrt{d_\lambda}}{|G|} \left[\, \mathbf{r}_\lambda(g) \,\right]_{i,j} |\lambda,i,j\rangle \otimes \rho(g)|\psi\rangle. \tag{8.6}$$

We can simplify this using Eq. (8.5) and the orthogonality relations for irrep matrix elements[GW98] to reexpress Eq. (8.6) as

$$|\phi_{\mathrm{out}}\rangle = \sum_{\lambda\in\hat{\mathcal{G}}} \sum_{\alpha=1}^{m_\lambda} \sum_{i,j=1}^{d_\alpha} \frac{c_{\lambda,\alpha,i}}{\sqrt{d_\lambda}} |\lambda,i,j\rangle \otimes |\lambda,\alpha,j\rangle_{\mathrm{V}}\,. \tag{8.7}$$

The output $|\phi_{\mathrm{out}}\rangle$ has several interesting properties which we can now exploit. Measuring the first register (the irrep label index) produces outcome λ with the correct probability $\sum_{j=1}^{m_\lambda}\sum_{k=1}^{d_\lambda}|c_{\lambda,j,k}|^2$. Remarkably, this is achieved independent of the basis in which $\mathbf{C}_\rho$ is implemented. As mentioned above, this reduces to measuring the irrep label λ in the Schur basis when $G=\mathcal{S}_n$ and $(\rho,V)=(\mathbf{P},(\mathbb{C}^d)^{\otimes n})$. In this case, the circuit requires running time poly(n) for the $\mathcal{S}_n$ QFT[Bea97] and $O(n\log d)$ time for the controlled permutation $C_{\mathbf{P}}$, comparable to the efficiency of the Schur transform given in the last chapter.

This circuit also allows us to perform arbitrary instruments on the irrep spaces V_λ; for example, we could perform a complete measurement, or could perform a unitary rotation conditioned on λ. This is because Eq. (8.7) has extracted the irrep basis vector from $|\psi\rangle$ into the $|i\rangle$ register. We can perform an arbitrary instrument on this V_λ register, and then return the information to the V register by performing $\hat{C}_\rho^\dagger$.

To put this more formally, suppose we want to perform an instrument with operation elements

$$\sum_{\lambda\in\hat{G}} |\lambda\rangle\langle\lambda| \otimes I_{m_\lambda} \otimes A_\lambda^{(x)} \tag{8.8}$$

on V, where x labels the outcomes of the instrument and the normalization condition is that $\sum_x (A_\lambda^{(x)})^\dagger A_\lambda^{(x)} = I_{d_\lambda}$ for each λ. Then this can be effected by performing the instrument

$$\hat{C}_\rho^\dagger \left(\sum_{\lambda\in\hat{G}} |\lambda\rangle\langle\lambda| \otimes A_\lambda^{(x)} \otimes I_{d_\lambda} \otimes I_V \right) \hat{C}_\rho. \tag{8.9}$$

This claim can be verified by explicit calculation and use of the orthogonality relations, but we will give a simpler proof in Section 8.1.3.

To recap, so far we have shown how GPE can be used to efficiently:

- measure the $|\lambda\rangle$ and $|p\rangle$ registers, or perform general instruments of the form of Eq. (8.8), in the Schur basis of $(\mathbb{C}^d)^{\otimes n}$ using $\text{poly}(n) + O(n \log d)$ gates; and

- perform instruments of the form of Eq. (8.8) for any group G and representation (ρ, V) such that the QFT on G and the controlled-ρ operation can be implemented efficiently.

8.1.2 Using GPE to measure $\mathcal{Q}^d_\lambda$

In this section we specialize to the case of the Schur basis and show how GPE can be adapted to measure the $|q\rangle$ register. This allows us to perform a complete measurement in the $|\lambda\rangle|p\rangle|q\rangle_{\text{Sch}}$ basis, or more generally, to perform instruments with operation elements

$$\sum_{\lambda \in \mathcal{I}_{d,n}} \sum_{q \in Q^d_\lambda} U_{\text{Sch}} \left(|\lambda\rangle\langle\lambda| \otimes |q\rangle\langle q| \otimes A^{(x)}_{\lambda,q} \right) U^\dagger_{\text{Sch}}. \tag{8.10}$$

Here Q^d_λ is the GZ basis defined in Section 7.1.2. We will find that the running time is $d\,\text{poly}(n, \log d, \log 1/\epsilon)$, which is comparable to the running time of the circuits in Section 7.2, but has slightly less dependence on d and slightly more dependence on n.* More importantly, it gives a conceptually independent method for a Schur basis measurement.

The main idea is that we can measure $|q\rangle \in Q^d_\lambda$ by measuring the irrep label q_c for each subgroup $\mathcal{U}_c \subset \mathcal{U}_d$, $c = 1, \ldots, d-1$. We can measure q_c by performing GPE in a way that only looks at registers in states $|1\rangle, \ldots, |c\rangle$. As these measurements commute[Bie63]—in fact, they are simultaneously diagonalized by the GZ basis[GZ50]—we can perform them sequentially without worrying about the disturbance that they cause. After performing this modified GPE $d-1$ times, we can extract the register $|q\rangle$ in addition to the $|\lambda\rangle|p\rangle$ that we get from the first application of GPE.

We now describe this modification of GPE in more detail. To do so, we will need to consider performing GPE on a variable number of qubits. Define $U^{(d,n)}_{\text{GPE}}$ by

$$U^{(d,n)}_{\text{GPE}} = \sum_{\lambda \in \mathcal{I}_{d,n}} |\lambda\rangle \otimes (U^{(d,n)}_{\text{Sch}})^\dagger \left(|\lambda\rangle\langle\lambda| \otimes I_{\mathcal{Q}^d_\lambda} \otimes I_{\mathcal{P}_\lambda} \right) U^{(d,n)}_{\text{Sch}} = \sum_{\lambda \in \mathcal{I}_{d,n}} |\lambda\rangle \otimes \Pi^{(d,n)}_\lambda \tag{8.11}$$

This coherently extracts the $|\lambda\rangle$ register from $(\mathbb{C}^d)^{\otimes n}$. Here we have also explicitly written out the dependence of $\Pi^{(d,n)}_\lambda$ and $U^{(d,n)}_{\text{Sch}}$ on d and n. Also, we have expressed $U^{(d,n)}_{\text{GPE}}$ as an isometry to avoid writing out the ancilla qubits initialized to zero, but it is of course a reversible unitary transform. For example, we use GPE to construct $\Pi^{(d,n)}_\lambda$ by performing $U^{(d,n)}_{\text{GPE}}$, measuring λ and then undoing $U^{(d,n)}_{\text{GPE}}$:

$$\Pi^{(d,n)}_\lambda = (U^{(d,n)}_{\text{GPE}})^\dagger \left(|\lambda\rangle\langle\lambda| \otimes I^{\otimes n}_d \right) U^{(d,n)}_{\text{GPE}}. \tag{8.12}$$

The observables we want to measure correspond to determining the irrep label of $\mathcal{U}_c$. For any $\mathcal{U}_c$-representation $(\mathbf{q}, \mathcal{Q})$, we define the $\mathcal{Q}^c_\mu$-isotypic subspace of $\mathcal{Q}$ to be the direct sum of the irreps in the decomposition of $\mathcal{Q}$ that are isomorphic to $\mathcal{Q}^c_\mu$ (cf. Eq. (5.2)). The projector onto this subspace

*If d is much larger than n, then it is always possible (even with the CG-based Schur transform) to reduce the time for a Schur basis measurement to $\text{poly}(n, \log 1/\epsilon) + O(n \log d)$. This is because, given a string $|i_1, \ldots, i_n\rangle \in (\mathbb{C}^d)^{\otimes n}$, we can first measure the type in time $O(n \log d)$ and then unitarily map $|i_i, \ldots, i_n\rangle$ to $|i'_i, \ldots, i'_n\rangle$, where $i'_j \in [n]$ and $i'_j = i'_k$ iff $i_j = i_k$. Measuring $|i'_i, \ldots, i'_n\rangle$ in the Schur basis then requires $\text{poly}(n, \log 1/\epsilon)$ time, and the measured value of $|q\rangle$ can be translated to the proper value by replacing each instance i'_j in the Young tableau with i_j. Moreover, the final answer can be used to uncompute the type, so this modification also works when implementing U_{Sch} rather than simply a Schur basis measurement.

can be given explicitly (though we will not need the exact formula) in terms of $\mathbf{q}$ as follows:

$$\pi^c_\mu(\mathbf{q}) = \dim \mathcal{Q}^c_\mu \int_{U\in\mathcal{U}_c} dU \left(\mathrm{Tr}\,\mathbf{q}^c_\mu(U)\right)^* \mathbf{q}(U), \tag{8.13}$$

where dU is a Haar measure on $\mathcal{U}_c$. Define the $\mathcal{U}_c$-representation $(\mathbf{Q}^n_d, (\mathbb{C}^d)^{\otimes n})$ by $\mathbf{Q}^n_d(U) = (U \oplus I_{d-c})^{\otimes n}$, where $(U \oplus I_{d-c})$ is the embedding of $\mathcal{U}_c$ in $\mathcal{U}_d$ given by Eq. (7.3). Our goal is to perform the projective measurements $\{\pi^c_\mu(\mathbf{Q}^n_d)\}_{\mu\in\mathbb{Z}^c_{++},|\mu|\leq n}$ for $c = 1, \ldots, d$. Since for each c, $\pi^c_\mu(\mathbf{Q}^n_d)$ is diagonal in the GZ basis Q^d_λ, the projectors commute and can be measured simultaneously.

For the special case of $|\lambda| = m = n$ and $c = d$, $\mathbf{Q}^n_d$ is the same as the $\mathbf{Q}$ defined in Eq. (5.11) and we have $\Pi^{(d,n)}_\lambda = \pi^d_\lambda(\mathbf{Q}^n_d)$. In this case, Eq. (8.12) tells us how to perform the projective measurement $\{\pi^d_\lambda(\mathbf{Q}^n_d)\}_{\lambda\in\mathcal{I}_{d,n}}$. We now need to extend this to measure $\{\pi^c_\mu(\mathbf{Q}^n_d)\}_{\mu\in\mathcal{I}_{c,m},m\leq n}$ for any $c \in [d]$.

The first step in doing so is to measure the number of positions in $|i_1, \ldots, i_n\rangle$ where $i_j \in \{1, \ldots, c\}$. Call this number m. Though we will measure m, we will not identify which j have $i_j \in [c]$. Instead we will coherently separate them by performing the unitary operation $U^{(c)}_{\mathrm{sel}}$ which implements the isomorphism:

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{m=0}^{n} (\mathbb{C}^c)^{\otimes m} \otimes (\mathbb{C}^{d-c})^{\otimes n-m} \otimes \mathbb{C}^{\binom{n}{m}} \tag{8.14}$$

It is straightforward to implement $U^{(c)}_{\mathrm{sel}}$ in time linear in the size of the input (i.e. $O(n \log d)$), though we will have to pad quantum registers as in the original Schur transform.

We can use $U^{(c)}_{\mathrm{sel}}$ to construct $\pi^c_\mu(\mathbf{Q}^n_d)$ in terms of $\pi^c_\mu(\mathbf{Q}^m_c)$ as follows:

$$\pi^c_\mu(\mathbf{Q}^n_d) = (U^{(c)}_{\mathrm{sel}})^\dagger \left(|m\rangle\langle m| \otimes \pi^c_\mu(\mathbf{Q}^m_c) \otimes I^{\otimes n-m}_{d-c} \otimes I_{\binom{n}{m}}\right) U^{(c)}_{\mathrm{sel}}. \tag{8.15}$$

If $m = |\mu|$, then we can use Eq. (8.12) to construct the projection on the RHS of Eq. (8.15), obtaining

$$\pi^c_\mu(\mathbf{Q}^n_d) = (U^{(c)}_{\mathrm{sel}})^\dagger \left(|m\rangle\langle m| \otimes \left[\left(U^{(c,m)}_{\mathrm{GPE}}\right)^\dagger (|\mu\rangle\langle\mu| \otimes I^{\otimes m}_c)\, U^{(c,m)}_{\mathrm{GPE}}\right] \otimes I^{\otimes n-m}_{d-c} \otimes I_{\binom{n}{m}}\right) U^{(c)}_{\mathrm{sel}}. \tag{8.16}$$

This gives a prescription for measuring $\pi^c_\mu(\mathbf{Q}^n_d)$, whose output μ corresponds to the component q_c of the GZ basis element. First measure $m = |\mu|$ by counting the number of qudits that have values in $\{1, \ldots, c\}$. Then select only those m qudits using $U^{(c)}_{\mathrm{sel}}$ and perform GPE on them to find μ.

Finally, measuring the commuting observables $\{\pi^c_\mu(\mathbf{Q}^n_d)\}_{\mu\in\mathbb{Z}^c_{++}}$ for $c = 1, \ldots, d$ yields a complete von Neumann measurement of the $\mathcal{Q}^d_\lambda$ register with operation elements as follows:

$$(U^{(d,n)}_{\mathrm{Sch}})^\dagger\, (|\lambda\rangle\langle\lambda| \otimes |q\rangle\langle q| \otimes I_{\mathcal{P}_\lambda})\, U^{(d,n)}_{\mathrm{Sch}} = \prod_{c=1}^{d} \pi^c_{q_c}(\mathbf{Q}^n_d), \tag{8.17}$$

with q ranging over Q^d_λ.

Combined with the results of the last section, we now have an algorithm for performing a complete measurement in the Schur basis, or more generally an arbitrary instrument with operation elements

$$(U^{(d,n)}_{\mathrm{Sch}})^\dagger \left(\sum_{\lambda\in\mathcal{I}_{d,n}} |\lambda\rangle\langle\lambda| \otimes |q\rangle\langle q| \otimes A^{(x)}_{\lambda,q}\right) U^{(d,n)}_{\mathrm{Sch}}, \tag{8.18}$$

where the $A^{(x)}_{\lambda,q}$ are arbitrary operators on $\mathcal{P}_\lambda$ and x labels the measurement outcomes. In particular, if we let x range over triples (λ, q, p) with $\lambda \in \mathcal{I}_{d,n}$, $q \in Q^d_\lambda$ and $p \in P_\lambda$; and set $A^{(\lambda,q,p)}_{\lambda',q'} = \delta_{\lambda,\lambda'}\delta_{q,q'}\, |p\rangle\langle p|$, then Eq. (8.18) corresponds to a complete von Neumann measurement in the Schur basis. The general algorithm is as follows:

```
Algorithm: Complete Schur basis measurement using GPE
  Input: A state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$.
  Output: $\sum_{\lambda\in\mathcal{I}_{d,n}} \sum_{q\in Q_\lambda^d} \sum_x |\lambda\rangle|q\rangle|x\rangle(U_{\rm Sch}^{(d,n)})^\dagger \left(\sum_{\lambda\in\mathcal{I}_{d,n}} |\lambda\rangle\langle\lambda| \otimes |q\rangle\langle q| \otimes A_{\lambda,q}^{(x)}\right) U_{\rm Sch}^{(d,n)}|\psi\rangle$
          corresponding to the coherent output of the instrument in Eq. (8.18)
  Runtime: $d \cdot O(T_{\rm QFT}(\mathcal{S}_n) + n\log d)$
  Procedure:
  1. For $c = 1, \ldots, d-1$:
  2.     Apply $U_{\rm sel}^{(c)}$ to $|\psi\rangle$, outputting superpositions of $|m\rangle|\alpha_c^m\rangle|\beta_{d-c}^{n-m}\rangle|\gamma\rangle_{\binom{n}{m}}$.
  3.     Perform $\sum_{m=0}^n |m\rangle\langle m| \otimes U_{\rm GPE}^{(c,m)}$ on $|m\rangle|\alpha_c^m\rangle$ to output $|m\rangle \sum_{\mu_c\in\mathcal{I}_{c,m}} |\mu_c\rangle\pi_{\mu_c}^c(\mathbf{Q}_d^n)|\alpha_c^m\rangle$.
  4.     Apply $(U_{\rm sel}^{(c)})^\dagger$.
  5. Set $|q\rangle := |\mu_1\rangle \ldots |\mu_{d-1}\rangle$.
  (Steps 6–10 are based on Eq. (8.9).)
  6. Add registers $|\mu\rangle = |(n)\rangle$ and $|i\rangle = |j\rangle = \left|\boxed{1}\boxed{2}\cdots\boxed{n}\right\rangle$ corresponding to the trivial irrep of $\mathcal{S}_n$.
  7. Perform $\hat{C}_{\mathbf{P}}$ $(:= (U_{\rm QFT}\otimes I)C_{\mathbf{P}}(U_{\rm QFT}^\dagger \otimes I))$ on $|\mu\rangle|i\rangle|j\rangle|\psi\rangle$ to output $|\lambda\rangle|i'\rangle|j'\rangle|\psi'\rangle$.
  8. Perform the instrument $\left\{\sum_{\lambda\in\mathcal{I}_{d,n}} \sum_{q\in Q_\lambda^d} |q\rangle\langle q| \otimes |\lambda\rangle\langle\lambda| \otimes A_{\lambda,q}^{(x)} \otimes I_{\mathcal{P}_\lambda} \otimes I_d^{\otimes n}\right\}_x$.
  9. Apply $\hat{C}_{\mathbf{P}}^\dagger$ to output $|\mu\rangle|i\rangle|j\rangle|\psi''\rangle$.
  10. The registers $|\mu\rangle|i\rangle|j\rangle$ are always in the state $|(n)\rangle\left|\boxed{1}\boxed{2}\cdots\boxed{n}\right\rangle^{\otimes 2}$ and can be discarded.
  11. Reverse steps 1–5.
```

Generalizations to other groups: Crucial to this procedure is not only that $\mathcal{U}_d$ and $\mathcal{S}_n$ form a dual reductive pair (cf. Section 5.4), but that both groups have GZ bases, and their canonical towers of subgroups also form dual reductive pairs. These conditions certainly exist for other groups (e.g. $\mathcal{U}_{d_1} \times \mathcal{U}_{d_2}$ acting on polynomials of $\mathbb{C}^{d_1+d_2}$), but it is an open problem to find useful applications of the resulting algorithms.

8.1.3 Connection to the Clebsch-Gordan transform

In this section, we explain how GPE can be thought of in terms of the CG transform on G, or on $\mathcal{S}_n$ when we specialize to the case of the Schur basis. The goal is to give a simple representation-theoretic interpretation of the measurements described in Section 8.1.1 as well as pointing out relations between the QFT and the CG transform.

We begin with a quick review of GPE. We have two registers, $\mathbb{C}[G]$ and V, where (ρ, V) is a representation of G. Assume for simplicity that the isomorphism in Eq. (8.1) is an equality:

$$V = \bigoplus_{\nu\in\hat{G}} V_\nu \otimes \mathbb{C}^{m_\nu}. \tag{8.19}$$

This means that the controlled-ρ operation C_ρ is given by

$$C_\rho = \sum_{g\in G} |g\rangle\langle g| \otimes \rho(g) = \sum_{g\in G} |g\rangle\langle g| \otimes \sum_{\nu\in\hat{G}} |\nu\rangle\langle\nu| \otimes \mathbf{r}_\nu(g) \otimes I_{m_\nu}. \tag{8.20}$$

The GPE prescription given in Section 8.1.1 began with initializing $\mathbb{C}[G]$ with the trivial irrep (or equivalently a uniform superposition over group elements). We will relax this condition and analyze the effects of $\hat{C}_\rho$ (the Fourier-transformed version of C_ρ, cf. Eq. (8.3)) on arbitrary initial states in $\mathbb{C}[G]$ in order to see how it acts like the CG transform.

There are two equivalent ways of understanding how $\hat{C}_\rho$ acts like the CG transform; in terms of representation spaces or in terms of representation matrices. We first present the explanation based on representation matrices. Recall from Section 5.2.3 that $\mathbb{C}[G]$ can be acted on by either the left or the right representation ($\mathbf{L}(h)|g\rangle = |hg\rangle$ and $\mathbf{R}(h)|g\rangle = |gh^{-1}\rangle$). The controlled-$\rho$ operation acts on these matrices as follows

$$C_\rho\left(\mathbf{L}(g)\otimes I_V\right)C_\rho^\dagger = \mathbf{L}(g)\otimes\rho(g) \tag{8.21}$$

$$C_\rho^\dagger\left(\mathbf{R}(g)\otimes I_V\right)C_\rho = \mathbf{R}(g)\otimes\rho(g) \tag{8.22}$$

The proofs of these claims are straightforward[*]. If we combine them, we find that

$$C_\rho(\mathbf{L}(g_1)\mathbf{R}(g_2)\otimes\rho(g_2))C_\rho^\dagger = \mathbf{L}(g_1)\mathbf{R}(g_2)\otimes\rho(g_1). \tag{8.25}$$

Let us examine how C_ρ transforms the left and right representations, any observable on $\mathbb{C}[G]$ can be constructed out of them. Focus for now on the action of C_ρ on the left representation. Eq. (8.21) says that conjugation by C_ρ maps the left action on $\mathbb{C}[G]$ to the tensor product action on $\mathbb{C}[G]\otimes V$. To see how this acts on the representation spaces, we conjugate each operator by U_{QFT}, replacing C_ρ with $\hat{C}_\rho$, $\mathbf{L}$ with $\hat{\mathbf{L}}$ and $\mathbf{R}$ with $\hat{\mathbf{R}}$. Then $\hat{C}_\rho$ couples an irrep V_μ from $\mathbb{C}[G]\cong\bigoplus_\mu V_\mu\otimes V_\mu^*$ with an irrep V_ν from V and turns this into a sum of irreps V_λ. This explains how GPE can decompose V into irreps: if initialize μ to be the trivial irrep, then only $\lambda=\nu$ appears in the output and measuring λ has the effect of measuring the irrep label of V. The right representation is acted on by $\hat{C}_\rho$ in the opposite manner; we will see that conjugating by $\hat{C}_\rho$ corresponds to the inverse CG transform, mapping V_μ^* in the input to $V_\lambda^*\otimes V_\nu$ in the output.

To make this concrete, Fourier transform each term in Eq. (8.25) to obtain

$$\hat{C}_\rho^\dagger\left(\sum_{\mu\in\hat{G}}|\mu\rangle\langle\mu|\otimes\mathbf{r}_\mu(g_1)\otimes\mathbf{r}_\mu(g_2^{-1})\otimes\sum_{\nu\in\hat{G}}|\nu\rangle\langle\nu|\otimes\mathbf{r}_\nu(g_1)\otimes I_{m_\nu}\right)\hat{C}_\rho \tag{8.26}$$

$$=\left(U_{\text{QFT}}\otimes I_V\right)C_\rho^\dagger\left(\mathbf{L}(g_1)\mathbf{R}(g_2)\otimes\rho(g_1)\right)C_\rho\left(U_{\text{QFT}}^\dagger\otimes I_V\right) \tag{8.27}$$

$$=\left(U_{\text{QFT}}\otimes I_V\right)\left(\mathbf{L}(g_1)\mathbf{R}(g_2)\otimes\rho(g_2)\right)\left(U_{\text{QFT}}^\dagger\otimes I_V\right) \tag{8.28}$$

$$=\sum_{\lambda\in\hat{G}}|\lambda\rangle\langle\lambda|\otimes\mathbf{r}_\lambda(g_1)\otimes\mathbf{r}_\lambda(g_2^{-1})\otimes\sum_{\nu\in\hat{G}}|\nu\rangle\langle\nu|\otimes\mathbf{r}_\nu(g_2)\otimes I_{m_\nu} \tag{8.29}$$

To understand this we need to work backwards. Measuring an observable on the V_λ register of the final state corresponds to measuring that observable on the V_λ-isotypic subspace of the original $V_\mu\otimes V_\nu$ inputs. On the other hand, the initial V_μ^* register splits into V_λ^* and V_ν registers. We can see an example of this in Eq. (8.7), where the V_μ register ($|i\rangle$) has been transferred to V_λ, while V_λ^* and

[*] Here we prove Eqns. (8.21) and (8.22):

$$C_\rho\left(\mathbf{L}(h)\otimes I_V\right)C_\rho^\dagger = \left(\sum_{g_1\in G}|hg_1\rangle\langle hg_1|\otimes\rho(hg_1)\right)\left(\sum_{g_2\in G}|hg_2\rangle\langle g_2|\otimes I_V\right)\left(\sum_{g_3\in G}|g_3\rangle\langle g_3|\otimes\rho(g_3^{-1})\right)$$

$$=\sum_{g\in G}|hg\rangle\langle g|\otimes\rho(h)=\mathbf{L}(h)\otimes\rho(h) \tag{8.23}$$

$$C_\rho^\dagger\left(\mathbf{R}(h)\otimes I_V\right)C_\rho = \left(\sum_{g_1\in G}|g_1h^{-1}\rangle\langle g_1h^{-1}|\otimes\rho(hg_1^{-1})\right)\left(\sum_{g_2\in G}|g_2h^{-1}\rangle\langle g_2|\otimes I_V\right)\left(\sum_{g_3\in G}|g_3\rangle\langle g_3|\otimes\rho(g_3)\right)$$

$$=\sum_{g\in G}|gh^{-1}\rangle\langle g|\otimes\rho(h)=\mathbf{R}(h)\otimes\rho(h) \tag{8.24}$$

V_ν are in the maximally entangled state $|\Phi_\lambda\rangle = d_\lambda^{-1/2}\sum_{j=1}^{d_\lambda}|jj\rangle$ corresponding to the trivial irrep that V_μ^* was initialized to.

Thus $\hat{C}_\rho$ corresponds to a CG transform from $V_\mu \otimes V_\nu$ to V_λ and an inverse CG transform from V_μ^* to $V_\lambda^* \otimes V_\nu$. These maps are sketched in Fig. 8-2.

$$\begin{tikzcd}
 & V_\nu \\
V_\mu^* \arrow[ur] \arrow[dr] & \\
V_\mu \arrow[dr] & V_\lambda^* \\
 & V_\lambda \\
V_\nu \arrow[ur] &
\end{tikzcd}$$

Figure 8-2: Performing $\hat{C}_\rho$ combines V_μ and V_ν to form V_λ and splits V_μ^* into V_λ^* and V_ν. Here $V_\mu \otimes V_\mu^*$ and $V_\lambda \otimes V_\lambda^*$ come from the decomposition of $\mathbb{C}[G]$ and V_ν comes from the decomposition of V.

We can verify the two maps separately by replacing $\hat{C}_\rho$ in Eq. (8.26) with U_{CG} or $U_{\text{CG}}^\dagger$ acting on the appropriate registers and checking that the representation matrices transform appropriately. There are two details here which still need to be explained. First, our description of the CG transform has not accounted for the multiplicity spaces that are generated. Second, we have not explained how the inverse CG transform always creates the correct irreps $V_\lambda^* \otimes V_\nu$ when λ is a label output by the first CG transform. To explain both of these, we track the representation spaces through a series of transformations equivalent to C_ρ. As with C_ρ, we will begin and end with $\mathbb{C}[G] \otimes V$, but we will show how the component irreps transform along the way. Here, $\overset{U}{\cong}$ is used to mean that the unitary operation U implements the isomorphism; all of the isomorphisms respect the action of the group G.

$$\begin{aligned}
\mathbb{C}[G]\otimes V \quad &\overset{U_{\text{QFT}}}{\cong} \left(\bigoplus_{\mu\in\hat{G}} G_\mu \otimes G_\mu^*\right) \otimes \left(\bigoplus_{\nu\in\hat{G}} G_\nu \otimes \mathbb{C}^{m_\nu}\right) &\qquad (8.30)\\
&\overset{U_{\text{CG}}}{\cong} \bigoplus_{\mu,\nu,\lambda\in\hat{G}} G_\lambda \otimes \operatorname{Hom}(G_\lambda, G_\mu \otimes G_\nu)^G \otimes G_\mu^* \otimes \mathbb{C}^{m_\nu} &\qquad (8.31)\\
&\cong \bigoplus_{\mu,\nu,\lambda\in\hat{G}} G_\lambda \otimes \operatorname{Hom}(G_\mu^*, G_\lambda^* \otimes G_\nu)^G \otimes G_\mu^* \otimes \mathbb{C}^{m_\nu} &\qquad (8.32)\\
&\overset{U_{\text{CG}}^\dagger}{\cong} \bigoplus_{\nu,\lambda\in\hat{G}} G_\lambda \otimes G_\lambda^* \otimes G_\nu \otimes \mathbb{C}^{m_\nu} &\qquad (8.33)\\
&\overset{U_{\text{QFT}}^\dagger}{\cong} \mathbb{C}[G] \otimes \bigoplus_{\nu\in\hat{G}} G_\nu \otimes \mathbb{C}^{m_\nu} = \mathbb{C}[G]\otimes V &\qquad (8.34)
\end{aligned}$$

The isomorphism in Eq. (8.32) is based on repeated application of the identity $\operatorname{Hom}(A,B) \cong A^* \otimes B$. This equivalence between $\operatorname{Hom}(G_\lambda, G_\mu \otimes G_\nu)^G$ and $\operatorname{Hom}(G_\mu^*, G_\lambda^* \otimes G_\nu)^G$ is the reason that a CG transform followed by an inverse CG transform on different registers can yield the correct representations in the output.

Application: using $U_{\mathbf{QFT}}$ to construct $U_{\mathbf{CG}}$

So far the discussion in this section has been rather abstract: we have shown that $\hat{C}_\rho$ acts in a way analogous to $U_{\rm CG}$, but have not given any precise statement of a connection. To give the ideas in this section operational meaning, we now show how $U_{\rm QFT}$ can be used to perform $U_{\rm CG}$ on an arbitrary group. This idea is probably widely known, and has been used for the dihedral group in [Kup03], but a presentation of this form has not appeared before in the literature.

The algorithm for $U_{\rm CG}$ is depicted in Fig. 8-3 and is described as follows:

Algorithm: Clebsch-Gordan transform using GPE

Input: $|\mu\rangle^{A_1}|v_\mu\rangle^{A_2}|\nu\rangle^{B_1}|v_\nu\rangle^{B_2}$, where $\mu,\nu \in \hat{G}$, $|v_\mu\rangle \in V_\mu$ and $|v_\nu\rangle \in V_\nu$.

Output: $|\lambda\rangle^{A_1}|v_\lambda\rangle^{A_2}|\nu\rangle^{B_1}|\alpha\rangle^{C}$ with $\lambda \in \hat{G}$, $|v_\lambda\rangle \in V_\lambda$ the irrep of the combined space and $|\alpha\rangle \in (V_\mu \otimes V_\nu \otimes V_\lambda^*)^G$ the multiplicity label.

Runtime: $4T_{\rm QFT} + T_{C_{\mathbf{L}}}$ where $T_{C_{\mathbf{L}}}$ is the time of the controlled-**L** operation.

Procedure:

1. Add states $|\Phi_\mu\rangle^{A_3A_4}$ and $|v_\nu^*\rangle^{B_3}$, where $|\Phi_\mu\rangle$ is the unique state (up to phase) in the one-dimensional space $(V_\mu^* \otimes V_\mu)^G$ (cf. Eq. (6.33)) and $|v_\nu^*\rangle \in V_\nu^*$ is arbitrary.

2. Perform the inverse QFT on $A_1A_2A_3$ (yielding output A) and on $B_1B_2B_3$ (yielding output B); i.e.

$$U_{\rm QFT}^{A_1A_2A_3\to A} \otimes U_{\rm QFT}^{B_1B_2B_3\to B}.$$

Registers A and B now contain states in $\mathbb{C}[G]$.

3. Apply $C_{\mathbf{L}}^{AB}$, mapping $|g_1\rangle^A|g_2\rangle^B$ to $|g_1\rangle^A|g_1g_2\rangle^B$.

4. Perform the QFT on A and B, yielding output $A_1A_2A_3$ and $B_1B_2B_3$.

5. Discard the register B_3, which still contains the state $|v_\nu^*\rangle$.

6. A_1 now contains the combined irrep label, which we call λ. The irrep space V_λ is in A_2, while the multiplicity space $(V_\mu \otimes V_\nu \otimes V_\lambda^*)^G$ is in $A_4B_2A_3$, which we relabel as C.

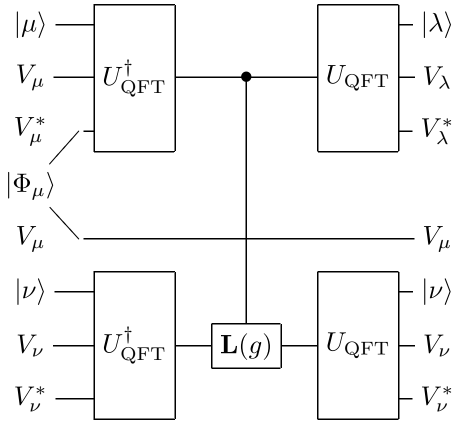


Figure 8-3: Using $U_{\rm QFT}$ to construct $U_{\rm CG}$ for an arbitrary group. The inputs to the CG transform are put in the $|\mu\rangle$, V_μ, $|\nu\rangle$ and V_ν registers. The V_ν^* register is not affected by the circuit, but is included so that the QFT will have valid inputs. The output of the CG transform is in the $|\lambda\rangle$ and V_λ registers. $|\nu\rangle$ saves the irrep label of one of the inputs and $(V_\mu \otimes V_\nu \otimes V_\lambda^*)^G$ is the multiplicity space.

The representations are transformed in the same way as in Fig. 8-2 with the addition of V_ν^* which is left unchanged. This is because we only act on the second register using left multiplication, which acts as the identity on V_ν^*.

Thus, we can perform the CG transform efficiently whenever we can efficiently perform the QFT, with the small caveat that efficiently manipulating the multiplicity space may take additional effort. One application of this reduction is to choose $G = \mathcal{U}_d$ and thereby replace the $\mathcal{U}_d$ CG construction of Section 7.3 with a CG transform based on the $\mathcal{U}_d$ QFT. Unfortunately, no fast quantum algorithms are known for the $\mathcal{U}_d$ and is is not immediately clear how quantizing $\mathbb{C}[\mathcal{U}_d]$ should correspond to cutting off the irreps that appear in the decomposition $\bigoplus_\lambda \mathcal{Q}_\lambda^d \otimes (\mathcal{Q}_\lambda^d)^*$. However, QFTs are known for discrete matrix groups such as $GL_n(\mathbb{F}_q)$ (running in time $q^{O(n)}$)[MRR04] and classical fast Fourier transforms are known for $\mathcal{U}_d$ and other compact groups[MR97].

A problem related to the $\mathcal{U}_d$ QFT was also addressed in [Zal04], which sketched an algorithm for implementing $\mathbf{q}_\lambda^2(U)$ in time polylogarithmic in $|\lambda|$, though some crucial details about efficiently integrating Legendre functions remain to be established. In contrast, using the Schur transform to construct $\mathbf{q}_\lambda^2(U)$ would require poly($|\lambda|$) gates. The idea behind [Zal04] is to embed $\mathcal{Q}_\lambda^2$ in $\mathcal{Q}_*^d := \bigoplus_\mu \mathcal{Q}_\mu^2$, where $\mu \in \mathbb{Z}_{++}^2$ and $|\mu|$ can be exponentially large. Then $\mathcal{Q}_*^d$ corresponds to functions on the 2-sphere which can be discretized and efficiently rotated, though some creative techniques are necessary to perform this unitarily. It is possible that this approach could ultimately yield efficient implementations of U_{CG} and U_{QFT} on $\mathcal{U}_d$.

8.2 Deriving the $\mathcal{S}_n$ QFT from the Schur transform

We conclude the chapter by showing how U_{Sch} can be used to construct U_{QFT}. Of course, an efficient algorithm for U_{QFT} already exists[Bea97], but the circuit we present here appears to be quite different. Some of the mathematical principles behind this connection are in Thm. 9.2.8 of [GW98] and I am grateful to Nolan Wallach for a very helpful conversaton on this subject.

The algorithm is based on the embedding of $\mathcal{S}_n$ in $[n]^n$ given by $s \rightarrow (s(1), \ldots, s(n))$. This induces a map from $\mathbb{C}[\mathcal{S}_n] \rightarrow (\mathbb{C}^n)^{\otimes n}$. More precisely, if 1^n denotes the weight $(1, \ldots, 1)$ with n ones, then we have a unitary map between $\mathbb{C}[\mathcal{S}_n]$ and $(\mathbb{C}^n)^{\otimes n}(1^n)$. This is the natural way we would represent a permutation on a computer (quantum or classical): as a string of n distinct numbers from $\{1, \ldots, n\}$. Similarly, we can embed $\mathcal{S}_n$ in $\mathcal{U}_n$ by letting a permutation s denote the unitary matrix $\sum_{i=1}^n |s(i)\rangle\langle i|$.

Using this embedding, the algorithm for U_{QFT} is as follows:

Algorithm: $\mathcal{S}_n$ QFT using the Schur transform

```
  Input: $\mathbb{C}[\mathcal{S}_n]$
  Output: $\bigoplus_{\lambda \in \mathcal{I}_n} \mathcal{P}_\lambda^* \otimes \mathcal{P}_\lambda$.
  Runtime: poly$(n, \log 1/\epsilon)$.
  Procedure:
  1. Embed $\mathbb{C}[\mathcal{S}_n]$ in $(\mathbb{C}^n)^{\otimes n}(1^n)$.
  2. Perform $U_{\mathrm{Sch}}^{(n,n)}$ on $(\mathbb{C}^n)^{\otimes n}(1^n)$ to output $|\lambda\rangle|q\rangle|p\rangle$.
  3. Output $|\lambda\rangle$ as the irrep label, $|q\rangle$ as the state of $\mathcal{P}_\lambda^*$ and $|p\rangle$ for $\mathcal{P}_\lambda$.
```

First we need to argue that setting $|q\rangle$ to be the $\mathcal{P}_\lambda^*$ output is well-defined. Note that $|q\rangle \in \mathcal{Q}_\lambda^n(1^n)$, so if $|q\rangle$ is a GZ basis vector, then its branching pattern $(q_1, \ldots, q_n)$ satisfies $q_i \in q_{i+1} - \Box$, and thus $|q_1, \ldots, q_n\rangle \in P_\lambda$.

Now to prove that this algorithm indeed performs a Fourier transform on $\mathcal{S}_n$, we examine a series of isomorphisms. The Fourier transform relates $\mathbb{C}[\mathcal{S}_n]$ to $\bigoplus_\lambda \mathcal{P}_\lambda \otimes \mathcal{P}_\lambda$. Since weights are determined by the action of the unitary group, restricting Eq. (5.16) on both sides to the 1^n weight space gives the relation

$$(\mathbb{C}^n)^{\otimes n}(1^n) \overset{\mathcal{U}_d \times \mathcal{S}_n}{\cong} \bigoplus_{\lambda \in \mathcal{I}_n} \mathcal{Q}_\lambda^n(1^n) \hat{\otimes} \mathcal{P}_\lambda \tag{8.35}$$

Thus we have the isomorphisms:

$$\begin{tikzcd}
\mathbb{C}[\mathcal{S}_n] \arrow[r, "\text{embed}", "(1)"'] \arrow[d, "U_{\text{QFT}}", "(2)"'] & (\mathbb{C}^n)^{\otimes n}(1^n) \arrow[d, "U_{\text{Sch}}", "(3)"'] \\
\bigoplus_{\lambda\in\mathcal{I}_n} \mathcal{P}_\lambda \otimes \mathcal{P}_\lambda \arrow[r, "(4)"] & \bigoplus_{\lambda\in\mathcal{I}_n} \mathcal{Q}^n_\lambda(1^n) \otimes \mathcal{P}_\lambda
\end{tikzcd}$$

Our goal is to understand the isomorphism (4) by examining how the other isomorphisms act on representation matrices. First we look at how (1) relates $\mathbf{P}, \mathbf{Q}$ with $\mathbf{L}, \mathbf{R}$. Note that $\mathbf{Q}(\mathcal{S}_n)$ and $\mathbf{P}(\mathcal{S}_n)$ act on $(\mathbb{C}^n)^{\otimes n}(1^n)$ according to

$$\mathbf{Q}(\pi) \bigotimes_{j=1}^n |s(j)\rangle = \bigotimes_{j=1}^n |\pi(s(j))\rangle \qquad \text{and} \qquad \mathbf{P}(\pi) \bigotimes_{j=1}^n |s(j)\rangle = \bigotimes_{j=1}^n |s(\pi^{-1}(j))\rangle \tag{8.36}$$

And from the definition of multiplying permutations, $\mathbf{L}$ and $\mathbf{R}$ act on $\mathbb{C}[\mathcal{S}_n]$ according to

$$\mathbf{L}(\pi) \bigotimes_{j=1}^n |s(j)\rangle = \bigotimes_{j=1}^n |\pi(s(j))\rangle \qquad \text{and} \qquad \mathbf{R}(\pi) \bigotimes_{j=1}^n |s(j)\rangle = \bigotimes_{j=1}^n |s(\pi^{-1}(j))\rangle, \tag{8.37}$$

if we write permutations as elements of $[n]^n$. Thus the embedding map relates $\mathbf{L}$ and $\mathbf{R}$ to $\mathbf{Q}$ and $\mathbf{P}$ respectively.

This means that for any $\pi_1, \pi_2 \in \mathcal{S}_n$, the isomorphism (4) maps $\sum_\lambda |\lambda\rangle\langle\lambda| \otimes \mathbf{p}_\lambda(\pi_1) \otimes \mathbf{p}_\lambda(\pi_2)$ to $\sum_\lambda |\lambda\rangle\langle\lambda| \otimes \mathbf{q}^n_\lambda(\pi_1)|_{\mathcal{Q}^n_\lambda(1^n)} \otimes \mathbf{p}_\lambda(\pi_2)$. This proves that $\mathcal{Q}^n_\lambda(1^n) \overset{\mathcal{S}_n}{\cong} \mathcal{P}_\lambda$ (cf. Thm 9.2.8 of [GW98]).

Moreover, it is straightforward to verify that the GZ basis of $\mathcal{Q}^n_\lambda(1^n)$ corresponds to the same chain of partitions that labels the GZ basis of $\mathcal{P}_\lambda$; one need only look at which weights appear in the restriction to $\mathcal{S}_{n-1} \subset \mathcal{U}_{n-1}$. This establishes that the representation matrices are the same, up to an arbitrary phase difference for each basis vector. The existence of this phase means that we have constructed a slightly different Fourier transform than [Bea97], and it is an interesting open question to calculate this phase difference and determine its significance.

Bibliography

[AH03] A. Abeyesinghe and P. Hayden. Generalized remote state preparation: Trading cbits, qubits and ebits in quantum communication. *Phys. Rev. A*, **68**, 062319, 2003. quant-ph/0308143.

[AHSW04] A. Abeyesinghe, P. Hayden, G. Smith, and A.J. Winter. Optimal superdense coding of entangled states, 2004. quant-ph/0407061.

[Art95] M. Artin. *Algebra*. Prentice Hall, New Jersey, 1995.

[AS04] A. Ambainis and A. Smith. Small pseudo-random families of matrices: Derandomizing approximate quantum encryption. In K. Jansen, S. Khanna, J.D.P. Rolim, and D. Ron, editors, *APPROX-RANDOM*, Volume 3122 of *Lecture Notes in Computer Science*, pp. 249–260. Springer, 2004. quant-ph/0404075.

[Bac01] D. Bacon. *Decoherence, Control, and Symmetry in Quantum Computers*. Ph.D. thesis, University of California at Berkeley, Berkeley, CA, 2001. quant-ph/0305025.

[Bar95] A. Barenco. A universal two-bit gate for quantum computation. *Proc. Roy. Soc. London Ser. A*, **449**, 679–683, 1995.

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing*, pp. 175–179, IEEE, New York, 1984.

[BBC$^+$93] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, **70**, 1895–1899, 1993.

[BBC98] G. Brassard, S.L. Braunstein, and R. Cleve. Teleportation as a quantum computation. *Physica D*, **120**, 43–47, 1998.

[BBPS96] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, **53**, 2046–2052, 1996. quant-ph/9511030.

[BCF$^+$01] H. Barnum, C.M. Caves, C.A. Fuchs, R. Jozsa, and B.W. Schumacher. On quantum coding for ensembles of mixed states. *J. Phys. A*, **34**(35), 6767–6785, 2001. quant-ph/0008024.

[BCH04] D. Bacon, I. L. Chuang, and A. W. Harrow. Efficient quantum circuits for Schur and Clebsch-Gordan transforms, 2004. quant-ph/0407082.

[BCH05a] D. Bacon, I. L. Chuang, and A. W. Harrow. The quantum schur transform: I. Efficient qudit circuits, 2005. In preparation.

[BCH05b] D. Bacon, I. L. Chuang, and A. W. Harrow. The quantum schur transform: II. Connections to the quantum Fourier transform, 2005. In preparation.

[BCL$^+$02] C. H. Bennett, J. I. Cirac, M. S. Leifer, D. W. Leung, N. Linden, S. Popescu, and G. Vidal. Optimal simulation of two-qubit Hamiltonians using general local operations. *Phys. Rev. A*, **66**, 012305, 2002. quant-ph/0107035.

[BDH$^+$05] C.H. Bennett, I. Devetak, A.W. Harrow, P.W. Shor, and A.J. Winter. The quantum reverse Shannon theorem, 2005. In preparation.

[BDSS04] C. H. Bennett, I. Devetak, P. W. Shor, and J. A. Smolin. Inequalities and separations among assisted capacities of quantum channels, 2004. quant-ph/0406086.

[BDSW96] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed state entanglement and quantum error correction. *Phys. Rev. A*, **52**, 3824–3851, 1996. quant-ph/9604024.

[Bea97] R. Beals. Quantum computation of Fourier transforms over symmetric groups. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation (STOC)*, pp. 48–53, ACM Press, El Paso, Texas, 1997.

[BGNP01] D. Beckman, D. Gottesman, M.A. Nielsen, and J. Preskill. Causal and localizable quantum operations. *Phys. Rev. A*, **64**, 052309, 2001. quant-ph/0102043.

[Bha97] R. Bhatia. *Matrix Analysis.* Number 169 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1997.

[BHL$^+$05] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. J. Winter. Remote preparation of quantum states. *IEEE Trans. Inf. Theory*, **51**(1), 56–74, 2005. quant-ph/0307100.

[BHLS03] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin. On the capacities of bipartite Hamiltonians and unitary gates. *IEEE Trans. Inf. Theory*, **49**(8), 1895–1911, 2003. quant-ph/0205057.

[Bie63] L. C. Biedenharn. On the representations of the semisimple lie groups: I. the explicit construction of invariants for the unimodular unitary group in n dimensions. *J. Math. Phys.*, **4**(3), 1963.

[BKN00] H. Barnum, E. Knill, and M. A. Nielsen. On quantum fidelities and channel capacities. *IEEE Trans. Inf. Theory*, **46**, 1317–1329, 2000. quant-ph/9809010.

[BL68] L. C. Biedenharn and J. D. Louck. A pattern calculus for tensor operators in the unitary groups. *Comm. Math. Phys.*, **8**, 89–131, 1968.

[BOM04] M. Ben-Or and D. Mayers. General security definition and composability for quantum & classical protocols, 2004. quant-ph/0409062.

[BRS03] S.D. Bartlett, T. Rudolph, and R.W. Spekkens. Classical and quantum communication without a shared reference frame. *Phys. Rev. Lett.*, **91**, 027901, 2003. quant-ph/0302111.

[BRS04] S.D. Bartlett, T. Rudolph, and R.W. Spekkens. Decoherence-full subsystems and the cryptographic power of a private shared reference frame. *Phys. Rev. A*, **70**, 032307, 2004. quant-ph/0403161.

[BS03a] D. W. Berry and B. C. Sanders. Relation between classical communication capacity and entanglement capability for two-qubit unitary operations. *Phys. Rev. A*, **68**, 032312, 2003. quant-ph/0207065.

[BS03b] D. W. Berry and B. C. Sanders. Relations for classical communication capacity and entanglement capability of two-qubit operations. *Phys. Rev. A*, **67**, 040302(R), 2003. quant-ph/0205181.

[BSST02] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Inf. Theory*, **48**, 2637–2655, 2002. quant-ph/0106052.

[BV93] E. Bernstein and U. Vazirani. Quantum complexity theory. In *Proceedings of the 25th Annual ACM Symposium on the Theory of Computation (STOC)*, pp. 11–20, ACM Press, El Paso, Texas, 1993.

[BW92] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Phys. Rev. Lett.*, **69**, 2881–2884, 1992.

[BW05] C. H. Bennett and A.J. Winter, 2005. In preparation.

[CA97] N. J. Cerf and C. Adami. Negative entropy and information in quantum mechanics. *Phys. Rev. Lett.*, **79**, 5194–5197, 1997. quant-ph/9512022.

[CD96] R. Cleve and D.P. DiVincenzo. Schumacher's quantum data compression as a quantum computation. *Phys. Rev. A*, **54**(4), 2636–2650, 1996. quant-ph/9603009.

[CDKL01] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein. Entangling operations and their implementation using a small amount of entanglement. *Phys. Rev. Lett.*, **86**, 544–547, 2001. quant-ph/0007057.

[CGB00] Anthony Chefles, Claire R. Gilson, and Stephen M. Barnett. Entanglement and collective quantum operations, 2000. quant-ph/0003062.

[Chu36] A. Church. An unsolvable problem of elementary number theory. *Am. J. Math.*, **58**, 345–363, 1936.

[CK81] I. Csiszár and J. Körner. *Information Theory: coding theorems for discrete memoryless systems.* Academic Press, New York–San Francisco–London, 1981.

[CLL05] A. M. Childs, D. W. Leung, and H.-K. Lo. Two-way quantum communication channels, 2005. quant-ph/0506039.

[CLP01] D. Collins, N. Linden, and S. Popescu. The non-local content of quantum operations. *Phys. Rev. A*, **64**, 032302, 2001. quant-ph/0005102.

[CLS02] A. M. Childs, D. W. Leung, and J. A. Smolin, 2002. private communication.

[CLVV03] A. M. Childs, D. W. Leung, F. Verstraete, and G. Vidal. Asymptotic entanglement capacity of the Ising and anisotropic Heisenberg interactions. *Quantum Inf. Comput.*, **3**, 97–105, 2003. quant-ph/0207052.

[CM04] M. Christandl and G. Mitchison. The spectra of density operators and the kronecker coefficients of the symmetric group, 2004. quant-ph/0409016.

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Series in Telecommunication. John Wiley and Sons, New York, 1991.

[DB01] I. Devetak and T. Berger. Low-entanglement remote state preparation. *Phys. Rev. Lett.*, **87**, 197901, 2001. quant-ph/0102123.

[Dev05a] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, **51**(1), 44–55, 2005. quant-ph/0304127.

[Dev05b] I. Devetak. A triangle of dualities: reversibly decomposable quantum channels, source-channel duality, and time reversal, 2005. quant-ph/0505138.

[DHLS05] I. Devetak, P. Hayden, D. W. Leung, and P.W. Shor. Triple trade-offs in quantum Shannon theory, 2005. In preparation.

[DHW04] I. Devetak, A. W. Harrow, and A. J. Winter. A family of quantum protocols. *Phys. Rev. Lett.*, **93**, 239503, 2004. quant-ph/0308044.

[DHW05] I. Devetak, A. W. Harrow, and A. J. Winter. Quantum Shannon theory, resource inequalities, and optimal trade-offs for a family of quantum protocols, 2005. In preparation.

[DN05] C.M. Dawson and M.A. Nielsen. The Solovay-Kitaev algorithm, 2005. quant-ph/0505030.

[DS03] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information, 2003. quant-ph/0311131.

[DSS98] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum channel capacity of very noisy channels. *Phys. Rev. A*, **57**, 830–839, 1998. quant-ph/9706061.

[DVC$^+$01] W. Dür, G. Vidal, J. I. Cirac, N. Linden, and S. Popescu. Entanglement capabilities of non-local Hamiltonians. *Phys. Rev. Lett.*, **87**, 137901, 2001. quant-ph/0006034.

[DW03a] I. Devetak and A.J. Winter. Classical data compression with quantum side information. *Phys. Rev. A*, **68**, 042301, 2003. quant-ph/0209029.

[DW03b] I. Devetak and A.J. Winter. Distilling common randomness from bipartite quantum states. *IEEE Trans. Inf. Theory*, **50**, 3138–3151, 2003. quant-ph/0304196.

[DW04] I. Devetak and A.J. Winter. Relating quantum privacy and quantum coherence: an operational approach. *Phys. Rev. Lett.*, **93**, 080501, 2004. quant-ph/0307053.

[DW05a] I. Devetak and A.J. Winter. Distillation of secret key and entanglement from quantum states. *Proc. R. Soc. Lond. A*, **461**, 207–235, 2005. quant-ph/0306078.

[DW05b] I. Devetak and A.J. Winter. Maximal and average error capacity regions coincide—under randomised encodings, 2005. In preparation.

[EHK97] M. Ettinger, P. Høyer, and E. Knill. Hidden subgroup states are almost orthogonal, 1997. quant-ph/9901034.

[Fan73] M. Fannes. A continuity property of the entropy density for spin lattices. *Commun. Math. Phys.*, **31**, 291–294, 1973.

[FH91] W. Fulton and J. Harris. *Representation Theory – A First Course*. Springer–Verlag, 1991.

[FvdG99] C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum mechanical states. *IEEE Trans. Inf. Theory*, **45**(4), 1216–1227, 1999. quant-ph/9712042.

[GC99] D. Gottesman and I.L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, **402**, 390–393, 1999. quant-ph/9908010.

[GC01] D. Gottesman and I.L. Chuang. Quantum digital signatures, 2001. quant-ph/0105032.

[Geo99] H. Georgi. *Lie Algebras in Particle Physics*. Perseus Books Group, 1999.

[GM02] R. Gill and S. Massar. State estimation for large ensembles. *Phys. Rev. A*, **61**, 042312, 2002. quant-ph/9902063.

[Got99] D. Gottesman. *Group 22: Proc. XXII International Colloquium on Group Theoretical Methods in Physics*. International Press, Cambridge, MA, 1999.

[Gro96] L. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computation (STOC)*, pp. 212–219, ACM Press, El Paso, Texas, 1996. quant-ph/9605043.

[GW98] R. Goodman and N.R. Wallach. *Representations and Invariants of the Classical Groups*. Cambridge University Press, 1998.

[GZ50] I.M. Gelfand and M.L. Zetlin. Matrix elements for the unitary groups. *Dokl. Akad. Nauk.*, **71**, 825–828, 1950.

[Har04] A. W. Harrow. Coherent communication of classical messages. *Phys. Rev. Lett.*, **92**, 097902, 2004. quant-ph/0307091.

[Hay01] M. Hayashi. Asymptotics of quantum relative entropy from representation theoretical viewpoint. *J. Phys. A*, **34**, 3413–3419, 2001. quant-ph/9704040.

[Hay02a] M. Hayashi. Exponents of quantum fixed-length pure state source coding, 2002. quant-ph/0202002.

[Hay02b] M. Hayashi. Optimal sequence of quantum measurements in the sense of stein's lemma in quantum hypothesis testing. *J. Phys. A*, **35**, 10759–10773, 2002. quant-ph/0208020.

[Hel76] C. W. Helstrom. *Quantum Detection and Estimation Theory*. Academic, New York, 1976.

[HHH$^+$01] M. Horodecki, P. Horodecki, R. Horodecki, D. W. Leung, and B. M. Terhal. Classical capacity of a noiseless quantum channel assisted by noisy entanglement. *Quantum Inf. Comput.*, **1**(3), 70–78, 2001. quant-ph/0106080.

[HHH05] A. Hayashi, T. Hashimoto, and M. Horibe. Extended quantum color coding. *Phys. Rev. A*, **71**, 012326, 2005. quant-ph/0409173.

[HHL04] A. W. Harrow, P. Hayden, and D. W. Leung. Superdense coding of quantum states. *Phys. Rev. Lett.*, **92**, 187901, 2004. quant-ph/0307221.

[HJW02] P. Hayden, R. Jozsa, and A.J. Winter. Trading quantum for classical resources in quantum data compression. *J. Math. Phys.*, **43**(9), 4404–4444, 2002. quant-ph/0204038.

[HL04] A. W. Harrow and H.-K. Lo. A tight lower bound on the classical communication cost of entanglement dilution. *IEEE Trans. Inf. Theory*, **50**(2), 319–327, 2004.

[HL05] A. W. Harrow and D. W. Leung. Bidirectional coherent classical communication. *Quantum Inf. Comput.*, **5**(4–5), 380–395, 2005. quant-ph/0412126.

[HM01] M. Hayashi and K. Matsumoto. Variable length universal entanglement concentration by local operations and its application to teleportation and dense coding, 2001. quant-ph/0109028.

[HM02a] M. Hayashi and K. Matsumoto. Quantum universal variable-length source coding. *Phys. Rev. A*, **66**(2), 022311, 2002. quant-ph/0202001.

[HM02b] M. Hayashi and K. Matsumoto. Simple construction of quantum universal variable-length source coding. *Quantum Inf. Comput.*, **2**, 519–529, 2002. quant-ph/0209124.

[HM02c] M. Hayashi and K. Matsumoto. Universal distortion-free entanglement concentration, 2002. quant-ph/0209030.

[HN03] M. Hayashi and H. Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Trans. Inf. Theory*, **49**(7), 1753–1768, 2003. quant-ph/0206186.

[Hoë63] Wassily Hoëffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**(1), 13–30, March 1963.

[Hol73] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, **9**, 177–183, 1973.

[Hol98] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory*, **44**, 269–273, 1998. quant-ph/9611023.

[Hol01] A. S. Holevo. *Statistical Structure of Quantum Theory*, Volume 67 of *Lecture Notes in Physics*. Springer, Berlin, 2001.

[Hol02] A. S. Holevo. On entanglement assisted classical capacity. *J. Math. Phys.*, **43**(9), 4326–4333, 2002. quant-ph/0106075.

[How89] R. Howe. Transcending classical invariant theory. *J. Amer. Math. Soc.*, **2**(3), 535–552, 1989.

[HVC02] K. Hammerer, G. Vidal, and J. I. Cirac. Characterization of non-local gates. *Phys. Rev. A*, **66**, 062321, 2002. quant-ph/0205100.

[HW03] P. Hayden and A.J. Winter. On the communication cost of entanglement transformations. *Phys. Rev. A*, **67**, 012306, 2003. quant-ph/0204092.

[Jam72] A. Jamiolkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.*, **3**, 275–278, 1972.

[JK81] G. D. James and A. Kerber. *The representation theory of the symmetric group.* Addison-Wesley, Reading, Mass., 1981.

[Joz94] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Opt.*, **41**, 2315–2323, 1994.

[JS94] R. Jozsa and B. Schumacher. A new proof of the quantum noiseless coding theorem. *J. Mod. Opt.*, **41**, 2343–2349, 1994.

[KBG01] N. Khaneja, R. Brockett, and S. J. Glaser. Time optimal control in spin systems. *Phys. Rev. A*, **63**, 032308, 2001. quant-ph/0006114.

[KBLW01] J. Kempe, D. Bacon, D. A. Lidar, and K. B. Whaley. Theory of decoherence-free fault-tolerant quantum computation. *Phys. Rev. A*, **63**, 042307, 2001. quant-ph/0004064.

[KC01] B. Kraus and J. I. Cirac. Optimal creation of entanglement using a two–qubit gate. *Phys. Rev. A*, **63**, 062309, 2001. quant-ph/0011050.

[Key04] M. Keyl. Quantum state estimation and large deviations, 2004. quant-ph/0412053.

[KI01] M. Koashi and N. Imoto. Compressibility of quantum mixed-state signals. *Phys. Rev. Lett.*, **87**, 017902, 2001. quant-ph/0103128.

[Kit04] A. Kitaev, 2004. private communication.

[KLV00] E. Knill, R. Laflamme, and L. Viola. Theory of quantum error correction for general noise. *Phys. Rev. Lett.*, **84**, 2525–2528, 2000. quant-ph/9908066.

[Kly04] A. Klyachko. Quantum marginal problem and representations of the symmetric group, 2004. quant-ph/0409113.

[KM01] P. Kaye and M. Mosca. Quantum networks for concentrating entanglement. *J. Phys. A*, **34**, 6939–6948, 2001. quant-ph/0101009.

[Kni04] E. Knill. Fault-tolerant postselected quantum computation: Schemes, 2004. quant-ph/0402171.

[KNTSZ01] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *ACM Symposium on Theory of Computing*, pp. 124–133, 2001. quant-ph/0005106 and quant-ph/0004100.

[KR03] A. Klappenecker and M. Roetteler. Quantum software reusability. *International Journal on Foundations of Computer Science*, **14**(5), 777–796, 2003. quant-ph/0309121.

[KSV02] A. Yu Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*, Volume 47 of *Graduate Studies in Mathematics*. AMS, 2002.

[Kup03] G. Kuperberg. A subexponential-time quantum algorithm for the dihedral hidden subgroup problem, 2003. quant-ph/0302112.

[KW01] M. Keyl and R. F. Werner. Estimating the spectrum of a density operator. *Phys. Rev. A*, **64**, 052311, 2001. quant-ph/0102027.

[KW04] D. Kretschmann and R. F. Werner. *Tema Con Variazioni:* quantum channel capacity. *New J. Phys.*, **6**, 26, 2004. quant-ph/0311037.

[Leu02] D. W. Leung. Quantum vernam cipher. *Quantum Inf. Comput.*, **2**(1), 14–34, 2002. quant-ph/0012077.

[Leu04] D.W. Leung, 2004. private communication.

[LHL03] M. S. Leifer, L. Henderson, and N. Linden. Optimal entanglement generation from quantum operations. *Phys. Rev. A*, **67**, 012306, 2003. quant-ph/0205055.

[Llo96] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, **55**, 1613–1622, 1996. quant-ph/9604015.

[LNC03] LNCS 2989. *Commitment Capacity of Noisy Channels*, Berlin, 2003. Springer. cs.CR/0304014.

[Lou70] J. D. Louck. Recent progress toward a theory of tensor operators in unitary groups. *Am. J. Phys.*, **38**(1), 3–42, 1970.

[LP99] H.-K. Lo and S. Popescu. The classical communication cost of entanglement manipulation: Is entanglement an inter-convertible resource? *Phys. Rev. Lett.*, **83**, 1459–1462, 1999.

[LR73] E. H. Lieb and M. B. Ruskai. Proof of strong subadditivity of quantum-mechanical entropy. *J. Math. Phys.*, **14**, 1938–1941, 1973.

[LS03] D. W. Leung and P. W. Shor. Oblivious remote state preparation. *Phys. Rev. Lett.*, **90**, 127905, 2003. quant-ph/0201008.

[Mak02] Y. Makhlin. Nonlocal properties of two-qubit gate and mixed states and optimization of quantum computation. *Quantum Inf. Process.*, **1**, 243–252, 2002. quant-ph/0002045.

[Mes62] A. Messiah. *Quantum Mechanics, Vol. 2*, chapter Representation of Irreducible Tensor Operators: Wigner-Eckart Theorem, pp. 573–575. North-Holland, Amsterdam, Netherlands, 1962.

[Mor95] E. Moran. *Bradymania! (25th Anniversary Edition)*. Adams Media Corporation, Avon, MA, 1995.

[MR97] D.K. Maslen and D.N. Rockmore. Generalized FFTS - A survey of some recent results. In L. Finkelstein and W.M. Kantor, editors, *Groups and Computation II*, Volume 28 of *DIMACS Series in Disc. Math. and Theoret. Comput. Sci.*, pp. 183–237, 1997.

[MRR04] Cristopher Moore, Daniel Rockmore, and Alexander Russell. Generic quantum Fourier transforms. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 778–787, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004. quant-ph/0304064.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, New York, 2000.

[NDD$^+$03] M. A. Nielsen, C. M. Dawson, J. L. Dodd, A. Gilchrist, D. Mortimer, T. J. Osborne, M. J. Bremner, A. W. Harrow, and Andrew Hines. Quantum dynamics as a physical resource. *Phys. Rev. A*, **67**, 052301, 2003. quant-ph/0208077.

[Nie98] M. A. Nielsen. *Quantum information theory*. Ph.D. thesis, University of New Mexico, Albuquerque, NM, 1998.

[Nie99a] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, **83**, 436–439, 1999. quant-ph/9811053.

[Nie99b] M. A. Nielsen. Majorization and its applications to quantum information theory, 1999. Lecture notes for class at Caltech.

[Nie00] M. A. Nielsen. Continuity bounds for entanglement. *Phys. Rev. A*, **61**(6), 064301, 2000. quant-ph/9908086.

[Nie05] M. A. Nielsen. A geometric approach to quantum circuit lower bounds, 2005. quant-ph/0502070.

[Per93] A. Peres. *Quantum theory: concepts and methods*. Kluwer Academic, Dordrecht, 1993.

[Pin64] M.S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, 1964.

[PRB99] M. Püschel, M. Rötteler, and T. Beth. Fast quantum fourier transforms for a class of non-abelian groups. In M. P. C. Fossorier, H. Imai, S. Lin, and A. Poli, editors, *AAECC*, Volume 1719 of *Lecture Notes in Computer Science*, pp. 148–159. Springer, 1999. quant-ph/9807064.

[Pre98] J. Preskill. Caltech ph229 lecture notes. unpublished lecture notes, 1998.

[Pre99] J. Preskill. Plug-in quantum software. *Nature*, **402**, 357–358, 1999.

[RG02] T. Rudolph and L. Grover. Quantum searching a classical database (or how we learned to stop worrying and love the bomb), 2002. quant-ph/0206066.

[RZBB94] M. Reck, A. Zeilinger, H.J. Bernstein, and P. Bertani. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, **73**, 58–61, 1994.

[SBM04] V.V. Shende, S.S. Bullock, and I.L. Markov. Synthesis of quantum logic circuits, 2004. quant-ph/0406176.

[Sch95] B. Schumacher. Quantum coding. *Phys. Rev. A*, **51**, 2738–2747, 1995.

[Sch96] B. Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, **54**, 2614–2628, 1996. quant-ph/9604023.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. Jnl.*, **27**, 379–423, 623–656, 1948.

[Sha61] C. E. Shannon. *Proc. 4th Berkeley Symp. Math. Stat. Prob.* UC Press, Berkeley, CA, 1961.

[Sho94] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In S. Goldwasser, editor, *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, pp. 124–134, IEEE Computer Society, Los Alamitos, CA, 1994.

[Sho96] P. W. Shor. Fault tolerant quantum computation. In *Proceedings of the 37th Symposium on the Foundations of Computer Science*, pp. 56–65, IEEE, Los Alamitos, CA, 1996. quant-ph/9605011.

[Sho02] P. W. Shor. The quantum channel capacity and coherent information. MSRI workshop on quantum computation, 2002.

[Sho03] P. W. Shor. Equivalence of additivity questions in quantum information theory, 2003. quant-ph/0305035.

[Sho04a] P. W. Shor, 2004. private communication.

[Sho04b] P. W. Shor. The classical capacity achievable by a quantum channel assisted by limited entanglement, 2004. quant-ph/0402129.

[SN96] B. Schumacher and M. A. Nielsen. Quantum data processing and error correction. *Phys. Rev. A*, **54**, 2629–2635, 1996. quant-ph/9604022.

[SST01] P. W. Shor, J. A. Smolin, and Barbara M. Terhal. Nonadditivity of bipartite distillable entanglement follows from conjecture on bound entangled werner states. *Phys. Rev. Lett.*, **86**, 2681–2684, 2001. quant-ph/0010054.

[Sti55] W. F. Stinespring. Positive functions on $\mathbb{C}^*$-algebras. *Proc. Amer. Math. Soc.*, **6**, 211–216, 1955.

[SW97] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, **56**, 131–138, 1997.

[Sze04] M. Szegedy. Spectra of quantized walks and a $\sqrt{\delta\epsilon}$ rule, 2004. quant-ph/0401053.

[TH00] B. M. Terhal and P. Horodecki. A Schmidt number for density matrices. *Phys. Rev. A*, **61**, 040301, 2000. quant-ph/9911117.

[Tha99] A. V. Thapliyal. On multipartite pure-state entanglement. *Phys. Rev. A*, **59**, 3336–3342, 1999. quant-ph/9811091.

[Tur36] A.M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proc. London Math. Soc.*, **42**, 230–265, 1936.

[Uhl76] A. Uhlmann. The 'transition probability' in the state space of a $*$-algebra. *Rep. Math. Phys.*, **9**, 273–279, 1976.

[Unr04] D. Unruh. Simulatable security for quantum protocols, 2004. quant-ph/0409125.

[VC02a] G. Vidal and J. I. Cirac. Catalysis in non–local quantum operations. *Phys. Rev. Lett.*, **88**, 167903, 2002.

[VC02b] G. Vidal and J. I. Cirac. Optimal simulation of nonlocal Hamiltonians using local operations and classical communication. *Phys. Rev. A*, **66**, 022315, 2002. quant-ph/0108076.

[vDH03] W. van Dam and P. Hayden. Universal entanglement transformations without communication. *Phys. Rev. A*, **67**(6), 060302(R), 2003. quant-ph/0201041.

[vE05] S.J. van Enk. Quantifying the resource of sharing a reference frame. *Phys. Rev. A*, **71**, 032339, 2005. quant-ph/0410083.

[VHC02] G. Vidal, K. Hammerer, and J. I. Cirac. Interaction cost of non-local gates. *Phys. Rev. Lett.*, **88**, 237902, 2002. quant-ph/0112168.

[vKK04] J. von Korff and J. Kempe. Quantum advantage in transmitting a permutation. *Phys. Rev. Lett.*, **93**(46), 260502, 2004. quant-ph/0405086.

[VLPT99] G. Vidal, J.I. Latorre, P. Pascual, and R. Tarrach. Optimal minimal measurements of mixed states. *Phys. Rev. A*, **60**, 126–135, 1999. quant-ph/9812068.

[Win99a] A.J. Winter. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory*, **45**(7), 2481–2485, 1999.

[Win99b] A.J. Winter. *Coding Theorems of Quantum Information Theory.* Ph.D. thesis, Universität Bielefeld, Germany, 1999. quant-ph/9907077.

[Win02] A.J. Winter. Compression of sources of probability distributions and density operators, 2002. quant-ph/0208131.

[Win04] A.J. Winter. “Extrinsic” and “intrinsic” data in quantum measurements: asymptotic convex decomposition of positive operator valued measures. *Comm. Math. Phys.*, **244**(1), 157–185, 2004. quant-ph/0109050.

[Wyn75] A. Wyner. The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, **21**(2), 163–179, 1975.

[WZ82] W.K. Wootters and W.H. Zurek. A single quantum cannot be cloned. *Nature*, **299**, 802–803, 1982.

[Yar05] J. Yard. *Simultaneous classical-quantum capacities of quantum multiple access channels.* Ph.D. thesis, Stanford University, Stanford, CA, 2005. quant-ph/0506050.

[YDH05] J. Yard, I. Devetak, and P. Hayden. Capacity theorems for quantum multiple access channels - part I: Classical-quantum and quantum-quantum capacity regions, 2005. quant-ph/0501045.

[Zal04] C. Zalka. Implementing high dimensional unitary representations of su(2) on a quantum computer, 2004. quant-ph/0407140.

[ZR97] P. Zanardi and M. Rasetti. Error avoiding quantum codes. *Mod. Phys. Lett. B*, **11**(25), 1085–1093, 1997. quant-ph/9710041.