Entanglement and interference resources in quantum computation and communication

by

Daniel Lee Stahlke

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
Department of Physics
Pittsburgh, Pennsylvania

Advised by Professor Robert Griffiths

July 11, 2014

Abstract

This dissertation contains results on three quite different topics. First, I investigate the entanglement resources required for two parties to jointly implement a unitary operation using local operations and classical communication (LOCC). If this resource is of smallest feasible Schmidt rank then it must be maximally entangled. If the Schmidt rank is higher, less entanglement may suffice.

Second, I investigate the source of the "quantum speedup". I quantify quantum interference and show that in order for a quantum computer to be significantly faster than a classical computer, it must make use of operations capable of producing large amounts of interference (or a large number of operations that can produce small amounts of interference). A quantum computer not making use of such a resource can be efficiently simulated by a classical computer.

Third, I investigate zero-error source-channel coding. In this scenario, Alice wishes to convey a message to Bob through a noisy channel, with zero chance of error, in the case where Bob already has some prior knowledge regarding the message that is to be sent. For classical messages and classical channels, it was known that three graph invariants of Lovász, Szegedy, and Schrijver provide necessary conditions for this task. We show that this applies also when Alice and Bob make use of an entangled state, unifying and extending a series of previous results.

Finally, I introduce a fully quantum version of source-channel coding where the message to be sent, the channel, and the side information are all quantum. Whereas the classical case makes use of graphs, the quantum case makes use of non-commutative graphs. I generalize the concept of graph homomorphism, as well as the Szegedy and Schrijver numbers, to non-commutative graphs and show that the necessary conditions for the classical case generalize to the quantum case. Using this theory, I construct a quantum channel whose one-shot zero-error entanglement assisted capacity can only be unlocked using a non-maximally entangled state, showing that in this case less is more when it comes to entanglement resources.

Acknowledgments

I would like to thank most of all my family, especially my wife Sarah, for support, encouragement, and patience. I thank my advisor Robert B. Griffiths, and the quantum information group at Carnegie Mellon University, for guidance and for many helpful discussions. Finally, I would like to thank Simone Severini for introducing me to a wonderful group of collaborators.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

This introductory chapter begins with a brief overview regarding the history of and contemporary issues in the field of quantum information, in relation to the results of this dissertation (section 1.2). Some mathematics and physics background will be introduced in section 1.3. Section 1.4 consists of a brief overview of the technical results of this dissertation.

1.2 Historical background and context

1.2.1 Quantum computing

The idea that quantum mechanics could be harnessed to solve problems (specifically, simulation of quantum systems) that are infeasible on traditional computers was discussed over 30 years ago by Feynman [Fey82], though some information theoretical investigation of quantum mechanics predates this [Wik14].

Deutsch and Jozsa developed the first quantum algorithm that is provably much faster than any possible classical algorithm [DJ92]. "Faster" in this context means that the runtime of the quantum algorithm scales in a better way as the problem size grows; although the quantum algorithm may be slower for smaller problem instances due to, for instance, quantum gates being slower than classical digital logic gates, for larger problems the quantum algorithm will be superior. In the Deutsch–Jozsa algorithm, one is given access to some function f which takes as input integers from 1 to 2^n as produces as output 0 or 1. Access to f is via a sort of black box in which one is allowed to query $f(x)$ for a given x but is not allowed to peak inside the box to see how it works. The goal is to determine whether f is a constant function or whether it takes value 1 exactly half of the time. The Deutsch–Jozsa algorithm solves this problem using only a single query, provided that the box may be queried using a quantum superposition. Any classical algorithm would require at least $2^{n-1}+1$ queries in the worst case, in order to solve the problem with certainty.

A weakness of this result is that a classical algorithm could solve the problem nearly as fast as Deutsch and Jozsa's quantum algorithm, provided one allows a very small probability of error. Simon improved upon this result, constructing a slightly more complicated problem which can be solved using order of n queries by a quantum computer but which requires at least order of $2^{n/2}$ queries on a classical computer, even if one allows some probability of error [Sim94]. As was the case with the Deutsch–Jozsa problem, Simon's problem involves a function f which one can only access in a black-box manner (i.e. one is given a box which, given x, produces $f(x)$). Simon's problem is to determine the value s such that $f(x) = f(y)$ if and only if $y = x \oplus s$, given that such an s exists. The operator $\oplus$ is bitwise addition modulo 2. Simon's problem is of little practical use, but served as the inspiration for Shor's algorithm for finding the prime factors of a number [Sho99]. Shor's algorithm runs in time order of n^3 where n is the length of the number to be factored, whereas the fastest known classical algorithm runs in time order of $2^{cn^{1/3}(\log n)^{2/3}}$ for $c \approx 1.92$ [Wei14]. Shor's algorithm is important since, if implemented, it would negate the security of many widely used public key cryptography schemes. For example, the RSA [RSA78] and elliptic curve [Kob87, Mil86] cryptosystems are vulnerable to Shor's algorithm (or variations thereof). The McEliece [McE78] cryptosystem is not known to be vulnerable. For more information regarding which cryptographic schemes are or are not vulnerable to quantum algorithms, refer to [BLCP14].

Grover devised a quantum algorithm for the problem of unstructured search ("searching for a needle in a haystack") [Gro96]. Here the goal is to find an input x such that $f(x) = 1$, given black box access to some function f which takes as input numbers from 1 to N and produces as output 0 or 1. As was the case with the Deutsch–Jozsa problem, "black box access" here means that one may query f to obtain $f(x)$ for a given x but one is not for example given any description of a circuit which implements f. Grover's algorithm requires only order of $\sqrt{N}$ queries of f whereas classically an average of $N/2$ queries are needed. Grover's algorithm solves a very widely applicable problem;

however, the gains are modest unless N is extremely large. Because the quantum advantage is so modest for Grover's algorithm when N is small, it will probably not have practical implications until it is possible to build a quantum computer with gate speed and gate count comparable to that of today's classical computers.

Although these algorithms, and others, demonstrate that quantum computers can be more powerful than classical computers, the source of this "quantum speedup" remains a bit of a mystery. Ad-hoc explanations abound in relation to particular algorithms (e.g. Grover's or Shor's), but general principles are lacking. Entanglement, measured in various ways, has been shown to be necessary for quantum speedup [Vid03, JL03, Eas10]; however, universal quantum computation (and hence quantum speedup) is possible using states arbitrarily close to having no entanglement at all [Nes12]. Other explanations have been proposed and also found to be problematic. For an overview, see section 9 of [FRS12].

In chapter 3 I propose quantum interference as a necessary resource for quantum speedup. Previously interference has been mentioned as being important for quantum speedup, but only in a vague and non-quantitative way. I define a measure of interference and show that quantum operations incapable of producing much interference (those of low *interference producing capacity*) cannot produce much quantum speedup. This opens up a new previously unexplored avenue in the study of quantum speedup. An intriguing open question is whether interference itself, rather than interference producing capacity, is necessary for quantum speedup.

1.2.2 Entanglement

Here we provide an informal overview of historical developments regarding entanglement. For a formal definition of entanglement see section 1.3.7.

In 1935, Einstein, Podolsky, and Rosen (EPR) proposed a thought experiment in which two spatially separated parties (Alice and Bob, in modern terminology) each measure one of a pair of entangled particles [EPR35]. If the two parties perform similar measurements then the outcome of one perfectly predicts the outcome of the other, similar to a situation in which a coin is cut in half lengthwise with the two pieces (heads and tails) randomly distributed to the two parties: if Alice sees heads, she knows Bob has tails. Now, for the entangled particles, suppose Alice measures position. The outcome of this measurement determines the position of Bob's particle. So Bob's particle must have a well defined position. On the other hand, Alice could measure momentum, and this would determine the momentum of Bob's particle, so Bob's particle must have a well defined momentum. EPR conclude that both the position and the momentum of Bob's particle have well defined values, even though we might not know ahead of time what these values are. Since this is in apparent violation of Heisenberg's uncertainty principle, the theory of quantum mechanics must be incomplete. Bohm (chapter 22 of [Boh51]) applied the conceptual ideas from the EPR thought experiment to a simpler scenario involving spins rather than position and momentum.

The notion that all measurable properties of particles have well defined values (even before measurement), and that these values don't depend upon spatially distant events, is now referred to as a *local hidden variable model*. It turns out such local hidden variable models cannot describe quantum mechanics, since the correlations in EPR-like experiments go beyond what is allowed in a local hidden variable model. This is a subtle point and went unnoticed for nearly thirty years. In 1964, Bell formalized the local hidden variables model by assuming all properties of each particle are well defined (although they may be random due to our lack of knowledge) and measurements on one particle do not affect the properties of the other. He showed the correlations predicted by quantum mechanics fall outside of the set of such correlations [Bel64]. Specifically, he produced an inequality (*Bell's inequality*) involving the probabilities of various measurement outcomes which must be satisfied by any local hidden variables model. The measurement outcomes predicted by quantum mechanics violate this inequality. These beyond-classical correlations due to quantum mechanics have since been experimentally verified in a great variety of ways (see,

e.g., [FC72, AGR81, WJS+98, RKM+01, GMR+13]), so it is widely believed that nature cannot be described in terms of local hidden variable models. The philosophical implications of this result have been widely studied, and are a subject of debate even to this day.

There are also a number of practical applications of entanglement. If a sender and a receiver possess an entangled state, the sender may convey two bits of classical information by transmitting only a single qubit to the receiver; this is known as *superdense coding* [BW92]. Conversely, the sender may transmit the state of a single qubit by communicating two classical bits; this is known as *teleportation* [BBC+93]. These two protocols will be explained in section 1.3.13. More recently, entanglement has found application in the creation of certifiably random numbers [VV12].

Entanglement can be helpful for distributed quantum computation. Suppose one desires to perform a quantum computation on the joint state of two spatially separated subsystems. Further, suppose that one has access to a classical communication channel (e.g. telephone) between these subsystems but cannot communicate quantum information such as qubits. In this situation, performing the joint computation (a unitary operation) requires use of some extra resource such as entanglement.

In chapter 2 I investigate the amount of entanglement needed for this task. One may expect that implementing an operation very close to identity (i.e. one that changes the input state by only a small amount) would not require much entanglement. This is not the case: I prove that if the entanglement resource is of smallest possible Schmidt rank then it must be maximally entangled, regardless of the unitary operation being performed. This extends (via a different and simpler technique) a recent result which only applied for Schmidt rank 2 [STM11a]. If the resource state is of larger Schmidt rank, then less entanglement suffices in some cases. So there is a trade-off between two resources: Schmidt rank and entropy of entanglement.

1.2.3 Quantum generalizations of graph theory

In addition to the philosophical and the practical aspects of entanglement, there are many interesting mathematical aspects. Suppose that Alice and Bob are locked in separate rooms and are not allowed to communicate. A referee asks each of them a question, and Alice and Bob (who are in alliance and can discuss their strategy ahead of time) receive some payoff depending upon their answer. Alice and Bob know ahead of time the types of questions that may be asked and the payoffs corresponding to the various possible answers. If Alice and Bob are in possession of an entangled state, they can sometimes achieve a larger average payoff. The CHSH [CHSH69] variant of Bell's inequality, and many of its generalizations, are of this form (even if not originally stated as such). Specifically, these variants of Bell's inequality place limits upon the maximum possible payoff in the case where no entanglement is involved, and with entanglement these limits are surpassed. For the entanglement enhanced case, Alice and Bob each perform some measurement on their share of the entangled state. The choice of measurement depends upon the question asked by the referee, and Alice and Bob's answers depend upon the outcomes of these measurements. For an overview, see [BCP+14].

One can also consider variations of this game in which certain pairs of answers are absolutely forbidden. Bell's original inequality was of this form: when Alice and Bob are asked to perform the same measurement, their answers must be perfectly anti-correlated. In other words, Bell only considered cases where Alice and Bob never provide the same answer when asked the same question.

Many problems from mathematics and computer science can be phrased in terms of such games. For example, consider the *chromatic number* of a graph. This is the minimum number of colors needed in order to color the vertices of the graph in such a way that adjacent vertices have different colors (a *proper coloring*). Suppose Alice and Bob claim (perhaps deceitfully) to have a proper coloring of a graph using n colors. A skeptical referee may then put them in separate interrogation rooms and ask them each for the color of one of the vertices. If the referee asks them both about the same vertex, their answers must coincide. If the referee asks Alice about one vertex and asks Bob about an adjacent vertex, then they must answer with unequal colors. The referee asks such a question only a single time. Classically, Alice and Bob can win this game (supply correct answers)

with no chance of failure if and only if the graph can indeed be colored by n colors—in other words, if n is at least as large as the chromatic number of the graph.

In 2002, Galliard and Wolf [GW02] showed if Alice and Bob are in possession of an appropriate entangled state then for certain graphs they can win this game even if n is less than the chromatic number; the minimum number of colors needed in this case is now called the *quantum chromatic number*. Other graph concepts, namely clique number and homomorphism (which will be defined in section 1.3.10), can also be cast as games similar to the above and so have quantum generalizations (*quantum clique number* and *quantum homomorphism*). These have been well studied, see [RM12] and the references therein for details.

In general it may be very difficult to determine whether there is a quantum homomorphism between two given graphs; in fact this is not even known to be computable at all. A common technique in such situations is to relax the conditions in order to create a more tractable problem. Any necessary conditions that apply to the relaxed problem apply also to the original problem. In chapter 4 we adopt such a strategy, studying a semidefinite relaxation of the quantum homomorphism problem. We find that this semidefinite relaxation is closely related to necessary conditions given in [RM12, Rob13] involving the Lovász ϑ number and its variants. We also show that this semidefinite relaxation is not tight: there are graphs for which the relaxed conditions can be satisfied but the original conditions cannot.

Recently, even the graphs themselves have gained a quantum generalization [DSW13]. We will expand upon this at the end of the next subsection.

1.2.4 Zero-error information theory

In 1948, Shannon developed a mathematical theory of communication in which he studied the rate at which data can be transmitted through a noisy channel [Sha48]. A telephone wire, a radio transmitter and receiver, or even the postal service can be considered as channels. The formal definition of channels will be given in sections 1.3.8 and 1.3.9. One may consider the number of bits that can be reliably transmitted with a single use of the channel (*one-shot capacity*) or the average number of bits that can be reliably transmitted per use of a channel, in the limit of many uses (*asymptotic capacity*). The entire input is encoded using some compression or error correction scheme before being sent through the channel, and is decoded by the receiver. Typically, one defines "reliable" transmission by mandating the error rate be very small, approaching zero as the number of channel uses increases.

Alternatively, one may require absolutely zero chance of error. This zero-error case was also studied by Shannon [Sha56]. One advantage of studying zero-error capacity is that one can obtain strong and often easily interpreted guarantees on accuracy. For instance, it is possible to store information on an array of hard disks in such a way that if any two disks fail the information can be perfectly recovered with zero chance of error. Here, the array of hard disks may be considered as a channel, transmitting information from yesterday to today; the "noise" of this channel corresponds to loss of any two disks. Zero-error information theory can be used to determine the maximum possible amount of information that can be stored on the disks while still maintaining this level of resilience.

Shannon [Sha56] showed that zero-error capacity for a single use of a channel can be computed via a graph derived from the channel, and the capacity for many parallel uses of the channel is a function of a suitably defined product graph. The definition of this graph, called the *confusability graph*, will be given in section 1.3.11. One may then speak of the *Shannon capacity of a graph*, with the understanding that this corresponds to the zero-error capacity of the underlying channel. The simplest nontrivial case is the pentagon graph C_5. The Shannon capacity of C_5 remained unsolved for over two decades, finally being resolved by Lovász in 1979 [Lov79]. The quantity therein defined is now known as the Lovász ϑ number of a graph (we give the definition in section 1.3.10). The Shannon capacity of the 7-cycle graph (the heptagon) is still an open question.

Recently, a study has begun of the zero-error capacity of channels in the case where sender and receiver are in possession of an entangled state. In 2010, Cubitt, et al. [CLMW10] showed

entanglement can increase the number of error-free messages that can be sent through a classical channel. A similar result for the asymptotic capacity (the limit of many channel uses) appeared in 2012 [LMM+12]. It turns out Lovász's bound on the capacity of a channel still applies in the presence of entanglement [Bei10, DSW13].

One may consider the case where the receiver already has some side information regarding the message to be sent; this is known as *zero-error source-channel coding*. Again the problem is analyzed using graphs, now with one graph representing the source and another graph representing the channel [NTR06]. Very recently, the entanglement assisted version of this problem has been studied [BBL+13]. There, the Lovász number (in fact, even a strengthened version due to Szegedy) was proved to provide a bound on entanglement assisted zero-error source-channel coding in the case where the channel is noiseless.

In chapter 4 we extend this result to the case of a noisy channel, as well as providing other novel results and reproducing many existing results under a unified framework. Specifically, we consider a semidefinite relaxation of this problem, similar to the one described in the previous subsection (in fact, entanglement assisted zero-error source-channel coding is mathematically very similar to the quantum homomorphisms described in the previous subsection). We prove a theorem which encompasses and generalizes the results of both [Bei10] and [BBL+13]. Further, we show that these results are the best that can be obtained by this sort of semidefinite relaxation, answering an open question posed in [Bei10].

The capacity of quantum channels has also been widely studied. Quantum channels take as input quantum states (such as photons) and produce quantum states as output. An example of such a channel is a fiber optic cable that maintains the polarization of light. In 2010, Duan, Severini, and Winter showed that a generalization of the Lovász number provides an upper bound on the zero-error entanglement assisted asymptotic capacity of quantum channels [DSW10, DSW13]. Whereas classical channels are analyzed using graphs, quantum channels are analyzed using what they termed *non-commutative graphs*, which take the form of operator subspaces. These authors concluded their paper with the suggestion that it may be possible to develop a richer theory of non-commutative graphs.

In chapter 5 I define and investigate a fully quantum version of zero-error source-channel coding, and use this to further develop the theory of non-commutative graphs. Specifically, I consider a source which produces quantum states, possibly entangled, with the sender being in possession of one subsystem of the state and the receiver being in possession of the other. The sender makes use of a (possibly noisy) classical or quantum channel to send a message that allows the receiver to decode the original state. In addition to being an interesting quantum information problem in its own right, general enough to include for example teleportation and dense coding, this has provided a path towards further investigation of the theory of non-commutative graphs, a mathematically interesting topic which has not otherwise progressed since its introduction in 2010 [DSW10].

1.3 Mathematics and physics background

This section contains a brief overview of the math and physics background relevant to this dissertation. Many of these topics are covered in more detail in [NC00, Wil13].

1.3.1 Linear algebra terminology

Vector spaces will generally be denoted by upper case letters, although in chapter 2 we will use, e.g., $\mathcal{H}_A$. Unless otherwise noted, vector spaces will be finite dimensional complex Hilbert spaces. The space of linear operators (matrices) on space A will be denoted $\mathcal{L}(A)$. The *adjoint* (i.e. conjugate transpose) of an operator M is written $M^\dagger$. An operator $U \in \mathcal{L}(A)$ is *unitary* if $U^\dagger U = I$. An operator $J : A \to B$ is an *isometry* if $J^\dagger J = I$. An operator M is *Hermitian* if $M = M^\dagger$. An operator M is *positive semidefinite* (PSD) if $\langle\psi|M|\psi\rangle \geq 0$ for all vectors $|\psi\rangle$. Such operators are necessarily

Hermitian. An operator M is *positive definite* if $\langle\psi|M|\psi\rangle > 0$ for all $|\psi\rangle \neq 0$, equivalently if it is PSD and full rank. We write $L \succeq M$ if $L - M$ is PSD and $L \succ M$ if $L - M$ is positive definite.

The dimension of a vector space A is denoted $\dim(A)$. For each vector space, we will single out a particular basis and call it the *standard basis*. The standard basis vectors for space A will be written $|0\rangle_A, \ldots, |\dim(A) - 1\rangle_A$ or, sometimes, $|1\rangle_A, \ldots, |\dim(A)\rangle_A$ when this is more convenient. The subscript A will be left off when the space can be inferred from context. All matrices will be expressed in terms of this basis, and we will write for example $M_{ij} = \langle i|M|j\rangle$.

1.3.2 Qubits and qudits

The simplest quantum system possible is one with two orthogonal states, which in the quantum information community is known as a *qubit*. For example, these two states could be the up or down spins of an electron or the horizontal or vertical polarizations of a photon. Mathematically, such a system is described by a two dimensional normalized complex vector $|\psi\rangle \in \mathbb{C}^2$. This can be written $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ where $|0\rangle$ and $|1\rangle$ are orthonormal basis states (corresponding to e.g. spin up or down). Since the state is normalized, $|\alpha|^2 + |\beta|^2 = 1$.

The term *qudit* refers to a quantum system with d orthogonal states, for some $d \geq 2$. We will only consider finite dimensional systems. For example, an atom that may be in d different Rydberg states is a qudit, as is a particle that can be in d different spacial positions. In all cases, these systems are allowed to be in a superposition of these basis states, so the state can be any vector in $\mathbb{C}^d$.

In this dissertation we will not have occasion to consider the physical realizations of qubits or qudits. Instead we deal with these as abstract concepts.

1.3.3 Tensor product

Multiple subsystems (qubits or qudits) exist within a vector space (a Hilbert space) that is a tensor product of the spaces associated with each subsystem. For example, if one qudit is in state $|\psi\rangle \in \mathbb{C}^{d_1}$ and another is in state $|\phi\rangle \in \mathbb{C}^{d_2}$ then the joint system is in state $|\psi\rangle \otimes |\phi\rangle \in \mathbb{C}^{d_1 d_2}$. States of the form $|\psi\rangle \otimes |\phi\rangle$ are *product states* and those not of this form are *entangled*. We will explore this further in section 1.3.7. The standard basis for a tensor product space $A \otimes B$ consists of the vectors $\{|i\rangle \otimes |j\rangle\}_{ij}$ where $\{|i\rangle\}_i$ is the standard basis of A and $\{|j\rangle\}_j$ is the standard basis of B.

Since the joint state of two qubits exists within a four dimensional space, it is possible to consider the pair as a ququart (qudit with $d = 4$). Likewise, ten qubits can be considered as a qudit with $d = 2^{10}$. The joint state of a hundred qubits is a vector of such high dimension that storing the coefficients of such a vector on a computer would not be possible. For this reason naïve simulation of quantum systems on a classical computer is not possible for more than a handful of qubits, although in some situations special tricks are available to simulate such systems. Such a technique will be investigated in chapter 3.

1.3.4 Unitary operations

The Schrödinger equation produces a unitary time evolution in a closed system. Quantum computing is founded on the assumption that one day it will be feasible to perform arbitrary unitary operations on any qubit or pair of qubits (or even qudits), with a very high degree of accuracy. This will likely require a very complicated process involving several layers of error correction, but this is outside the scope of this dissertation.

Consider three qubits, on vector spaces A, B, and C. Their joint state will be some $|\psi\rangle \in A \otimes B \otimes C$. A unitary U_A acting on only the first of these will have action $(U_A \otimes I_B \otimes I_C)|\psi\rangle$ on the joint state, where I_B is the identity matrix on space B (and similarly for I_C). A unitary U_{AB} acting jointly on qubits A and B will have action $(U_{AB} \otimes I_C)|\psi\rangle$.

Arbitrary unitaries acting on an arbitrary number of qubits can always be composed from a sequence of unitaries acting on one or two qubits. It may be that this sequence of operations is extremely long. Sometimes we will be concerned only with whether a certain task is possible, not with the level of difficulty. In these cases the assumption is that any arbitrary unitary operation on any arbitrary number of qubits is possible.

1.3.5 Measurements

The most basic type of measurement is the *projective measurement*. The outcomes are labeled by projectors that sum to the identity, $\sum_x P_x = I$. Such a collection is known as a *projective decomposition of the identity*. When a state $|\psi\rangle$ is measured, outcome x will occur with probability $p_x = \langle\psi|P_x|\psi\rangle$ (the Born rule). Since $\sum_x P_x = I$, these probabilities add up to 1.

Typically, one assumes that the particles being measured are consumed by the measuring device. One may also consider a *non-destructive measurement* in which the particles being measured are not consumed. In general, it is not possible to perform a measurement without disturbing a system. If the outcome of a non-destructive measurement is x, the system will be left in the state $P_x|\psi\rangle/p_x$ where again $p_x = \langle\psi|P_x|\psi\rangle$.

A more general form of (destructive) measurement is the *positive-operator valued measurement* (POVM). Here, the outcomes are labeled by positive semidefinite (PSD) operators that sum to the identity, $\sum_x M_x = I$. Again, outcome x occurs with probability $\langle\psi|M_x|\psi\rangle$. Physically, a POVM can be achieved by introducing some number of ancillary qubits all in the $|0\rangle$ state, and doing a projective measurement on the joint system.

1.3.6 Density operators

So far we have been discussing what are known as *pure states*, which can be thought of as states of maximal knowledge. But suppose we only know that a system was in state $|\psi\rangle$ with probability p_ψ or state $|\phi\rangle$ with probability p_ϕ. This is known as a *mixed state*. Consider a POVM measurement $\{M_x\}_x$. If the system is in state $|\psi\rangle$ then outcome x occurs with probability $\langle\psi|M_x|\psi\rangle$. Given that we don't know what state the system is in, we can only say that outcome x occurs with probability

$$p_\psi \langle\psi|M_x|\psi\rangle + p_\phi \langle\phi|M_x|\phi\rangle . \tag{1.1}$$

There is a more succinct way to write this. Defining $\rho = p_\psi|\psi\rangle\langle\psi| + p_\phi|\phi\rangle\langle\phi|$, we may write (1.1) as $\mathrm{Tr}\{M_x\rho\}$. Note that ρ is an operator (a matrix). So if $|\psi\rangle \in A$ then $\rho \in \mathcal{L}(A)$ where $\mathcal{L}(A)$ is the set of linear operators on Hilbert space A.

The quantity ρ is known as a *density operator*. Density operators are always PSD and have trace 1. Conversely, any PSD operator with trace 1 is a density operator. In other words, if $\rho \succeq 0$ and $\mathrm{Tr}(\rho) = 1$ there is some probability distribution $\{p_i\}$ and some collection of states $\{|\psi_i\rangle\}$ such that $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$. This is just an eigenvector decomposition. If some of the eigenvalues are degenerate (i.e. $p_i = p_j$) then this decomposition is not unique. For instance, the qubit density operator $\rho = I/2$ could correspond either to an even (50% of each) mixture of $|0\rangle$ and $|1\rangle$ states or to a mixture of $|+\rangle = 2^{-1/2}(|0\rangle + |1\rangle)$ and $|-\rangle = 2^{-1/2}(|0\rangle - |1\rangle)$ states. There is no way to distinguish between the two cases; the density operator ρ represents all of the physically available information about a system.

Suppose that given a joint system in state $\rho \in \mathcal{L}(A) \otimes \mathcal{L}(B)$ we measure only subsystem A, using a POVM $\{M_x\}_x \subset \mathcal{L}(A)$. Outcome x will occur with probability

$$p_x = \mathrm{Tr}\{(M_x \otimes I_B)\rho\}. \tag{1.2}$$

In the process of the measurement, subsystem A is consumed. Conditioned on outcome x, subsystem B will remain in the *conditional state*

$$\rho_x = \mathrm{Tr}_A\{(M_x \otimes I_B)\rho\}/p_x \tag{1.3}$$

where Tr_A denotes partial trace over space A.[1]

Suppose we forget which outcome x occurred. We will only know that we have state ρ_x with probability x. In other words, we will be left with state $\rho' = \sum_x p_x \rho_x$. Plugging in (1.3) yields $\rho' = \mathrm{Tr}_A(\rho)$. Notice this does not depend on the choice of measurement $\{M_x\}_x$. Measuring a subsystem and forgetting the outcome is equivalent to just discarding that subsystem. The operation $\mathrm{Tr}_A(\rho)$ is referred to as *tracing out* A, with the physical interpretation being that subsystem A has been discarded or is somehow no longer available. For example, suppose Alice is in possession of subsystem A and Bob is in possession of subsystem B, with joint state ρ. From Alice's perspective, she has access to state $\rho_A = \mathrm{Tr}_B(\rho)$. The state ρ_A contains all of the information needed to compute the outcome probabilities for any measurement that Alice could do without Bob's cooperation.

Any density operator can be *purified*. That is to say, if $\rho \in \mathcal{L}(A)$ is a density operator then there is some space B, whose dimension is at most that of A, and a pure state $|\psi\rangle \in A \otimes B$, such that $\rho = \mathrm{Tr}_B\{|\psi\rangle\langle\psi|\}$. The purification is unique up to the application of an arbitrary unitary operator on system B. In other words, $|\psi'\rangle$ is a purification of ρ if and only if there is a unitary $U \in \mathcal{L}(B)$ such that $|\psi'\rangle = (I \otimes U)|\psi\rangle$.

Density operators can be considered as a quantum generalization of classical probability distributions. Specifically, restricting to density operators that are diagonal matrices exactly recovers classical probability theory, with the numbers down the diagonal corresponding to the probabilities. For example, suppose $\rho \in \mathcal{L}(A) \otimes \mathcal{L}(B)$ is given by $\rho = \sum_{ij} p_{ij} |i\rangle\langle i| \otimes |j\rangle\langle j|$. Then p_{ij} is a joint probability distribution since $p_{ij} \geq 0$ and $\sum_{ij} p_{ij} = \mathrm{Tr}(\rho) = 1$. Discarding the first subsystem yields the reduced distribution $\sum_i p_{ij}$, corresponding to the diagonal entries of $\mathrm{Tr}_A(\rho)$. Measuring the state of the first subsystem corresponds to measurement $M_i = |i\rangle\langle i|$ where $|i\rangle$ are the basis states. Then (1.2) gives $p_i = \sum_j p_{ij}$, and (1.3) gives the post-measurement state $p_{j|i} = p_{ij}/p_i$, just as in classical probability theory.

1.3.7 Entanglement, Schmidt rank, and separable states

As mentioned previously, a pure state of the form $|\psi\rangle \otimes |\phi\rangle$ is known as a product state, and states on two subsystems that are not of this form are entangled. Entangled states are in some situations considered a valuable resource, since they may be difficult to create and may grant important abilities. For example, Alice and Bob cannot create an entangled state if their only contact is through a classical channel (e.g. a telephone). On the other hand, if they are already in possession of an entangled state then Alice can transmit quantum states to Bob using only a classical channel. This is known as *teleportation*, and will be discussed in section 1.3.13.

Any pure state on two subsystems, $|\psi\rangle \in A \otimes B$, can be written as a linear combination of product states,

$$|\psi\rangle = \sum_i s_i |u_i\rangle \otimes |v_i\rangle . \tag{1.4}$$

Furthermore, there exists such a decomposition where $s_i > 0$, the $|u_i\rangle$ are orthonormal vectors on A, and the $|v_i\rangle$ are orthonormal vectors on B. This is known as a *Schmidt decomposition*. The *Schmidt coefficients* s_i satisfy $\sum_i s_i^2 = 1$. The number of Schmidt coefficients is the *Schmidt rank* of $|\psi\rangle$. Entangled states are those with Schmidt rank at least 2. States whose Schmidt rank r satisfies $r = \dim(A) = \dim(B)$ and whose Schmidt coefficients are all equal are *maximally entangled*. Note that this requires $s_i = \sqrt{r}$. The Schmidt decomposition for bipartite states is analogous to the singular value decomposition for operators.

The amount of entanglement for a pure state can be quantified in terms of the Von Neumann

[1] The partial trace of a product operator $\sigma \otimes \chi \in \mathcal{L}(A) \otimes \mathcal{L}(B)$ is defined as $\mathrm{Tr}_A(\sigma \otimes \chi) = \mathrm{Tr}(\sigma)\chi$. The definition is extended to general operators by linearity.

entropy of the reduced density operator,

$$S(\mathrm{Tr}_B(|\psi\rangle\langle\psi|)) = -\sum_i s_i \log s_i. \tag{1.5}$$

Here, and throughout this dissertation, logarithms are taken base 2. A maximally entangled state of dimension r has $\log r$ bits of Von Neumann entropy (termed *ebits*).

The theory of entanglement for mixed states is more complicated, but we will only be concerned with three classes of states. We focus on two subsystems, although the multipartite generalization is straightforward. The product states are those of the form $\rho \otimes \sigma$. The *separable states* are those of the form $\sum_i \rho_i \otimes \sigma_i$ where $\rho_i \succeq 0$ and $\sigma_i \succeq 0$ for all i. The set of all separable states is known as SEP. (Later we will discuss separable operations, also referred to as SEP. This notational ambiguity is resolved by considering the context.) It is computationally difficult to determine whether a mixed state is separable. The entangled states are those that are not separable.

Among the states whose density operators are diagonal in the standard basis, the product states correspond to uncorrelated classical probability distributions and the non-product separable states correspond to correlated probability distributions. There are no entangled states whose density operators are diagonal, so entanglement is a purely quantum phenomenon.

1.3.8 Classical channels

A *discrete memoryless channel* (which we will abbreviate as just *channel*, or *classical channel*), denoted $\mathcal{E} : A \to B$ where A and B are finite sets, takes input $a \in A$ and produces output $b \in B$ with conditional probability $\mathcal{E}(b|a)$. This is to be thought of as a formalization of the concept of a communication channel such as a copper telephone line, although realistically such a channel can be considered discrete and memoryless only in an approximate sense.

For a given channel, one may define the matrix with entries $E_{ba} = \mathcal{E}(b|a)$. Since $\mathcal{E}(b|a)$ is a conditional probability distribution, E is a stochastic matrix—it consists of nonnegative entries and each column sums to 1. Composition of channels corresponds to multiplication of these matrices:

$$(\mathcal{F} \circ \mathcal{E})(c|a) = \sum_b \mathcal{F}(c|b)\mathcal{E}(b|a) = (FE)_{ca}. \tag{1.6}$$

The *perfect channel* or *identity channel* is $\mathcal{I}(b|a) = \delta_{ba}$. Here δ is the Kronecker delta which takes value 1 when $b = a$ and zero otherwise. The corresponding stochastic matrix is the identity matrix.

1.3.9 Superoperators and quantum channels

Since a unitary operator acts on a pure state by the action $|\psi\rangle \to U|\psi\rangle$, it acts on a density operator by $\rho \to U\rho U^\dagger$. More general transformations are possible by introducing ancillary qubits (that are uncorrelated with the input), doing a unitary transformation, and then discarding a (possibly different) set of ancillaries. Such an operation is known as a *discrete memoryless quantum channel*, and may map density operators on one space, say $\mathcal{L}(A)$, to another space, say $\mathcal{L}(B)$. Since all channels considered in this dissertation will be discrete and memoryless, these will simply be called *quantum channels*, or even just *channels*.

If $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ is a quantum channel then there is an ancillary space C of dimension at most $\dim(A)\dim(B)$ and an isometry $J : A \to B \otimes C$, known as a *Stinespring isometry*, such that

$$\mathcal{E}(\rho) = \mathrm{Tr}_C\{J\rho J^\dagger\}. \tag{1.7}$$

The Stinespring isometry is not unique, although any two such isometries are related by a unitary transform on the ancillary space (as long as they both use the same dimension of ancillary).

It is useful to consider the *Kraus operators* $K_i : A \to B$ defined by $K_i = (I_B \otimes \langle i|_C)J$ for $i \in \{1, \ldots, \dim(C)\}$. Note that since the Stinespring isometry is not unique, neither are the Kraus operators. With these, (1.7) can be written

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger \tag{1.8}$$

Because J is an isometry, the Kraus operators satisfy the *closure condition*

$$\sum_i K_i^\dagger K_i = I. \tag{1.9}$$

In general, a linear transformation on the space of operators, $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$, is known as a *superoperator*. The tensor product of two superoperators has the action $(\mathcal{E} \otimes \mathcal{F})(\rho \otimes \sigma) = \mathcal{E}(\rho) \otimes \mathcal{F}(\sigma)$; this is extended by linearity to non-product states. In particular, we will have occasion to consider the tensor product of $\mathcal{E}$ with the identity channel, $(\mathcal{E} \otimes \mathcal{I})(\rho \otimes \sigma) = \mathcal{E}(\rho) \otimes \sigma$.

A superoperator is *positive* if it maps PSD operators to PSD operators. It is *completely positive* if it remains positive when tensored with an identity channel of arbitrary dimension, i.e. $\mathcal{E} \otimes \mathcal{I}$ is positive. By *Choi's theorem on completely positive maps*, completely positive superoperators correspond exactly to those of the form (1.8). A superoperator is *trace preserving* if $\mathrm{Tr}\{\mathcal{E}(\rho)\} = \mathrm{Tr}(\rho)$ for all ρ. The set of superoperators that are both completely positive and trace preserving (CPTP) corresponds exactly to the set of quantum channels. In other words, a superoperator is CPTP if and only if it is of the form (1.7) or, equivalently, of the form (1.8)-(1.9).

Again it is useful to relate to the classical case. Recall that density operators which are diagonal in the standard basis correspond to classical probability distributions. Let $\mathcal{E}$ be a quantum channel that maps diagonal matrices to diagonal matrices. The diagonal entries of a density matrix ρ, written $\mathrm{diag}(\rho)$, may be regarded as a vector (and interpreted as a probability distribution). The superoperator $\mathcal{E}$ may then be regarded as an operator E acting on such vectors, $E\,\mathrm{diag}(\rho) = \mathrm{diag}(\mathcal{E}(\rho))$. This operator is a stochastic matrix: since $\mathcal{E}$ is positive, E has non-negative entries, and since $\mathcal{E}$ is trace preserving, the columns of E each sum to 1. So the classical analogue of a CPTP map is a classical noisy channel.

Three special channels deserve mention. The *perfect quantum channel* is just the identity map: $\mathcal{E}(\rho) = \rho$. The *perfect classical channel* preserves the diagonal entries of its input (in the standard basis) but zeros out the off diagonal entries. The Kraus operators for this channel are $K_i = |i\rangle\langle i|$; the Stinespring isometry is $J = \sum_i (|i\rangle \otimes |i\rangle)\langle i|$. The *completely noisy channel* passes no information about its input: $\mathcal{E}(\rho) = |0\rangle\langle 0|$. The Kraus operators are $K_i = |0\rangle\langle i|$.

Every channel $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ has an associated *complementary channel* $\widehat{\mathcal{E}} : \mathcal{L}(A) \to \mathcal{L}(C)$. Let $J : A \to B \otimes C$ be a Stinespring isometry for $\mathcal{E}$, so that $\mathcal{E}(\rho) = \mathrm{Tr}_C\{J\rho J^\dagger\}$. The complementary channel is $\widehat{\mathcal{E}}(\rho) = \mathrm{Tr}_B\{J\rho J^\dagger\}$. In other words, the roles of B and C are switched. Since the Stinespring isometry is not unique, neither is the complementary channel; one may vary the output C by an arbitrary unitary. One may think of B as the receiver and C as the environment, so that $\widehat{\mathcal{E}}$ is in some sense the channel to the environment. The complementary channel of $\widehat{\mathcal{E}}$ is $\mathcal{E}$, modulo a unitary on B.

For example, the complement of the perfect quantum channel is the completely noisy channel and the complement of the completely noisy channel is the perfect quantum channel. The perfect classical channel is its own complement.

Every channel (indeed, every superoperator) $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ has an adjoint $\widehat{\mathcal{E}}^* : \mathcal{L}(B) \to \mathcal{L}(A)$. By definition, the adjoint satisfies $\mathrm{Tr}\{\widehat{\mathcal{E}}^*(\rho)^\dagger \sigma\} = \mathrm{Tr}\{\rho^\dagger \mathcal{E}(\sigma)\}$. If a completely positive map has Kraus operators $\{K_i\}$, its adjoint has Kraus operators $\{K_i^\dagger\}$. The adjoint of a channel is not necessarily a channel since it might not be trace preserving.

Although the ancillary C was discarded (dumped to the environment) to arrive at (1.7), it is sometimes useful to consider a scenario in which C is measured in the standard basis. Measurement

outcome i will occur with probability

$$p_i = \mathrm{Tr}\{K_i \rho K_i^\dagger\}. \tag{1.10}$$

Conditioned on outcome i, the input state $\rho \in \mathcal{A}$ will transform into the *conditional density operator* $\rho_i \in \mathcal{B}$ of the form

$$\rho_i = K_i \rho K_i^\dagger / p_i. \tag{1.11}$$

If the outcome i is then forgotten, this reduces to (1.8).

1.3.10 Graphs

A *graph* G consists of a set of *vertices* $V(G)$ along with a set of *edges* between vertices. We write $x \sim_G y$ when there is an edge between x and y, and say that x and y are *adjacent*. We abbreviate $x \sim y$ when the graph can be inferred from context. An edge from a vertex to itself, $x \sim x$, is called a *loop*. In chapter 4 only graphs without loops (*simple graphs*) will be considered; in chapter 5 we will have occasion to allow loops (*loop graphs*).

The *complement* of a graph G, denoted $\overline{G}$, has the same set of vertices but has edges between distinct pairs of vertices which are not adjacent in G. I.e.

$$x \sim_{\overline{G}} y \iff x \not\sim_G y \text{ and } x \neq y. \tag{1.12}$$

For loop graphs the complement is typically defined as $x \sim_{\overline{G}} y \iff x \not\sim_G y$. Fortunately, in this dissertation we will only consider the complement of loop graphs which have loops on every vertex, and in this case the two conventions coincide.

A set of vertices no two of which are adjacent is known as an *independent set*; the size of the largest independent set is the *independence number* $\alpha(G)$. A set of vertices such that every pair is adjacent is known as a *clique*; the size of the largest clique is the *clique number* $\omega(G)$. Clearly $\omega(G) = \alpha(\overline{G})$. An assignment of colors to vertices such that adjacent vertices are given distinct colors is called a *proper coloring*; the minimum number of colors needed is the *chromatic number* $\chi(G)$.

A function mapping the vertices of one graph to those of another, $f : V(G) \to V(H)$, is a *homomorphism* if $x \sim_G y \implies f(x) \sim_H f(y)$. In general f could map two different vertices of G to the same vertex of H, but note that since vertices are not adjacent to themselves in a simple graph, (1.12) gives $f(x) \neq f(y)$ when $x \sim_G y$. If such a function exists, we say that G *is homomorphic to* H and write $G \to H$.

The *complete graph* on n vertices, denoted K_n, has an edge between every pair of vertices. It is not hard to see that $\omega(G)$ is equal to the largest n such that $K_n \to G$, and $\chi(G)$ is equal to the smallest n such that $G \to K_n$. Many other graph properties can be expressed in terms of homomorphisms; for details see [HT97, HN04].

The *strong product* of two graphs, $G \boxtimes H$, has vertex set $V(G) \times V(H)$ and has edges

$$(x_1, y_1) \sim (x_2, y_2) \iff (x_1 = x_2 \text{ and } y_1 \sim y_2) \text{ or} \tag{1.13}$$
$$(x_1 \sim x_2 \text{ and } y_1 = y_2) \text{ or} \tag{1.14}$$
$$(x_1 \sim x_2 \text{ and } y_1 \sim y_2). \tag{1.15}$$

The n-fold strong product is written $G^{\boxtimes n} := G \boxtimes G \boxtimes \cdots \boxtimes G$. The *disjunctive product* $G * H$ has edges

$$(x_1, y_1) \sim (x_2, y_2) \iff x_1 \sim x_2 \text{ or } y_1 \sim y_2. \tag{1.16}$$

It is easy to see that $\overline{G * H} = \overline{G} \boxtimes \overline{H}$. The n-fold disjunctive product is written $G^{*n} := G * G * \ldots * G$.

The clique number and chromatic number are difficult to compute, or even to approximate, unless P = NP; however, there is an efficiently computable quantity that is sandwiched between these: the Lovász ϑ number. It will be more convenient to instead work with the Lovász number of the complement of a graph: $\bar{\vartheta}(G) = \vartheta(\overline{G})$. There are many equivalent formulations for this quantity [Lov79, Knu94, Lov03, dCST13]; the ones most relevant to this dissertation are[2]

$$\bar{\vartheta}(G) = \max\{\|I + T\| : I + T \succeq 0, T_{ij} = 0 \text{ for } i \not\sim j\} \tag{1.17}$$

$$\bar{\vartheta}(G) = \max\left\{\sum_{ij} B_{ij} : B \succeq 0, \mathrm{Tr}B = 1, B_{ij} = 0 \text{ for } i \not\sim j, i \neq j\right\} \tag{1.18}$$

$$\bar{\vartheta}(G) = \min\{\lambda : \exists Z \succeq 0, Z_{ii} = \lambda - 1, Z_{ij} = -1 \text{ for } i \sim j\}. \tag{1.19}$$

The variables T, B, and Z are Hermitian matrices. They may be taken to be real or complex, as this choice doesn't affect the optimal value. It is not hard to see that $\bar{\vartheta}(K_n) = n$. The $\bar{\vartheta}$ number is monotone under graph homomorphisms in the sense that $G \to H \implies \bar{\vartheta}(G) \leq \bar{\vartheta}(H)$ [dCST13]. Consequently, we have $n = \omega(G) \implies K_n \to G \implies n \leq \bar{\vartheta}(G)$. Similarly, $n = \chi(G) \implies G \to K_n \implies \bar{\vartheta}(G) \leq n$. This gives the *Lovász sandwich theorem*

$$\omega(G) \leq \bar{\vartheta}(G) \leq \chi(G). \tag{1.20}$$

Two related quantities, $\bar{\vartheta}'$ due to Schrijver and $\bar{\vartheta}^+$ due to Szegedy (see definition 4.5) are also homomorphism monotones and so satisfy similar sandwich theorems. These are no greater and no less, respectively, than $\bar{\vartheta}$:

$$\omega(G) \leq \bar{\vartheta}'(G) \leq \bar{\vartheta}(G) \leq \bar{\vartheta}^+(G) \leq \chi(G). \tag{1.21}$$

An important property of ϑ is that is is multiplicative under the strong and the disjunctive graph products:

$$\bar{\vartheta}(G \boxtimes H) = \bar{\vartheta}(G)\bar{\vartheta}(H) \tag{1.22}$$

$$\bar{\vartheta}(G * H) = \bar{\vartheta}(G)\bar{\vartheta}(H). \tag{1.23}$$

The $\bar{\vartheta}^+$ number is not multiplicative under strong product and the $\bar{\vartheta}'$ number is not multiplicative under the disjunctive product, but $\bar{\vartheta}'$ is multiplicative under the strong product (see appendix 4.A).

1.3.11 Zero-error classical channel capacity

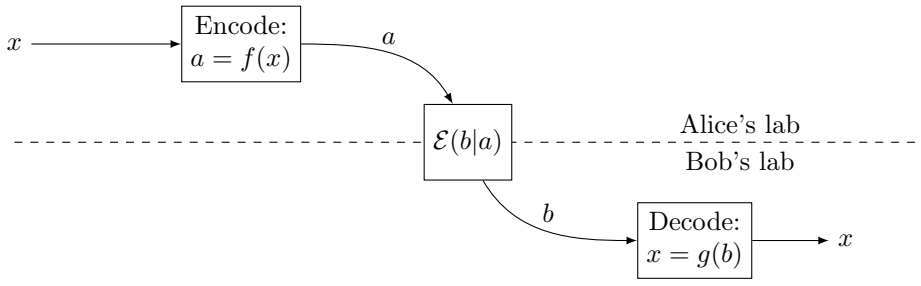


Figure 1.1: Zero-error channel coding.

[2] The first of these follows from theorem 6 of [Lov79] by setting $T = A/|\lambda_n(A)|$ (note that in [Lov79] vertices are considered adjacent to themselves). The second is theorem 4 of [Lov79]. The third comes from page 167 of [Lov03], or from theorem 3 of [Lov79] by taking $Z = \lambda I - A$ with λ being the maximum eigenvalue of A.

The capacity of a channel is the number of bits that can be sent, per channel use, in the limit of many parallel uses of a channel. The entire collection of bits to be sent is encoded in some way, then sent through the many parallel channels, then decoded. The capacity is taken in the limit of infinitely many parallel channels, and the probability of faithfully recovering the original message must approach 1. Rather than several parallel instances of a channel, one may consider a single channel that is used many times in sequence. Indeed, this is the more applicable scenario. The mathematical formalism is exactly the same.

The zero-error capacity of a channel is defined similarly, except that the probability of faithfully recovering the message must be *exactly* 1. We distinguish between *one-shot capacity*, which can be achieved with a single use of a channel, and *asymptotic capacity*, which is achieved in the limit of an infinite number of parallel channels.

For classical channels, this problem was first considered by Shannon in [Sha56]. Let $\mathcal{E} : A \to B$ be a noisy classical channel. Alice, who wishes to send a value x, encodes x using some encoding scheme agreed to ahead of time, and passes the encoded value $a \in A$ into the channel. Since the channel is noisy, Bob receives some value $b \in B$ which is possibly different from a. The probability of receiving b given input a is denoted $\mathcal{E}(b|a)$. Bob decodes the received value b to recover the value x. This protocol is schematically represented in fig. 1.1.

Codewords a and $a' \in A$ can be unambiguously distinguished by Bob if and only if they cannot map to the same output value. That is to say, if $\mathcal{E}(b|a)\mathcal{E}(b|a') = 0$ for all $b \in B$. This motivates definition of the *distinguishability graph* of $\mathcal{E}$, having vertex set $V(G) = A$ and edges

$$a \sim a' \iff \mathcal{E}(b|a)\mathcal{E}(b|a') = 0 \text{ for all } b \in B. \tag{1.24}$$

Two codewords are then unambiguously distinguishable if and only if they are adjacent in this graph. The largest set of such codewords is $\omega(G)$, so the number of bits that can be sent with a single channel use is $\log \omega(G)$.

The more traditional construct is the *confusability graph* with edges $a \sim a' \iff (\mathcal{E}(b|a)\mathcal{E}(b|a') \neq 0$ for some $b \in B)$, of which (1.24) is the complement. The distinguishability graph turns out to be more convenient for our purposes.

It is not too hard to see that the distinguishability graph corresponding to n parallel uses of the channel is G^{*n}. Briefly, two codewords $(a_1, \ldots, a_n)$ and $(a'_1, \ldots, a'_n)$ are distinguishable when a_i and a'_i are distinguishable for some i. So, the maximum number of codewords that can be sent is $\omega(G^{*n})$, corresponding to $n^{-1} \log \omega(G^{*n}) = \log \sqrt[n]{\omega(G^{*n})}$ bits per channel use. One can now take the limit $n \to \infty$. For convenience, the logarithm is dropped in this expression, and we define

$$\bar{\Theta}(G) = \lim_{n\to\infty} \sqrt[n]{\omega(G^{*n})}. \tag{1.25}$$

So $\log \bar{\Theta}(G)$ is the asymptotic capacity of the channel. This is the *Shannon capacity* of $\overline{G}$. (Shannon used the symbol Θ for his capacity. Since we use the distinguishability graph rather than the confusability graph as Shannon used, it is natural to use $\bar{\Theta}(G) = \Theta(\overline{G})$.)

Shannon [Sha56] considered the example of a channel which takes as input the symbols $\{1, 2, 3, 4, 5\}$ and produces as output either a faithful rendition of the input, or a version of the input corrupted by adding 1 modulo 5. For clarity, consider the output symbols to be primed: $\{1', 2', 3', 4', 5'\}$. The distinguishability graph of this channel is the five cycle (pentagon) C_5. For example, since inputs 1 and 2 can both map to output $2'$, they are not distinguishable and so have no edge in the distinguishability graph. On the other hand, 1 can map to $1'$ or $2'$, and 3 maps to $3'$ or $4'$, so inputs 1 and 3 are distinguishable; there is therefore an edge between 1 and 3 in the distinguishability graph. This is illustrated in fig. 1.2.

The one-shot capacity of this channel is $\omega(C_5) = 2$, the codewords $\{1, 3\}$ being distinguishable. For two uses of the channel we have $\omega(C_5^{*2}) = 5$, the codewords $\{(1, 1), (2, 3), (3, 5), (4, 2), (5, 4)\}$ being distinguishable. Since the limit in (1.25) can easily be shown to be at least as large as its

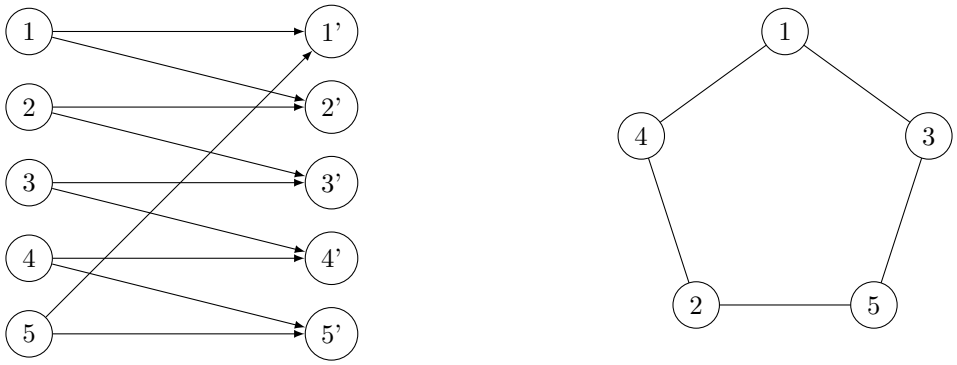


Figure 1.2: The possible channel outputs for each input (left) and the distinguishability graph (right).

argument for any particular n, it is clear that $\bar{\Theta}(G) \geq \sqrt{5}$. The question of equality here remained an open question for over two decades until Lovász defined his ϑ number. Since $\omega(G) \leq \bar{\vartheta}(G)$, we get

$$\bar{\Theta}(G) \leq \lim_{n\to\infty} \sqrt[n]{\bar{\vartheta}(G^{*n})} = \lim_{n\to\infty} \sqrt[n]{\bar{\vartheta}(G)^n} = \bar{\vartheta}(G). \tag{1.26}$$

So $\bar{\Theta}(C_5) \leq \bar{\vartheta}(C_5) = \sqrt{5}$ [Lov79].

1.3.12 Zero-error quantum channel capacity

Again we will be concerned only with the zero-error capacity. In the case of a quantum channel $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$, two input codewords $|\psi_x\rangle, |\psi_y\rangle \in A$ can be unambiguously distinguished if and only if they map to orthogonal states. Here, orthogonality is with respect to the Hilbert–Schmidt inner product $\langle X, Y\rangle = \mathrm{Tr}(X^\dagger Y)$. With $K_i : A \to B$ being the Kraus operators for $\mathcal{E}$, the condition for distinguishability can be expanded

$$0 = \langle \mathcal{E}(|\psi_x\rangle\langle\psi_x|), \mathcal{E}(|\psi_y\rangle\langle\psi_y|)\rangle \tag{1.27}$$

$$= \mathrm{Tr}\left\{ \left(\sum_i K_i |\psi_x\rangle\langle\psi_x| K_i^\dagger \right) \left(\sum_j K_j |\psi_y\rangle\langle\psi_y| K_j^\dagger \right) \right\} \tag{1.28}$$

$$= \sum_{ij} \left| \left\langle \psi_x \middle| K_i^\dagger K_j \middle| \psi_y \right\rangle \right|^2 . \tag{1.29}$$

Since each term here is nonnegative, they all must vanish:

$$\left\langle \psi_x \middle| K_i^\dagger K_j \middle| \psi_y \right\rangle = 0 \text{ for all } i, j. \tag{1.30}$$

So the maximum set of codewords that can be unambiguously distinguished is the largest set of states such that each pair satisfies (1.30). The *one-shot zero-error capacity* is the log of the size of this set.

There is a useful intuition regarding (1.30). The action $\mathcal{E}(|\psi\rangle\langle\psi|) = \sum_i K_i|\psi\rangle\langle\psi|K_i^\dagger$ can be interpreted in the following way: the state $|\psi\rangle$ gets acted upon by some Kraus operator K_i, but we don't know which one. This follows from plugging $\rho = |\psi\rangle\langle\psi|$ into (1.11). So the receiver knows that codeword $|\psi_x\rangle$ will map to $K_i|\psi_x\rangle$ (disregarding normalization) for some i, and $|\psi_y\rangle$ will map to $K_j|\psi_y\rangle$ for some j. These two sets can be distinguished if and only if all of the $K_i|\psi_x\rangle$ are orthogonal to all of the $K_j|\psi_y\rangle$, which is just a restatement of (1.30).

Whereas the preceding discussion was regarding the transmission of classical information (the codewords) through a channel, we will also be interested in transmitting quantum information—qubits rather than bits. As before, the sender encodes the state in some way before sending through the channel. It can be shown that a qudit of dimension d (or $\log d$ qubits) can be perfectly transmitted if and only if there is a subspace $Q \subseteq A$ such that for all $|\psi\rangle, |\phi\rangle \in Q$ with $\langle\psi|\phi\rangle = 0$ we have

$$\left\langle \psi \middle| K_i^\dagger K_j \middle| \phi \right\rangle = 0 \text{ for all } i, j. \tag{1.31}$$

This is (one form of) the *Knill–Laflamme error correction condition*, and follows from theorem III.2 of [KL97]. The *one-shot zero-error quantum capacity* is the maximum possible value of $\log \dim(Q)$.

As was the case with classical channels, it is important to consider the asymptotic capacity—the average number of bits or qubits that can be sent per channel use as the number of parallel instances of a channel goes to infinity. With n parallel channels, the composite channel is $\mathcal{E}^{\otimes n}$ and the codewords (for classical capacity) or the subspace Q (for quantum capacity) are on $A^{\otimes n}$.

For both classical and quantum channels, one may also consider the zero-error *entanglement assisted* capacity (one-shot or asymptotic). In this scenario, sender and receiver make use of an arbitrary entangled state which they have set up ahead of time. The math in this case is slightly more complicated, and will be derived when the need arises in chapters 4 and 5.

1.3.13 Dense coding and teleportation

Dense coding (also known as *superdense coding*) is a method for Alice to convey two classical bits (i.e. a number from 1 to 4) to Bob by transmitting one qubit. To carry out the protocol, Alice and Bob must be in prior possession of a maximally entangled state. The protocol goes as follows: Alice performs one of four unitaries on the qubit that makes up her half of the entangled state, and sends the resulting qubit to Bob over a perfect quantum channel. Bob performs a joint measurement on this qubit and on his half of the entangled state, getting one of four outcomes. This measurement outcome tells him which of the four unitaries Alice used, thus conveying two bits of classical information. The four unitaries must be orthogonal under the Hilbert–Schmidt inner product $\mathrm{Tr}(U^\dagger V)$. The canonical choice consists of the four Pauli matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.32}$$

Teleportation is in some sense the reverse of dense coding: Alice sends one qubit to Bob by transmitting two classical bits. Again, a maximally entangled state is consumed. The protocol goes as follows. Alice performs a joint measurement on the qubit she wishes to send and on her half of the entanglement resource. This measurement has four outcomes, or two classical bits. She transmits these two bits to Bob, who then performs one of four unitary operations (typically the Pauli matrices listed above) on his half of the entanglement resource. Bob's half of the entanglement resource is now in the same state as the qubit that Alice originally had. Alice's qubit was destroyed by her measurement.

Dense coding and teleportation are useful subroutines for communication protocols. For example, if the entanglement assisted classical capacity (one-shot or asymptotic; zero-error or not) of a noisy channel is 6 bits, then it must be that its entanglement assisted quantum capacity is at least 3 qubits since teleportation can be used to send 3 qubits using 6 bits of classical communication. Conversely, the quantum capacity cannot be more than 3 qubits since otherwise dense coding could be used to send more than 6 classical bits.

1.3.14 LOCC and SEP

Local operations and classical communication (LOCC) is the class of quantum transformations that can be achieved by cooperating parties who are physically separated but can communicate over a

classical channel (e.g. a telephone).

An LOCC protocol proceeds as follows. Alice and Bob (for simplicity we assume only two parties) each possess a subsystem of a possibly entangled input state. Alice performs a quantum operation of the form (1.10)-(1.11) on her subsystem, resulting in a transformation of her quantum subsystem and generation of a measurement outcome i. She transmits the classical information i to Bob. Depending on the value i, Bob will choose some quantum operation to perform on his subsystem. Without loss of generality this can be chosen deterministically; any random choice can be absorbed into the quantum operation itself. This operation will result in a measurement outcome j, which Bob transmits to Alice. Alice performs some operation depending upon the values of i and j, resulting in a measurement outcome k which she transmits to Bob. Alice and Bob may proceed for many rounds, adapting their strategy depending on the entire record of messages that have been passed back and forth during prior rounds. When they are finished communicating, they may each perform some final post-processing (by a transformation of the form (1.8)-(1.9)) on their subsystem to produce their final output states.

Of special note is *one-way LOCC* (LOCC-1) in which Alice sends a message to Bob, but Bob is not allowed to send a message to Alice. The teleportation protocol of section 1.3.13 is of this form.

The mathematical structure of LOCC is difficult to work with, is not very well understood, and is notationally tedious. A simpler class of operations, which includes LOCC as a special case, is the set of *separable operations* (SEP). The term SEP also refers to separable states, but it is possible to determine which concept is being referred to by context. The separable operations are those of the form (1.8)-(1.9) with the added restriction that each K_i is a product operator: $K_i = E_i \otimes F_i$. Then (1.8)-(1.9) take the form

$$\rho \to \sum_i (E_i \otimes F_i)\rho(E_i^\dagger \otimes F_i^\dagger), \tag{1.33}$$

$$\sum_i E_i^\dagger E_i \otimes F_i^\dagger F_i = I_{AB}. \tag{1.34}$$

Since LOCC $\subset$ SEP, any necessary conditions proved for SEP also apply for LOCC. The relaxation of LOCC to the much simpler SEP can be thought of as the quantum analogue of relaxing classical communication protocols to what are called combinatorial rectangles.

1.3.15 Computational complexity and Bachmann–Landau notation

The amount of time an algorithm requires in order to complete some task depends upon the hardware on which it is running: all else being equal a computer that is twice as fast will complete a task in half the time. On the other hand, the way in which runtime scales as a function of the size of a problem is thought to be more or less independent of the specifics of the hardware. The exception to this is that quantum computers (if they can be built) are widely believed to be capable of solving certain problems asymptotically faster than any classical computer. Note that in these cases, the quantum computer will be running an algorithm quite different from that used on the classical computer. For example, Shor's quantum algorithm for finding the prime factors of a number runs asymptotically faster than any known classical algorithm.

Statements regarding the way a function (runtime in this case) scales as a function of input are made using Bachmann–Landau notation. There are three notations that we will make use of. We will write $f(n) \in \mathrm{O}(g(n))$ to mean that there is a k such that for sufficiently large n, $f(n) \leq k \cdot g(n)$. In other words, $f(n)$ is asymptotically upper bounded by $g(n)$. We will write $f(n) \in \Omega(g(n))$ to mean that there is a k such that for sufficiently large n, $f(n) \geq k \cdot g(n)$. In other words, $f(n)$ is asymptotically lower bounded by $g(n)$. We will write $f(n) \in \Theta(g(n))$ to mean that there are k_1 and k_2 such that for sufficiently large n, $k_1 \cdot g(n) \leq f(n) \leq k_2 \cdot g(n)$. In other words, $f(n)$ is asymptotically bounded from above and below by $g(n)$. We will follow the almost universally used abuse of notation by writing, for example, $f(n) = \mathrm{O}(g(n))$ instead of $f(n) \in \mathrm{O}(g(n))$. The notation $\mathrm{poly}(n)$ means a

polynomial in the variable n. So, for instance, $f(n) = \mathrm{O}(\mathrm{poly}(n))$ means that $f(n)$ is asymptotically bounded from above by some polynomial in n.

For example, a problem of size n is said to be solvable in $\mathrm{O}(n^2)$ time if there is some algorithm that runs in time *at most* kn^2 for some k (and for all sufficiently large n). The specifics of the hardware do not matter, since the constant factor k absorbs any speed advantage a fast computer would have compared to a slow computer. A problem is said to require $\Omega(\sqrt{n})$ memory if any possible algorithm would require *at least* $k\sqrt{n}$ bytes of memory, for some k.

In general, it is very difficult to prove that a classical (or quantum) computer cannot solve a problem in a certain amount of time. In fact, it is not even known whether there is any problem that a quantum computer can solve in polynomial time, $\mathrm{O}(\mathrm{poly}(n))$, that a classical computer cannot solve in polynomial time. This is widely believed to be the case, as Shor's algorithm can efficiently factor numbers on a quantum computer and this is thought to be very time consuming on a classical computer. But so far there is no proof, and solving this requires proving P $\neq$ PSPACE, a well known and extremely difficult open problem in computer science.

This difficulty can be overcome by introducing an *oracle*, which can be thought of as a box that computes, in unit time, some unknown function. The box cannot be opened, rather it can only be queried. One then asks questions such as how many oracle queries are needed in order to determine whether there is a value x for which some function f gives $f(x) = 1$ when the only access to f is through the oracle which, when queried with x, reveals the value $f(x)$. This particular problem is known as *unstructured search*. One is not given any other information about f, in particular one does not have access to a description of a circuit which implements f. A query to the oracle is considered to only take unit time. Quantum circuits are granted access to the oracle in the form of a unitary operator; in particular the quantum version of an oracle can perform queries on superposition states. This will be explained in further detail in section 3.5.3. Techniques do exist to show that quantum computers are asymptotically faster than classical computers *relative to an oracle*. For example, the unstructured search problem on a domain of size N can be solved by a quantum computer using only $\mathrm{O}(\sqrt{N})$ queries whereas any classical algorithm would require $\Omega(N)$ queries. For certain other problems, quantum algorithms are known that exhibit exponential speedup as compared to any possible classical algorithm, relative to an oracle [Sim94].

1.3.16 Conic and semidefinite programming

A *closed convex cone* $\mathcal{K}$ is a closed subset of some vector space such that for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{K}$ and $\lambda \geq 0$ we have $\lambda\boldsymbol{x} \in \mathcal{K}$ and $\boldsymbol{x} + \boldsymbol{y} \in \mathcal{K}$. The *dual cone* is

$$\mathcal{K}^* = \{\boldsymbol{y} : \langle \boldsymbol{y}, \boldsymbol{x}\rangle \geq 0 \text{ for all } \boldsymbol{x} \in \mathcal{K}\}. \tag{1.35}$$

The notation $\langle \boldsymbol{y}, \boldsymbol{x}\rangle$ denotes inner product. This is again a closed convex cone. The dual of $\mathcal{K}^*$ is $\mathcal{K}$.

Conic programming involves optimization of a linear functional (which can always be expressed as an inner product) over a set defined in terms of two closed convex cones $\mathcal{K}$ and $\mathcal{L}$. Each such problem has a *dual* (the original problem is called the *primal*). The primal and dual take the form (section 4.7 of [GM12])

$$\text{(Primal)} \quad \max\{\langle \mathbf{c}, \boldsymbol{x}\rangle : \mathbf{b} - A\boldsymbol{x} \in \mathcal{L}, \boldsymbol{x} \in \mathcal{K}\} \tag{1.36}$$

$$\text{(Dual)} \quad \min\{\langle \mathbf{b}, \boldsymbol{y}\rangle : A^T\boldsymbol{y} - \mathbf{c} \in \mathcal{K}^*, \boldsymbol{y} \in \mathcal{L}^*\} \tag{1.37}$$

where $\mathcal{K}$ and $\mathcal{L}$ are closed convex cones, $\boldsymbol{x}$ and $\boldsymbol{y}$ are the vectors being optimized over, A is a matrix, and $\mathbf{b}$ and $\mathbf{c}$ are constant vectors. All of these quantities are real. Under rather general conditions (see [GM12] for details), the primal and dual will take the same value; this is called *strong duality*. A wide range of problems can be adapted to fit the above form. For instance, if one has more than one variable, say $\boldsymbol{x}_1 \in \mathcal{K}_1$ and $\boldsymbol{x}_2 \in \mathcal{K}_2$, then one can define $\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2$ and $\mathcal{K} = \mathcal{K}_1 \oplus \mathcal{K}_2$ to put the problem in the form (1.36), where $\oplus$ is the direct sum.

Although the above formulation describes optimization with vector variables, it is in fact also possible to use matrix variables. For instance, the set of 2-by-2 Hermitian matrices can be thought of as a four dimensional vector space over $\mathbb{R}$ with basis (1.32). In fact, a similar statement holds for Hermitian matrices of arbitrary dimension. Among Hermitian matrices of a given dimension, the set of positive semidefinite matrices forms a cone. Given superoperators Φ_1 and Φ_2 which map Hermitian matrices to Hermitian matrices, and given Hermitian matrices A, B_1, and B_2, it is possible to derive from (1.36)-(1.37) the following forms: [Wat11]

$$\text{(Primal)} \quad \max\{\langle A, X\rangle : \Phi_1(X) = B_1, \Phi_2(X) \preceq B_2, X \succeq 0\} \tag{1.38}$$

$$\text{(Dual)} \quad \min\{\langle B_1, Y_1\rangle + \langle B_2, Y_2\rangle : \Phi_1^*(Y_1) + \Phi_2^*(Y_2) \succeq A, Y_2 \succeq 0\}. \tag{1.39}$$

The variables being optimized over, X, Y_1, and Y_2, are all Hermitian. Such programs are known as *semidefinite programs*. Again, under rather general conditions these two forms will take the same value. To quote [Wat11], if one does not *try* to make strong duality fail, it will probably hold.

Semidefinite programs can typically be efficiently solved numerically, both in theory and in practice. Often, a problem which cannot be solved efficiently can in fact be efficiently approximated via a *semidefinite relaxation* in which discrete (often binary) variables are allowed to take vector values. This is called a relaxation because some of the constraints (that the variables be discrete) are relaxed.

As an example consider the clique number of a graph, $\omega(G)$. Computing this quantity requires solving an NP-complete problem, but it admits the following semidefinite relaxation. Consider a clique of G of size t, and let χ_i be the normalized indicator function for this clique (i.e. $\chi_i = 1/\sqrt{t}$ if $i \in G$ is part of the clique and $\chi_i = 0$ otherwise). Define the matrix with entries

$$B_{ij} = \chi_i \chi_j. \tag{1.40}$$

Since vertices in a clique must be adjacent by definition, whenever $i \not\sim j, i \neq j$ it cannot be that both $\chi_i = 1$ and $\chi_j = 1$; therefore $B_{ij} = 0$. Also we have $\mathrm{Tr}B = t$ and $\sum_{ij} B_{ij} = t^2$.

We can now formulate $\omega(G)$ as an optimization problem. The domain will consist of all possible assignments of 0 or 1 to each χ_i. It is not hard to see that

$$\max\left\{\frac{\sum_{ij} B_{ij}}{\mathrm{Tr}B} : B_{ij} = \chi_i\chi_j, \chi_i \in \{0,1\}, (B_{ij} = 0 \text{ whenever } i \not\sim j, i \neq j)\right\} \tag{1.41}$$

is equal to $\omega(G)$. In fact, the χ_i that achieves the maximum in (1.41) will be the indicator function for the maximum clique. But such an optimization cannot be computed efficiently.

Now relax the problem by allowing each χ_i to be an arbitrary vector, in some arbitrary Hilbert space. The matrix $B_{ij} = \langle \chi_i, \chi_j\rangle$, being a Gram matrix, is positive semidefinite (conversely, if $B \succeq 0$ then it is a Gram matrix). With this relaxation, the integer program (1.41) becomes the semidefinite program

$$\max\left\{\frac{\sum_{ij} B_{ij}}{\mathrm{Tr}B} : B \succeq 0, B_{ij} = 0 \text{ whenever } i \not\sim j, i \neq j\right\}. \tag{1.42}$$

Because the maximization (1.42) has more relaxed constraints than (1.41), it will achieve at least as large a value; in other words, it will be an upper bound on $\omega(G)$. Since the constraints and the objective value aren't affected by rescaling of B, we might as well add the constraint $\mathrm{Tr}B = 1$. Then (1.42) is seen to be equivalent to (1.18), the Lovász $\bar{\vartheta}$ number of G. The dual of this semidefinite program is (1.19).

Although the prior example involved relaxing binary variables by letting them take on vector values, there are other possible relaxation techniques. For example, one could relax an optimization over matrices by considering these matrices to be vectors and forgetting any of their matrix properties (such as semidefiniteness). We will consider such a relaxation in chapter 4.

1.4 Overview of dissertation chapters

Each chapter of this dissertation consists of a self contained paper that has been either published or submitted for publication to a peer reviewed journal. I was the sole author on two of these papers and the other two were collaborations, as described below. In this section I summarize the contents and discuss the relevance of each paper to the field of quantum information.

1.4.1 Chapter 2: Entanglement requirements for implementing bipartite unitary operations

This was a collaboration with Robert B. Griffiths and has been published in Physical Review A [SG11].

Suppose Alice and Bob each possess some quantum subsystem and they wish to collaborate to perform some unitary operation on the joint system (a *bipartite unitary*). They can communicate over a perfect channel (e.g. telephone), but are not able to communicate over any quantum channel. In other words, we are considering LOCC (see section 1.3.14).

For a nontrivial unitary (specifically, one that is not a product operator) this task cannot be completed unless Alice and Bob possess an entangled state. The question we investigate is how much entanglement is required. One might expect that if the unitary were very close to the identity then not much entanglement would be needed, since the operation is acting only very weakly. This is not the case: it had recently been shown [STM11a] that a certain class of unitaries (the controlled-unitary operations) acting on two qubits always require a maximally entangled state.

We found a much simpler proof that applies to arbitrary unitary operations on an arbitrary number of qubits or qudits. Specifically, we showed if the entanglement is of minimum feasible Schmidt rank then it must be maximally entangled. We also showed, by numerical search, that if a larger Schmidt rank is used then less entanglement suffices. In fact, in some cases it is even possible to use less than one bit of entanglement (in terms of Von Neumann entropy). This demonstrates an interesting tradeoff between two measures of entanglement: for this task minimal Schmidt rank requires maximal entropy, but larger Schmidt rank allows smaller entropy.

My contribution to this paper consisted of the main proof. Robert Griffiths worked with me to simplify the proof, and wrote the section on map-state duality.

1.4.2 Chapter 3: Quantum interference as a resource for quantum speedup

This work has been published in Physical Review A [Sta14a].

For certain problems there are quantum algorithms that run much faster than any known classical algorithm. Shor's algorithm for factoring numbers and Grover's algorithm for unstructured search are two notable examples. Other than a few ad-hoc explanations in reference to particular algorithms it is not really clear, in an intuitive sense, what makes quantum computers faster. The source of the "quantum speedup" is still not well understood.

One approach is to investigate necessary resources. In other words, which features unique to quantum mechanics, when disallowed, remove the quantum speedup? Previous work has focused on various measures of entanglement or various restrictions on the classes of gates used.

I considered quantum interference as a necessary resource. More precisely, I defined a measure of interference and considered the amount of interference each operation was capable of producing, a quantity I termed *interference producing capacity*. If the product of the interference producing capacities for all operations in a quantum circuit is small then there can be no quantum speedup, since such circuits can then be efficiently simulated by a classical computer. It remains an open question whether interference itself, rather than interference producing capacity, is a necessary resource.

1.4.3 Chapter 4: Bounds on Entanglement Assisted Source-channel Coding via the Lovász ϑ Number and its Variants

This was a collaboration with Toby Cubitt, Laura Mančinska, David Roberson, Simone Severini, and Andreas Winter. This work was posted on arXiv [CMR+13] and has been submitted to IEEE Transactions on Information Theory. It was presented at the Theory of Quantum Computation, Communication & Cryptography 2014 conference.

We investigate zero-error source-channel coding, assisted by an entanglement resource. In this scenario, Alice wishes to send a message to Bob using a noisy channel, making use of the fact that Bob already has some side information regarding Alice's message. Bob must be able to determine Alice's message with zero chance of error. The message and the channel are both classical; the only quantum aspect is that Alice and Bob possess an entangled state which can be used to their advantage. This is an interesting problem to investigate because it is a case where quantum entanglement provides an advantage for an otherwise purely classical task.

It was known that three graph invariants, the Lovász ϑ, Schrijver ϑ', and Szegedy ϑ^+ numbers, provide necessary conditions for source-channel coding in the absence of entanglement assistance [NTR06, dCST13]. Beigi showed the Lovász number provides a necessary condition when there is entanglement assistance but no side information. He did this by defining a quantity β which is an upper bound on one-shot entanglement assisted capacity and which is no greater than $\lfloor \vartheta \rfloor$. Beigi posed as an open question whether $\beta = \lfloor \vartheta \rfloor$ [Bei10]. This open question is significant because it determines whether β could be a useful quantity for finding a gap between ϑ and the asymptotic entanglement assisted capacity. Recently the Szegedy number was found to provide a necessary condition for entanglement assisted source coding with a perfect (i.e. not noisy) channel [BBL+13].

We unified and strengthened all these results by showing all three of the Lovász, Schrijver, and Szegedy numbers provide necessary conditions for the entanglement assisted source-channel coding problem. This was accomplished via two related semidefinite relaxations. Investigating these further, we answer Beigi's open question in the affirmative, $\beta = \lfloor \vartheta \rfloor$. Another group, working independently from and concurrently with us, investigated a similar semidefinite relaxation of the quantum chromatic number [PT13]. The authors posed the question of whether a gap existed between the quantum chromatic number and their relaxed quantity. We found that their math was similar to ours and amended our paper to answer their question in the affirmative.

Laura Mančinska and David Roberson proved that $\bar{\vartheta}$, $\bar{\vartheta}'$, and $\bar{\vartheta}^+$ are monotone under entanglement assisted homomorphisms. Toby Cubitt, Simone Severini, and Andreas Winter proved that $\beta = \lfloor \vartheta \rfloor$. I grafted the Cubitt, Severini, Winter proof onto a proof from [BBL+13]. Most of the intermediate steps canceled out and the result was a converse of the result of Mančinska and Roberson. David Roberson and I solved (mostly simultaneously and independently) the open question from [PT13].

1.4.4 Chapter 5: Quantum source-channel coding and non-commutative graph theory

This work was posted on arXiv [Sta14b] and has been submitted to IEEE Transactions on Information Theory.

I define and investigate a fully quantum version of zero-error source-channel coding. Alice and Bob receive a (possibly entangled) state from some finite collection or from some subspace known ahead of time. Alice is to send a message through a noisy quantum channel such that Bob may reproduce the input state, with zero chance of error. This framework encompasses, for example, teleportation, dense coding, entanglement assisted quantum channel capacity, and one-way communication complexity of function evaluation.

In [DSW13] zero-error quantum channel capacity was investigated in terms of *non-commutative graphs*, in which operator subspaces take the place of the adjacency matrix of a graph. They generalized the Lovász number for non-commutative graphs and showed that it bounds the asymptotic

entanglement assisted quantum channel capacity. I show that non-commutative graphs apply also to the quantum source-channel coding problem. This involves generalization of graph homomorphisms to non-commutative graphs, which I hope will pave the way toward a richer theory of non-commutative graphs, a program listed as an open problem in [DSW13]. I show the Lovász, Schrijver, and Szegedy numbers all provide bounds on quantum source-channel coding. The latter two quantities were generalized to non-commutative graphs for the first time in this paper.

As an application, I construct a quantum channel whose zero-error entanglement assisted one-shot capacity can only be achieved using a non-maximally entangled state. This is noteworthy for two reasons. First, maximally entangled states are typically considered to be a more powerful resource than non-maximally entangled states. This result provides a counterexample to this intuition. Secondly, if such an example (needing a non-maximally entangled state to achieve one-shot capacity) were to be found for a classical rather than a quantum channel, this would solve the open problem of whether the one-shot entanglement assisted capacity for classical channels is always equal to the quantum clique number of their distinguishability graphs. The quantum clique number is defined in a similar way as the quantum chromatic number of section 1.2.3.

Chapter 2

Entanglement requirements for implementing bipartite unitary operations[1]

[1] This work was published in: Dan Stahlke and Robert B. Griffiths, *Entanglement requirements for implementing bipartite unitary operations*, Phys. Rev. A **84**, 032316 (2011).

2.1 Abstract

We prove, using a new method based on map-state duality, lower bounds on entanglement resources needed to deterministically implement a bipartite unitary using separable (SEP) operations, which include LOCC (local operations and classical communication) as a particular case. It is known that the Schmidt rank of an entangled pure state resource cannot be less than the Schmidt rank of the unitary. We prove that if these ranks are equal the resource must be uniformly (maximally) entangled: equal nonzero Schmidt coefficients. Higher rank resources can have less entanglement: we have found numerical examples of Schmidt rank 2 unitaries which can be deterministically implemented, by either SEP or LOCC, using an entangled resource of two qutrits with less than one ebit of entanglement.

2.2 Introduction

It is possible to carry out nonlocal quantum operations on multipartite systems using only local quantum operations and classical communications (LOCC) provided that the parties involved have access to a suitable entangled state, referred to as a resource. Given a large enough resource it is always possible to use teleportation to send all inputs to one party, who performs the operation and then distributes the results to the other parties using teleportation. In some cases it is possible to perform a nonlocal operation with less entanglement than is required by teleportation [EJPP00, RAG02, YGC10, Coh10, GYC10, ZW08]. The question then arises as to how much entanglement is really necessary in order to implement a given nonlocal operation.

Our first result, that the Schmidt rank of the resource must be at least as great as that of the unitary [Theorem 1(a)], follows rather immediately from the fact that it is a separable (SEP) operation. This is analogous to the result given in [DVC02] in which probabilistic (i.e. SLOCC) implementations are considered. Since SEP is contained in SLOCC, our Theorem 1(a) can be seen as a consequence of the result in [DVC02], however we provide an independent proof along the way to our main result.

In contrast to the probabilistic case, the deterministic implementation of a unitary is only possible if the state meets certain entanglement requirements. For one thing, the entanglement of the resource must be at least as great as the entangling power of the unitary since entanglement cannot increase under SEP [GG08]. It has been shown that any deterministic controlled-unitary operator on two qubits implemented with bipartite LOCC using a resource of two entangled qubits necessarily requires a maximally entangled resource [STM11b]. Our paper takes a different approach to the problem, using SEP, and provides a proof applicable to general unitaries of arbitrary dimension. We show that if the resource has Schmidt rank equal to that of the unitary, the resource must be uniformly entangled in the sense that all its nonzero Schmidt coefficients are the same [Theorem 1(b)]. These same restrictions apply to LOCC, as it is a particular case of SEP.

It is not hard to see that if the Schmidt rank of the resource is greater than the Schmidt rank of the unitary, then the resource need not be uniformly entangled (e.g. a larger rank resource that is majorized by a smaller rank maximally entangled state). We have found that it is in fact possible for such a larger rank resource to have less entanglement than would be required for a resource of Schmidt rank equal to that of the unitary. We have found examples of protocols in both SEP and LOCC which deterministically implement a controlled phase operation using less than one ebit of entanglement. In this case the unitary has Schmidt rank two and the resource has Schmidt rank three. Although the nonlocal unitary protocol given in [CDKL01] can with certain probability consume less than one ebit of entanglement[2], we believe that ours is the first example of carrying out such a

[2] Although the protocol given in [CDKL01] is deterministic in the sense that it always succeeds in a finite number of steps, it is probabilistic in the amount of entanglement required. For any nontrivial unitary there is a chance that the protocol requires usage of the $|\psi_{\alpha_2}\rangle$ state, which has one ebit of entanglement. Thus, if the protocol only has access to a state with less than one ebit of entanglement there is a nonzero probability that the protocol cannot be carried out successfully.

protocol deterministically using less than one ebit of entanglement.

The remainder of this article is organized as follows. Section 2.3 sets up the problem of bipartite deterministic implementations of unitary operators using SEP. Section 2.4 provides the requisite background regarding map-state duality [ZB04, BZ06] and atemporal diagrams [GWYC06, Ste09, Cvi08]. Our main result is proved in Sec. 2.5 using what we believe to be a new method based on the use of map-state duality. In Sec. 2.6 we consider the case of a resource of larger Schmidt rank. There is a brief conclusion in Sec. 2.7. An appendix details the implementation of a controlled unitary using a qutrit resource state of less than one ebit of entanglement.

2.3 Nonlocal Unitaries Via Separable operations

We are interested in carrying out a bipartite unitary map $U : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_{\bar{A}} \otimes \mathcal{H}_{\bar{B}}$, using as a resource an entangled state $|\psi\rangle$ on two ancillary systems $\mathcal{H}_a$ and $\mathcal{H}_b$, by means of a separable operation $\{E_k \otimes F_k\}$, $k = 1, 2, \ldots$. Here $E_k : \mathcal{H}_A \otimes \mathcal{H}_a \to \mathcal{H}_{\bar{A}}$ and $F_k : \mathcal{H}_B \otimes \mathcal{H}_b \to \mathcal{H}_{\bar{B}}$ together form a product Kraus operator. For U to be unitary it is necessary that the dimensions of the Hilbert spaces satisfy $d_A d_B = d_{\bar{A}} d_{\bar{B}}$, but we do not require that $d_A = d_{\bar{A}}$ or $d_A = d_B$. The separable operation must satisfy the usual closure condition [HHHH09]

$$\sum_k (E_k \otimes F_k)^\dagger (E_k \otimes F_k) = I_A \otimes I_a \otimes I_b \otimes I_B \tag{2.1}$$

which is depicted in Fig. 2.1(a).

In addition, for $|\Phi\rangle$ any pure input state on $\mathcal{H}_A \otimes \mathcal{H}_B$, the outcome of the operation will be a pure state

$$U\Big(|\Phi\rangle\langle\Phi|\Big)U^\dagger = \sum_k \Big(E_k \otimes F_k\Big)\Big(|\Phi\rangle\langle\Phi| \otimes |\psi\rangle\langle\psi|\Big)\Big(E_k \otimes F_k\Big)^\dagger \tag{2.2}$$

on $\mathcal{H}_{\bar{A}} \otimes \mathcal{H}_{\bar{B}}$. Since the protocol is assumed to be deterministic, every term on the right side is proportional to the same pure state and it must be the case that

$$(E_k \otimes F_k)\,|\psi\rangle = \alpha_k U, \tag{2.3}$$

with α_k some complex number. Note that both sides of (2.3) are operators acting on $\mathcal{H}_A \otimes \mathcal{H}_B$; Fig. 2.2(a) will help interpreting it correctly.

The resource $|\psi\rangle$ is assumed to have a Schmidt rank of D_ψ, which means it can be written in the form

$$|\psi\rangle = \sum_{i=1}^{D_\psi} \lambda_i\,|a_i\rangle \otimes |b_i\rangle\,. \tag{2.4}$$

for suitable orthonormal bases $\{|a_i\rangle\}$ and $\{|b_i\rangle\}$ of $\mathcal{H}_a$ and $\mathcal{H}_b$, with Schmidt coefficients $\lambda_i > 0$ for $i \le D_\psi$.

Similarly, the bipartite operator U is assumed to have a Schmidt rank of D_U, meaning that it can be written in the form [Tys03]

$$U = \sum_{i=1}^{D_U} \mu_i \mathcal{A}_i \otimes \mathcal{B}_i, \tag{2.5}$$

where $\{\mathcal{A}_i\}$ and $\{\mathcal{B}_i\}$ are bases of the operator spaces $\mathcal{L}(\mathcal{H}_A, \mathcal{H}_{\bar{A}})$ and $\mathcal{L}(\mathcal{H}_B, \mathcal{H}_{\bar{B}})$, orthonormal under the Frobenius (Hilbert–Schmidt) inner product, and $\mu_i > 0$ for $i \le D_U$. Equivalently, D_U is the minimum number of terms needed in order to write U in the form $\sum \mathcal{C}_i \otimes \mathcal{D}_i$, without requiring $\mathcal{C}_i$ or $\mathcal{D}_i$ to be from an orthonormal basis.

2.4 Map-State Duality and Diagrams

Map-state duality [ZB04, BZ06] plays a central role in the proof that will follow. This is a general concept that is sometimes referred to as reshaping or a partial transpose [ZB04] and in a specific manifestation is known as the Jamiołkowski or sometimes the Choi–Jamiołkowski isomorphism. States and maps are considered to both be tensors, and when a choice of orthonormal basis is fixed there is a natural linear relation between bras and kets (i.e. $|i\rangle \leftrightarrow \langle i|$ for all basis vectors $|i\rangle$)[3].

With this identification between bras and kets in place, the bipartite state $|\psi\rangle$ on the Hilbert space $\mathcal{H}_a \otimes \mathcal{H}_b$ can be identified with the linear map $\psi' : \mathcal{H}_b \to \mathcal{H}_a$ obtained by turning kets into bras on the $\mathcal{H}_b$ space:

$$|\psi\rangle = \sum_{ij} \psi_{ij} |a_i\rangle \otimes |b_j\rangle \to \psi' = \sum_{ij} \psi_{ij} |a_i\rangle \langle b_j| \tag{2.6}$$

Similarly, the operators U, E_k, and F_k, give rise to $U' : \mathcal{H}_B \otimes \mathcal{H}_{\bar{B}} \to \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$ (by turning bras into kets on $\mathcal{H}_A$ and kets into bras on $\mathcal{H}_{\bar{B}}$), $E'_k : \mathcal{H}_a \to \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$ (by turning bras into kets on $\mathcal{H}_A$), and $F_k'^T : \mathcal{H}_B \otimes \mathcal{H}_{\bar{B}} \to \mathcal{H}_b$ (by turning bras into kets on $\mathcal{H}_b$ and kets into bras on $\mathcal{H}_{\bar{B}}$),

$$\begin{aligned} U' &= \sum_{ijmn} \langle \bar{A}_j, \bar{B}_n | U | A_i, B_m \rangle | A_i, \bar{A}_j \rangle \langle B_m, \bar{B}_n | , \\ E'_k &= \sum_{ijm} \langle \bar{A}_j | E_k | A_i, a_m \rangle | A_i, \bar{A}_j \rangle \langle a_m | , \\ F_k'^T &= \sum_{ijm} \langle \bar{B}_j | F_k | B_i, b_m \rangle | b_m \rangle \langle B_i, \bar{B}_j | . \end{aligned} \tag{2.7}$$

In the case of these three operators, *map-map duality* may be a more precise term, however we will use map-state duality to refer to any such partial transpose. The primed operator for F_k is denoted as $F_k'^T$ in order to draw attention to the fact that its domain and range are swapped in comparison to E'_k.

The equations introduced so far make use of six distinct Hilbert spaces and tensors of various rank. In such situations the underlying structure of equations can be somewhat hidden when expressed using Dirac notation. Abstract index notation is more transparent but can become unwieldy. For this reason we provide atemporal diagrams, similar to those found in [GWYC06], which should aid the reader in following the arguments in the text.

Operators are designated by squares or rectangular boxes. As a matter of style, the state $|\psi\rangle$ and its corresponding operator ψ' will be represented as a circle instead of a square. Lines between these boxes represent tensor contraction, and these lines are labeled by the Hilbert spaces which they correspond to. Open lines on the left of a diagram represent the input to the total linear operator defined by the diagram, and open lines on the right represent outputs. Putting the inputs on the left means that operators are to be applied in a left-to-right manner, opposite to how algebraic equations are interpreted. As has been so far described, our diagrams are to be interpreted in exactly the same way as traditional quantum circuits as used for example in Nielsen and Chuang [NC00]. The primary difference between our diagrams and traditional circuits is that in the latter the horizontal direction is understood to represent the passage of time whereas our diagrams make no reference to time. The presence of a summation symbol has the obvious meaning: the linear operator depicted in the diagram denotes the terms of a series. The trace or partial trace operation is just a special case of tensor contraction and is denoted by joining the relevant spaces with a line. The identity operator is represented by a line. With minor changes in style our diagrams are equivalent to the atemporal diagrams of [GWYC06], and resemble other such schemes [Ste09, Cvi08].

[3] It is also possible to formulate map-state duality in a basis independent manner [Tys03], however this is not necessary for the present work.

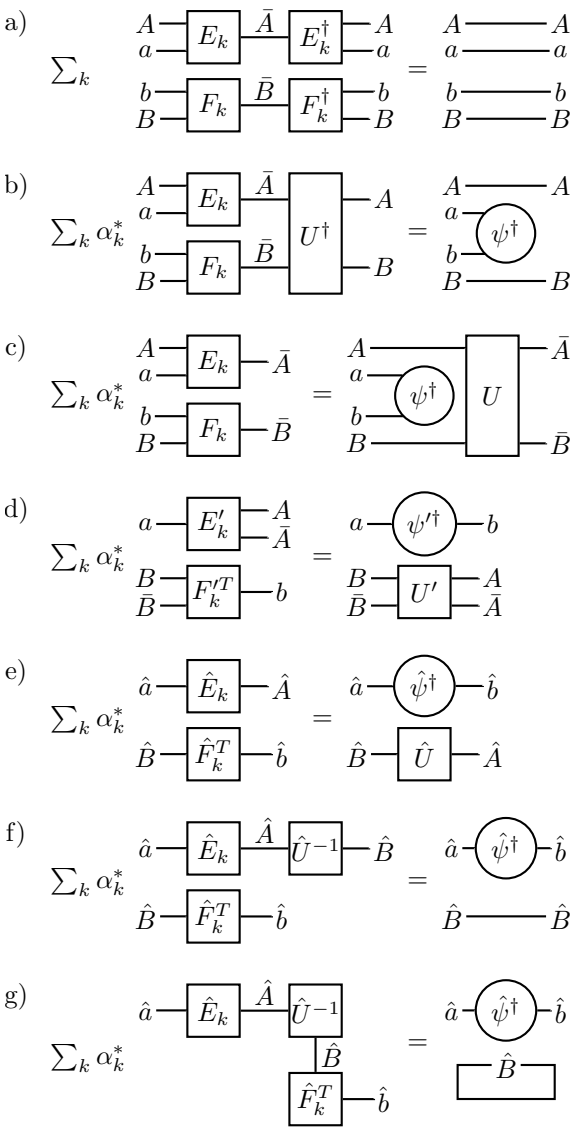


Figure 2.1: Atemporal diagrams, explained in Sec. 2.4. (a) Closure condition, (2.1). (b) Apply $\langle\psi|$ and simplify using the adjoint of Fig. 2.2(a) to get (2.9). (c) Multiply on the right by U to get (2.10). (d) Apply map-state duality to get (2.11). (e) Restrict spaces to supports and ranges of operators to get (2.12). (f) Multiply by $\hat{U}^{-1}$. (g) Trace over $\mathcal{H}_{\hat{B}}$ to get (2.13).

2.5 Entanglement Requirements

Our main result is the following:

Theorem 2.1. *Suppose that a unitary operator U is implemented deterministically by a separable operation that makes use of the pure state entanglement resource $|\psi\rangle$ [i.e. suppose that (2.1) and (2.3) hold]. Then*

(a) The Schmidt rank D_ψ of $|\psi\rangle$ is greater than or equal to the Schmidt rank D_U of U.

(b) If the Schmidt ranks are equal, $D_U = D_\psi$, then $|\psi\rangle$ must be a uniformly (maximally) entangled state: all the nonzero Schmidt coefficients are the same.

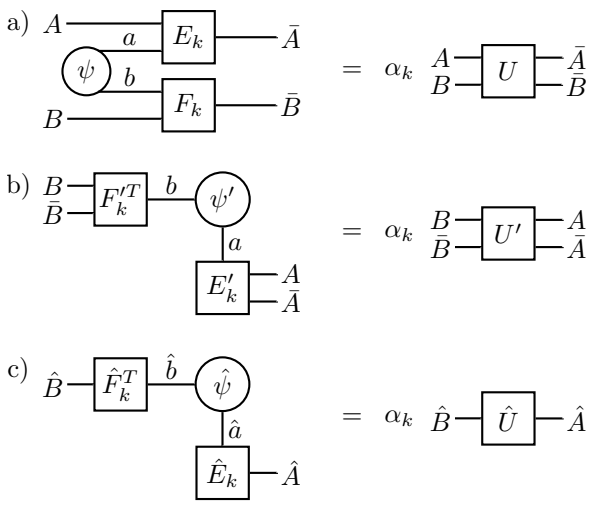


Figure 2.2: (a) Deterministic unitary operation, (2.3). (b) Apply map-state duality to get (2.8). (c) Restrict spaces to supports and ranges of operators to get (2.14).

Proof of (a). Making use of map-state duality and the operators defined in (2.6) and (2.7), equation (2.3) [Fig. 2.2(a)] can be rewritten as [Fig. 2.2(b)]

$$E_k' \psi' F_k'^T = \alpha_k U'. \tag{2.8}$$

Since the rank of a product of linear operators is at most the smallest of the ranks of the individual operators, it follows that $\mathrm{rank}(\psi') \geq \mathrm{rank}(U')$. The rank of an operator is equal to the number of its nonzero singular values. Since the Schmidt decompositions (2.4) and (2.5) are essentially singular value decompositions of ψ' and U', it is apparent that $\mathrm{rank}(\psi') = D_\psi$ and $\mathrm{rank}(U') = D_U$ and the inequality becomes $D_\psi \geq D_U$. Part (a) is proved. □

Proof of (b). Apply the closure condition (2.1) to $\langle\psi|$ and use the adjoint of (2.3) to obtain

$$\sum_k \alpha_k^* U^\dagger (E_k \otimes F_k) = \langle\psi| \otimes I_{AB}. \tag{2.9}$$

as shown in Fig. 2.1(b). Next, multiply both sides on the left by U to arrive at

$$\sum_k \alpha_k^* (E_k \otimes F_k) = \langle\psi| \otimes U, \tag{2.10}$$

as shown in Fig. 2.1(c). Making use of map-state duality gives [Fig. 2.1(d)]

$$\sum_k \alpha_k^* (E_k' \otimes F_k'^T) = \psi'^\dagger \otimes U'. \tag{2.11}$$

The map U' may in general have rank less than the dimension of $\mathcal{H}_B \otimes \mathcal{H}_{\bar{B}}$ or $\mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$ (which need not be equal to each other). In this case it will be useful to denote by $\mathcal{H}_{\hat{B}}$ the subspace of $\mathcal{H}_B \otimes \mathcal{H}_{\bar{B}}$ which forms the *support* (or co-image or row space) of U', the orthogonal complement of its kernel (null space), and by $\mathcal{H}_{\hat{A}}$ the subspace of $\mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$ that forms the *range* (or image) of U'. Each of these subspaces has a dimension equal to D_U, and U' is a nonsingular (invertible) linear map of $\mathcal{H}_{\hat{B}}$ onto $\mathcal{H}_{\hat{A}}$, which we hereafter denote by $\hat{U}$. In the same way one can introduce subspaces $\mathcal{H}_{\hat{b}}$ and $\mathcal{H}_{\hat{a}}$ of $\mathcal{H}_b$ and $\mathcal{H}_a$ which form the support and range of ψ', and define $\hat{\psi}$ to be the corresponding

nonsingular map of rank D_ψ from $\mathcal{H}_{\hat{b}}$ to $\mathcal{H}_{\hat{a}}$. Next, $\hat{E}_k$ is E'_k with its domain restricted to $\mathcal{H}_{\hat{a}}$, which can be strictly smaller than the support of E'_k, and with its range restricted to $\mathcal{H}_{\hat{A}}$, which could be smaller than the image of E'_k. Finally, $\hat{F}_k^T$ is $F_k'^T$ regarded as a map from $\mathcal{H}_B \otimes \mathcal{H}_{\bar{B}}$ to $\mathcal{H}_b$, but with domain and range restricted to $\mathcal{H}_{\hat{B}}$ and $\mathcal{H}_{\hat{b}}$, respectively[4].

The result of restricting (2.11) to the subspaces just defined is

$$\sum_k \alpha_k^* (\hat{E}_k \otimes \hat{F}_k^T) = \hat{\psi}^\dagger \otimes \hat{U}, \tag{2.12}$$

corresponding to Fig. 2.1(e). Multiplying on the left by $\hat{U}^{-1}$ and tracing over $\mathcal{H}_{\hat{B}}$ gives

$$\sum_k \alpha_k^* \hat{F}_k^T \hat{U}^{-1} \hat{E}_k = D_U \hat{\psi}^\dagger, \tag{2.13}$$

see Fig. 2.1(f) and (g). Restricting (2.8) to subspaces results in

$$\hat{E}_k \hat{\psi} \hat{F}_k^T = \alpha_k \hat{U}, \tag{2.14}$$

see Fig. 2.2(c). Here we have restricted the spaces over which matrix multiplications are being performed ($\mathcal{H}_{\hat{b}}$ and $\mathcal{H}_{\hat{a}}$ instead of $\mathcal{H}_b$ and $\mathcal{H}_a$), however equality is still maintained because the dimensions which have been eliminated correspond to the zero Schmidt coefficients of $|\psi\rangle$, which is to say the zero singular values of ψ'.

To complete the proof, make use of the assumption $D_\psi = D_U$. Then D_ψ is also the rank of $\hat{E}_k$ and $\hat{F}_k^T$: all four operators in (2.14) are full rank. Taking the inverse of both sides and inserting the result for $\hat{U}^{-1}$ in (2.13) leads to the result

$$\sum_k |\alpha_k|^2 \hat{\psi}^{-1} = \hat{\psi}^{-1} = D_\psi \hat{\psi}^\dagger, \tag{2.15}$$

where $\sum_k |\alpha_k|^2 = 1$ follows from (2.1), (2.3) and the normalization of $|\psi\rangle$. With $|\psi\rangle$ in Schmidt form, $\hat{\psi}$ is diagonal, so $\hat{\psi} = I/\sqrt{D_\psi}$. Therefore all the nonzero Schmidt coefficients of $|\psi\rangle$ are equal to $1/\sqrt{D_\psi}$.

□

2.6 Larger Rank Resource

We have proved that a resource that is of the smallest viable Schmidt rank must be maximally entangled, but it is also possible to use a resource that is of higher Schmidt rank that is not maximally entangled. For one thing, if such a state meets an appropriate majorization criterion it can be deterministically transformed into a maximally entangled state [Nie99]. In this case the larger rank initial resource would have greater entanglement than would be required if the smaller maximally entangled state had been used in the first place. There is however the possibility that some protocol could be devised to use a resource of larger Schmidt rank that has less entanglement than the maximally entangled state of smaller rank.

In fact, we have numerically found examples of such constructions in both SEP and LOCC. One solution in SEP uses a resource state $|\psi\rangle = \sqrt{0.81}\,|00\rangle + \sqrt{0.095}(|11\rangle + |22\rangle)$ on two qutrits to implement the two qubit controlled unitary operator $U = \mathrm{diag}\{1, 1, 1, e^{i\phi}\}$ with $\phi = 2\cos^{-1}(35/36)$. We have verified this to be an exact solution using a computer algebra system. This resource constitutes less than one ebit of entanglement: the Von Neumann entropy is approximately 0.89 ebits.

[4] It is significant that we define $\hat{F}_k^T$ as $F_k'^T$ restricted to subspaces. In general it is not the case that $\hat{F}_k$ is equal to F'_k restricted to subspaces.

Since entropy cannot increase under SEP [GG08] it is necessary for the resource that is consumed to have greater entanglement than the entangling capacity of the unitary being implemented. The entangling capacity of this unitary is shown in [LHL03] to be approximately 0.23 ebits. Since this is much less than the 0.89 ebits that we use, there remains the possibility that a different construction or an even larger rank resource could potentially lower the entanglement cost further.

We also found an LOCC protocol which, though less efficient than the SEP construction just described, allows one to carry out a bipartite unitary deterministically using a resource with less than one ebit of entanglement. The resource in this case is $|\psi\rangle = \sqrt{0.8}\,|00\rangle + \sqrt{0.1}(|11\rangle + |22\rangle)$ and the unitary implemented is $U = \mathrm{diag}\{1, 1, 1, e^{i\phi}\}$ with $\phi = 0.08\pi$. The Von Neumann entropy of this resource is approximately 0.92 ebits, and this is a four round protocol (Alice, Bob, Alice, Bob).

The constructions described above are instances of a more general continuous family of solutions that we have found, covering a range of controlled phase operations. As should be expected, a larger phase ϕ requires a larger entanglement resource. In both the SEP and the LOCC case only certain classes of solutions were searched for, so it is possible that a more thorough search would provide more efficient protocols. The details of our SEP construction are presented in Appendix 2.A. Our LOCC construction consists of a long list of Kraus operators in numerical form, which is available upon request.

2.7 Conclusion

We have shown that a unitary operator of Schmidt rank D implemented as a bipartite separable operation requires an entanglement resource of Schmidt rank at least D. If the Schmidt rank of the resource is exactly equal to D, the resource must be uniformly (maximally) entangled with equal nonzero Schmidt coefficients. These restrictions apply also to LOCC, which is a subset of SEP. The proof uses map-state duality in a way which has not (so far as we know) been previously applied to problems of this type, so might have other interesting applications.

Numerical results show that the amount of entanglement required for the resource can be lowered by using a resource of Schmidt rank larger than D. A four round LOCC protocol has been found which uses a two-qutrit resource state with less than one ebit of entanglement to implement a bipartite controlled phase gate (albeit with a small phase).

Although some large classes of unitaries are known to have implementations in LOCC using resources having the minimal Schmidt rank required by Theorem 1(a) [EJPP00, RAG02, YGC10, Coh10, ZW08], it is not known whether such minimal-rank implementations are possible for all unitaries. Given a unitary of Schmidt rank D_U it is always possible to find a collection of operators $\{E_k \otimes F_k\}$ such that (2.9) and (2.3) are satisfied with a resource of Schmidt rank $D_\psi = D_U$. But it is not known if there is a separable operation satisfying both (2.1) and (2.3). Consequently, it is possible that some unitaries may require a resource of greater rank than the lower bound given in Theorem 1(a). Even if such a minimal rank solution is always possible in SEP, it still might not be possible in LOCC. This stands in contrast to the case of SLOCC where it is known that any unitary can be implemented using a state of Schmidt rank equal to that of the unitary [DVC02].

2.8 Acknowledgments

We thank Vlad Gheorghiu for his comments. The research reported here was supported in part by the National Science Foundation through Grant No. 0757251.

2.A Less than one ebit in SEP

We performed a numerical search for solutions to (2.1) and (2.3) with the resource and unitary taking the forms

$$|\psi\rangle = \sqrt{c_0}\,|00\rangle + \sqrt{(1-c_0)/2}(|11\rangle + |22\rangle), \tag{2.16}$$

$$U = \mathrm{diag}\{1, 1, 1, e^{i\theta}\}. \tag{2.17}$$

In this case the unitary U is Schmidt rank 2 and the resource is Schmidt rank 3, so the spaces $\mathcal{H}_a$ and $\mathcal{H}_b$ are each 3 dimensional. In order to reduce the search space we looked for operators $\{E_k\}$ and $\{F_k\}$ of the form

$$E_k = E_* S_k \text{ and } F_k = F_* T_k \tag{2.18}$$

where $S_k : \mathcal{H}_a \to \mathcal{H}_c$, $T_k : \mathcal{H}_b \to \mathcal{H}_c$ with $\mathcal{H}_c$ being a two dimensional space, and

$$E_* = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_{\bar{A}A} \otimes \langle 0|_c + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}_{\bar{A}A} \otimes \langle 1|_c\,, \tag{2.19}$$

$$F_* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}_{\bar{B}B} \otimes \langle 0|_c + \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}_{\bar{B}B} \otimes \langle 1|_c\,. \tag{2.20}$$

It is possible to take advantage of the symmetry of the resource $|\psi\rangle$ by searching for operator sets of the form

$$\{S_k L^l M^m N^n\} \text{ and } \{T_k L^l M^m N^n\} \tag{2.21}$$

where $l, m, n \in \{0, 1\}$ and L, M, and N are defined by

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{2.22}$$

$M = \mathrm{diag}(1, 1, -1)$, and $N = \mathrm{diag}(1, -1, 1)$. There is no loss of generality in this assumption, since if $(\{S_k\}, \{T_k\})$ gives a solution to (2.1) and (2.3) then so does $(\{\frac{1}{\sqrt{8}} S_k L^l M^m N^n\}, \{T_k L^l M^m N^n\})$. This decreases the number of independent operators (indexed by k) that need to be solved for, and in fact it turns out to be sufficient to consider only two values of k.

Initially we searched for solutions with $\theta = \pi/4$ and $c_0 = 0.6$ which, although representing more than one ebit of entanglement, is not majorized by a fully entangled resource of Schmidt rank 2. Once a solution was found, the parameters were variated until a value of c_0 was reached which represented a resource of less than one ebit of entanglement. Further constraints were added and variations made to simplify the solution and identify relations between the parameters. A family of solutions was found of the form (2.21) with

$$S_0 = \begin{pmatrix} p & 1 & -p \\ e^{i\theta/2} & -p & -1 \end{pmatrix}, \tag{2.23}$$

$$S_1 = \begin{pmatrix} -1 & 1 & -p \\ -pe^{i\theta/2} & p & 1 \end{pmatrix}, \tag{2.24}$$

$$T_0 = \begin{pmatrix} -x-y & * & * \\ (x-y)e^{-i\theta/2} & * & * \end{pmatrix}, \tag{2.25}$$

$$T_1 = \begin{pmatrix} -x+y & * & * \\ (x+y)e^{-i\theta/2} & * & * \end{pmatrix}, \tag{2.26}$$

$$p = \sqrt{\frac{1-s}{s}}, \tag{2.27}$$

$$s = x^2\,(1 - cos\,(\theta/2)) + y^2\,(1 + cos\,(\theta/2))\,, \tag{2.28}$$

where the parameters x, y, c_0, and θ must be solved for numerically. The asterisks in T_0 and T_1 represent parameters that can be found using the relation $S_k \psi' T_k^T = I/4$.

A sequence of solutions for x, y, c_0, and θ were fed into an inverse symbolic calculator of our own design which uses a lookup table to convert floating point numbers into algebraic expressions. One of these solutions produced particularly simple algebraic expressions:

$$x = 9/5, \tag{2.29}$$
$$y = -3/5, \tag{2.30}$$
$$c_0 = 0.81, \tag{2.31}$$
$$\theta = 2\arccos(35/36). \tag{2.32}$$

With this algebraic solution in hand, we used the computer algebra package Sage [S+10] to verify that this indeed represented an exact (not just approximate to within floating point precision) solution to (2.1) and (2.3).

Chapter 3

Quantum interference as a resource for quantum speedup[1]

[1] This work was published in: Dan Stahlke, *Quantum interference as a resource for quantum speedup*, Phys. Rev. A **90**, 022302 (2014). The version in this thesis contains minor typesetting differences.

3.1 Abstract

Quantum states can in a sense be thought of as generalizations of classical probability distributions, but are more powerful than probability distributions when used for computation or communication. Quantum speedup therefore requires some feature of quantum states that classical probability distributions lack. One such feature is interference. We quantify interference and show that there can be no quantum speedup due to a small number of operations incapable of generating large amounts of interference (although large numbers of such operations can in fact lead to quantum speedup). Low-interference operations include sparse unitaries, Grover reflections, short time/low energy Hamiltonian evolutions, and the Haar wavelet transform. Circuits built from such operations can be classically simulated via a Monte Carlo technique making use of a convex combination of two Markov chains. Applications to query complexity, communication complexity, and the Wigner representation are discussed.

3.2 Introduction

It is well known that certain quantum algorithms, such as Shor's and Grover's, provide a speedup compared to classical algorithms. However, the source of such quantum speedup is still somewhat of a mystery. Insight can be gained by determining necessary resources. Suppose that any quantum circuit not making use of some resource X can be efficiently simulated. Being efficiently simulated, such circuits do not exhibit quantum speedup. One can then conclude that resource X is necessary for quantum speedup. Many such resources have been identified. For circuits on pure states there is no quantum speedup if at all times (i.e. before and after every unitary) the state has small Schmidt rank [Vid03] or factors into a product state on small subsystems [JL03]. For qubit circuits there is no quantum speedup if the discord across all bipartite cuts is zero at all times [Eas10]. There is no quantum speedup for circuits that use only Clifford gates [Got98], or matchgates [Val01], that have small tree width [MS08, Joz06], or that use only operations having nonnegative Wigner representation [VFGE12, VWFE13, ME12]. For a brief overview of resources identified as important for quantum speedup see section 9 of [FRS12].

A tempting but naive explanation for quantum speedup is the exponentially large dimensionality of Hilbert space (2^n for n qubits), combined with "quantum parallelism". Shor's algorithm begins by preparing a state $\sqrt{2^{-n}}\sum_x |x\rangle \otimes |f(x)\rangle$ which can be interpreted as simultaneously evaluating f for all 2^n values of x. However, this is not a satisfactory explanation for quantum speedup since classical probability distributions over n bits can also be considered as vectors of dimension 2^n, and allow a similar sort of parallelism. We show that the quantum speedup is connected to *interference*, something which classical probability distributions lack. Prior works have mentioned interference as being important for quantum speedup but without offering a quantitative definition [Ben95, For03, Llo99, VdN11] or have quantified interference without providing a strong connection to speedup [BG06].

We consider quantum circuits composed of an initial state, followed by several unitary operators, and terminated by measurement of a Hermitian observable. The expectation value of this measurement can be written as a sum of Feynman-like paths in the computational basis, and this sum can be estimated via a Monte Carlo technique that considers an ensemble of paths drawn according to a suitable probability distribution. The required size of the ensemble is lower bounded by the square of the interference, which we define as a sum of absolute values of the path amplitudes (definition 3.3). We are not able to reach this lower bound, however by using a convex combination of a pair of Markov chains we are able to provide a simulation algorithm that runs in time quadratic in the product of the *interference producing capacities* of each operator in the circuit, defined as the largest amount of interference an operator is capable of producing (definition 3.5). This ends up being equal to the largest singular value of the entrywise absolute value of the operator in the computational basis. Briefly, we can estimate expressions of the form $\langle\psi|A\cdots Z|\phi\rangle$, of which quantum circuits $\langle\psi|U^{(1)\dagger}\cdots U^{(T)\dagger}MU^{(T)}\cdots U^{(1)}|\psi\rangle$ are a special case, in time proportional to $\|\bar{A}\|_2^2\cdots\|\bar{Z}\|_2^2$ where

$\|\cdot\|_2$ denotes maximum singular value and where a bar over an operator denotes entrywise absolute value in the computational basis. This work was inspired by, and extends, [VdN11] which provides an efficient simulation when $A, \ldots, Z$ are all sparse.

Operations with small interference producing capacity include the *efficiently computable sparse* operations as defined in [VdN11] (e.g. permutation matrices and gates acting on a constant number of qubits), as well as the Grover reflection operation, short time/low energy Hamiltonian evolutions, and the Haar wavelet transform. Our simulation algorithm will generally be exponentially slow in the length of the circuit, but for the classes of gates listed in the previous sentence has only polynomial dependence on the number of qubits. An example of a circuit that apparently uses much "quantum magic," but which can nevertheless be simulated in time polynomial in the number of qubits, is depicted in fig. 3.1.

We (of course) cannot efficiently simulate Shor's algorithm. However, replacing the Fourier transform by the Haar transform, which has low interference producing capacity, yields a circuit that we can simulate (fig. 3.2). We show that there is no quantum advantage for communication protocols that use small interference, although curiously this result does not apply to one-round communication protocols. To our knowledge, interference producing capacity is the first continuous-valued quantity that has been shown necessary for quantum speedup, escaping the theorem of [Nes12] which shows that a large class of continuous-valued quantities, such as entanglement and discord, are not necessary for quantum speedup.

In sections 3.3 and 3.4 we explain our method for estimating expectation values using a Monte Carlo technique with Markov chains. In section 3.5 we formalize and extend this technique and provide guarantees on runtime. In section 3.6 we characterize the types of quantum circuits that our technique can efficiently simulate, and explore a variety of circuits that we cannot efficiently simulate. Section 3.7 discusses further applications, including the Wigner representation and communication complexity. In section 3.8 we formalize our conjecture that interference, rather than interference producing capacity, is required for quantum speedup. Nontrivial proofs are deferred to appendices.

3.3 Monte Carlo technique

3.3.1 Sampling of paths

We will make use of the following circuit model. Let ρ be an initial density operator. This state is acted upon by a sequence of unitaries $U^{(1)}, \ldots, U^{(T)}$. Finally, a Hermitian observable (e.g. a projector) M is measured. It is not assumed that the unitary operations or the final observable are local, they can be arbitrary operations potentially involving all qubits or qudits (e.g. a quantum Fourier transform). The expectation value of this final measurement is

$$\mathrm{Tr}\left\{U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)} \rho\right\}. \tag{3.1}$$

Our goal is to estimate this expectation value to within small additive error, using a classical computer. We allow the unitaries to be oracle operations (as in Grover's algorithm), in which case we grant the classical computer that runs the simulation access to an equivalent oracle (this is further discussed in section 3.5.3).

This is not the most general type of simulation. In particular, we do not consider the case of a many-outcome measurement (e.g. individual measurements on several qubits, or a measurement given by a projective decomposition of the identity) in which the simulation is required to produce individual outcomes according to the same probability distribution with which the quantum circuit produces those outcomes. The ability to estimate the expectation value of a projector to within small multiplicative error would allow simulation of such sampling, as discussed in [TD04], however the algorithm of the present paper only estimates to within additive error.

Although our primary goal is to estimate expressions of the form (3.1), we generalize the task by considering products of the form $\mathrm{Tr}\{A^{(1)} \cdots A^{(S)} \sigma\}$ where σ and the $A^{(s)}$ are matrices, not necessarily unitary or Hermitian, and possibly rectangular (we label σ separately from the $A^{(s)}$ in anticipation of the results of the next section). This product can be written as a sum over paths,

$$\mathrm{Tr}\{A^{(1)} \cdots A^{(S)} \sigma\} = \sum_{i_0 \ldots i_S} A^{(1)}_{i_0 i_1} \cdots A^{(S)}_{i_{S-1} i_S} \sigma_{i_S i_0}. \tag{3.2}$$

Or, by defining the tuple index $\pi = (i_0 \ldots i_S)$, this can be written as

$$\mathrm{Tr}\{A^{(1)} \cdots A^{(S)} \sigma\} = \sum_{\pi} V(\pi) \tag{3.3}$$

$$V(\pi) = A^{(1)}_{i_0 i_1} \cdots A^{(S)}_{i_{S-1} i_S} \sigma_{i_S i_0}. \tag{3.4}$$

Our strategy is to estimate this sum by drawing a reasonably small number of paths π according to a probability distribution, denoted $R(\pi)$. Any probability distribution can be used, although some are more suitable than others. Finding a good $R(\pi)$ will be a central goal of this section and the next. Denote by Π a random variable that takes value π with probability $R(\pi)$. Consider the expectation value of $V(\Pi)/R(\Pi)$.

$$\mathbb{E}\left[\frac{V(\Pi)}{R(\Pi)}\right] = \sum_{\pi} \frac{V(\pi)}{R(\pi)} R(\pi) \tag{3.5}$$

$$= \sum_{\pi} V(\pi). \tag{3.6}$$

By the weak law of large numbers, $\sum_\pi V(\pi)$ can be approximated to arbitrary accuracy by computing the mean of sufficiently many samples of $V(\Pi)/R(\Pi)$, however the efficiency of this strategy hinges on two things. First, it must be possible using a classical computer to efficiently draw random samples according to the probability distribution $R(\pi)$ and to compute the corresponding values $V(\pi)/R(\pi)$. This is an important point that we will return to throughout the paper. Second, the sample mean of $V(\Pi)/R(\Pi)$ must rapidly converge to its expectation value. The Chernoff–Hoeffding bound states that for a random variable whose magnitude is bounded by b, the mean of $\mathrm{O}(\epsilon^{-2} b^2)$ samples is very likely to approximate the expectation value to within additive error ϵ. Thus there is rapid convergence when $\max_\pi \{|V(\pi)|/R(\pi)\}$ is small. Note that this is a sufficient but not necessary condition for rapid convergence, for example considering the variance of $V(\Pi)/R(\Pi)$ could in some cases reveal that convergence happens more rapidly.

We now present the Chernoff–Hoeffding bound in one of its standard forms, along with a corollary that adapts it to our application.

Theorem 3.1 (Chernoff–Hoeffding bound [Hoe63])**.** *Let $X_1, \ldots, X_K$ be independent identically distributed real-valued random variables with expectation value $\mathbb{E}[X]$ and satisfying $|X_k| \leq b$. Let $\epsilon > 0$. Then*

$$\Pr\left\{\left|\frac{1}{K}\sum_{k=1}^{K} X_k - \mathbb{E}[X]\right| > \epsilon\right\} \leq 2e^{-K\epsilon^2/2b^2}. \tag{3.7}$$

Corollary 3.2. *Let $V(\pi)$ be a complex valued function of π and $R(\pi)$ be a probability distribution. Define*

$$b_{max} = \max_{\pi}\left\{\frac{|V(\pi)|}{R(\pi)}\right\}. \tag{3.8}$$

Let $\epsilon, \delta > 0$. Then, with probability less than δ of exceeding the error bound, $\sum_\pi V(\pi)$ can be estimated to within additive error ϵ using $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2} b_{max}^2)$ draws from the distribution $R(\pi)$ and the same number of evaluations of $V(\pi)/R(\pi)$.

Proof. It can be shown[2] that theorem 3.1 can be extended to complex variables at the expense of replacing the right hand side of (3.7) by $4e^{-K\epsilon^2/4b^2}$. Define the independent identically distributed random variables $X_k = V(\Pi_k)/R(\Pi_k)$ with $k \in \{1, \ldots, K\}$. Applying the complex valued version of theorem 3.1, and noting that $|X_k| \leq b_{\max}$ and $\mathbb{E}\left[V(\Pi)/R(\Pi)\right] = \sum_\pi V(\pi)$, we get

$$\Pr\left\{\left|\frac{1}{K}\sum_{k=1}^{K}\frac{V(\Pi_k)}{R(\Pi_k)} - \sum_\pi V(\pi)\right| > \epsilon\right\} \leq 4e^{-K\epsilon^2/4b_{\max}^2}. \tag{3.9}$$

Setting $K = \ln(4/\delta)4\epsilon^{-2}b_{\max}^2 = \mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b_{\max}^2)$ makes the right hand side of (3.9) equal to δ. □

Since the number of samples needed depends only logarithmically on δ, it is possible to choose δ to be extremely small (say, one part in a billion) while having only minimal impact on the number of samples needed. With such a small δ, the estimate will be very likely to be within additive error ϵ.

The number of samples needed for an accurate estimate is quadratic in $b_{\max}$, so finding an $R(\pi)$ for which $b_{\max}$ is small is of crucial importance. However, feasibility of the simulation also depends on the difficulty of drawing random paths π according to the distribution $R(\pi)$ and computing the corresponding values $V(\pi)/R(\pi)$. We will denote by the letter f the time needed to carry out these operations. Specifically, we require that sampling from $R(\pi)$ and computing $V(\pi)/R(\pi)$ can be carried out in average time $\mathrm{O}(f)$ where f is some function of the dimension or number of qubits of a quantum circuit. Since $\sum_\pi V(\pi)$ can be estimated by averaging $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b_{\max}^2)$ samples of $V(\Pi)/R(\Pi)$, each of which can be computed in time $\mathrm{O}(f)$, the total runtime of the algorithm is $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b_{\max}^2 f)$.

Some probability distributions are easier to sample from than others, and this needs to be decided on a case by case basis. For example, consider $R(i) = |\psi_i|^2$ where $|\psi\rangle$ is a quantum state. If $|\psi\rangle$ is a computational basis state then $R(i)$ is rather trivial and can be sampled by simply outputting the sole index i for which $R(i) \neq 0$. If $|\psi\rangle$ is a graph state on n qubits then $R(i)$ is the uniform distribution over the 2^n basis states. This can be sampled in time $\mathrm{O}(n)$ by tossing a fair coin n times, once for each qubit, so in this case $f = n$. On the other hand, if $|\psi\rangle$ is defined as being the state just before the final measurement in Shor's algorithm, then it is probably not feasible to sample from $R(i)$ efficiently on a classical computer.

For simplicity we will assume that all operations can be carried out with perfect computational accuracy, including the degree to which the probability distribution of the generated samples π agrees with an ideal distribution $R(\pi)$, and the precision of the computed $V(\pi)/R(\pi)$ values. Of course, computers can only compute with finite precision. However, since we are concerned only with approximating expectation values to within additive error ϵ, carrying out the computations to finite but high precision is sufficient as long as the total accumulated computational error is small compared to the error tolerance ϵ. This is discussed in more detail in appendix A of [VdN11].

3.3.2 Interference

An efficient simulation requires choosing a probability distribution $R(\pi)$ for which $b_{\max}$ of (3.8) is not large. A tempting choice is

$$R_{\mathrm{opt}}(\pi) := \frac{|V(\pi)|}{\sum_{\pi'}|V(\pi')|}. \tag{3.10}$$

[2] This is shown by applying theorem 3.1 separately to the real and imaginary parts and using the fact that the sample mean is within additive error ϵ of the expectation value as long as both the real and imaginary parts are within $\epsilon/\sqrt{2}$.

It can be shown[3] that this is the unique distribution yielding the minimum possible value of $b_{\max}$,

$$b_{\text{opt}} = \sum_{\pi} |V(\pi)| \,. \tag{3.11}$$

Being lowest possible value of $b_{\max}$, (3.11) represents a lower bound on the number of samples needed as guaranteed by the Chernoff–Hoeffding bound, although a more careful analysis of variances, for instance, could show that the algorithm actually produces a faster than expected convergence.

An efficient algorithm requires both that $b_{\max}$ be small and that $R(\pi)$ can be sampled from efficiently. We do not know of a way to efficiently sample from the probability distribution (3.10) in general, so this is not useful for computing the expectation value. Nevertheless, it is worthwhile to discuss for a moment the case where the one condition is met (small $b_{\max}$) even if the other condition is not met (ability to efficiently draw samples). For concreteness, consider a simple quantum circuit with only one unitary, $\text{Tr}\{U^\dagger M U \rho\}$. This can be written as a sum over paths

$$\text{Tr}\{U^\dagger M U \rho\} = \sum_{\pi} V(\pi) \tag{3.12}$$

with $\pi = (i, j, k, l)$ and $V(i, j, k, l) = U^\dagger_{ij} M_{jk} U_{kl} \rho_{li}$. Plugging this into (3.11) gives

$$b_{\text{opt}} = \text{Tr}\{\bar{U}^\dagger \bar{M} \bar{U} \bar{\rho}\} \tag{3.13}$$

where a bar over a vector or matrix denotes entrywise absolute value in the computational basis, a notation that will be used throughout this paper. This generalizes in the obvious way for circuits with more than one unitary.

Comparing (3.11) and (3.12), both are sums over paths but the latter involves an absolute value for each path. The sum (3.12) has magnitude bounded by 1 if the observable M has eigenvalues bounded in magnitude by 1. The sum (3.11) on the other hand can take a much larger value than (3.12) when the terms in the latter sum exhibit cancellations due to destructive interference. For example, consider the case $|\psi\rangle = N^{-1/2} \sum_i |i\rangle$, U the Fourier transform, and M the identity, giving $b_{\text{opt}} = \sqrt{N}$.

It may be enlightening to consider a physical example. To this end, we introduce a simple toy-model version of Young's double-slit experiment. Let states $|0\rangle$ and $|1\rangle$ represent a particle immediately exiting the upper and lower slits, respectively, and let $|x\rangle$ represent a particle impacting the detector at position x. The transfer operator representing passage of the particle from the slits to the detector will be some unitary U satisfying $U(\alpha |0\rangle + \beta |1\rangle) = \int_x (\alpha\psi_x + \beta\phi_x) |x\rangle \, dx$. A particle passing through the upper slit will impact the detector at position x with probability density $|\psi_x|^2$; for a particle passing through the lower slit the probability density is $|\phi_x|^2$. A particle in a superposition of passing through upper and lower slits, in state $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, will impact the screen at x with probability density

$$\left| \frac{1}{\sqrt{2}} \psi_x + \frac{1}{\sqrt{2}} \phi_x \right|^2 = \frac{1}{2} |\psi_x|^2 + \frac{1}{2} |\phi_x|^2 + \text{Re}(\psi_x^* \phi_x). \tag{3.14}$$

The first two terms on the right hand side represent the probability that would be expected if the particle were in a classical stochastic mixture of passing through one slit or the other. The third is the interference term. Integrating this term over x yields zero, as it must in order for the probabilities to sum to 1. The total amount of interference can be quantified by instead integrating the absolute value of this term. Similarly, if we were interested in only part of the detector, say $x \in [0, 1]$, the interference associated with that region could be defined by integrating only over this

[3] Let $R(\pi)$ be any probability distribution that differs from $R_{\text{opt}}(\pi)$ of (3.10). Then there must be a π' such that $R(\pi') < R_{\text{opt}}(\pi')$. It follows that $\max_\pi \{|V(\pi)| / R(\pi)\} > |V(\pi')| / R_{\text{opt}}(\pi') = \sum_\pi |V(\pi)|$.

range. It turns out to be more mathematically convenient to include all three terms in the definition of interference; for one thing the resulting quantity will be multiplicative when considering a system composed of non-interacting subsystems. The $|\psi_x|^2/2 + |\phi_x|^2/2$ terms contribute at most 1 (exactly 1 if integrating over the entire range). In summary, we may define the interference associated with the $x \in [0,1]$ region of the detector as

$$\mathcal{I} = \int_{x\in[0,1]} \left(\frac{1}{2}|\psi_x|^2 + \frac{1}{2}|\phi_x|^2 + |\psi_x^* \phi_x| \right) dx. \tag{3.15}$$

This is essentially what is done in (3.13). Specifically, setting $\rho = |+\rangle\langle+|$ and $M = \int_{x\in[0,1]} |x\rangle\langle x|\, dx$ in (3.13) yields

$$b_{\text{opt}} = \int_{x\in[0,1]} \left(\frac{1}{\sqrt{2}}|\psi_x| + \frac{1}{\sqrt{2}}|\phi_x| \right)^2 dx \tag{3.16}$$

$$= \int_{x\in[0,1]} \left(\frac{1}{2}|\psi_x|^2 + \frac{1}{2}|\phi_x|^2 + |\psi_x^* \phi_x| \right) dx. \tag{3.17}$$

Note that (3.13) depends upon the choice of basis since the entrywise absolute value is basis dependent. Typically one has some canonical basis in mind, for example when one says that the double slit experiment exhibits interference this is relative to the position basis. For quantum circuits there is the computational basis, although in the interest of efficient simulation one may choose to use some other basis.

For a more complicated apparatus, such as a network of beam splitters, similar arguments apply: we quantify interference by computing a sum over paths, summing the absolute value of each path contribution. This definition depends upon a choice of course graining. For instance, a box which simply passes a photon from input to output undisturbed could be said to contribute no interference. On the other hand, if one were to take a more detailed view of this box—suppose for example that it contains a perfectly balanced Mach–Zehnder interferometer—then one could conclude that there is in fact interference. The same applies to simulation of quantum circuits. Although our simulation technique has difficulty simulating the Fourier transform, a Fourier transform followed by its inverse presents no difficulty if one course grains the circuit by replacing $F^\dagger F$ by the identity.

The above considerations lead to the following definition.

Definition 3.3. *The interference of a quantum circuit with initial state ρ, unitary operators $U^{(1)}, \ldots, U^{(T)}$, and measurement M is*

$$\mathcal{I}\left(U^{(1)\dagger}, \ldots, U^{(T)\dagger}, M, U^{(T)}, \ldots, U^{(1)}, \rho\right) = Tr\left\{\bar{U}^{(1)\dagger} \cdots \bar{U}^{(T)\dagger} \bar{M} \bar{U}^{(T)} \cdots \bar{U}^{(1)} \bar{\rho}\right\}. \tag{3.18}$$

More generally, the interference of an arbitrary expression of the form $Tr\{A^{(1)} \cdots A^{(S)} \sigma\}$ is

$$\mathcal{I}(A^{(1)}, \ldots, A^{(S)}, \sigma) = Tr\{\bar{A}^{(1)} \cdots \bar{A}^{(S)} \bar{\sigma}\}. \tag{3.19}$$

This definition depends on the choice of basis. Unless otherwise specified the standard (a.k.a. computational) basis is used.

With this definition, we have that $b_{\max} \geq \mathcal{I}(U^\dagger, M, U, \rho)$ in (3.8) for any choice of probability distribution, with equality when the distribution (3.10) is used. Since the number of samples needed to estimate the expectation value using our technique is proportional to $b_{\max}^2$, any quantum circuit with very large interference could never feasibly be simulated with our technique, no matter the choice of $R(\pi)$.

While we don't know how to efficiently sample from the optimal probability distribution (3.10), we conjecture that there is still some way to efficiently estimate the expectation value of a quantum

circuit in cases where the interference is low. The precise statement of this conjecture is a delicate matter taken up in section 3.8. We will however show, by the end of the next section, that it is possible to simulate circuits in which each unitary as well as the final observable has a low *interference producing capacity* (definition 3.5).

A connection between $\mathcal{I}$ and the decoherence functional of Gell-Mann and Hartle is discussed in section 3.7.4.

3.4 Markov chains

3.4.1 Introduction

The problem with the probability distribution (3.10) is that there is no obvious way to efficiently sample from it using a classical computer. So while only $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}\mathcal{I}^2)$ samples are needed (with $\mathcal{I}$ given by definition 3.3), each sample may be very complicated to evaluate. The essence of the difficulty is that this distribution treats the circuit holistically, so drawing samples apparently requires an understanding of how all the factors of (3.2) interact with each other. In order to avoid this problem we instead use a probability distribution defined in terms of a time-inhomogeneous Markov chain with a transition corresponding to each operator in (3.2). More precisely, we take the convex combination of two (unrelated) Markov chains, one proceeding left-to-right and the other proceeding right-to-left. This way, it is only necessary to understand each individual operator, not the interactions between operators. The computation time of this simulation will end up being related not to the interference $\mathcal{I}$ but rather the product of the interference producing capacities of each factor (a term that will be defined at the end of this section).

The end result of this section will be an algorithm for estimating products of the form $\mathrm{Tr}\{A^{(1)}\cdots A^{(S)}\sigma\}$ where σ and the $A^{(t)}$ are matrices, not necessarily unitary or Hermitian and possibly rectangular. This includes as a special case quantum circuits of the form $\mathrm{Tr}\{U^{(1)\dagger}\cdots U^{(T)\dagger}MU^{(T)}\cdots U^{(1)}\rho\}$. We build the algorithm step by step, considering first an example that demonstrates why a convex combination of probability distributions is needed, second an example that explains how the Markov chains are built, and finally using a convex combination of Markov chains. The exposition in this section is meant to be instructive; formal theorems will be taken up in section 3.5.

3.4.2 Inner product

Consider the task of estimating the inner product $\langle\psi|\phi\rangle = \sum_i \psi_i^* \phi_i$ where the two vectors satisfy the property $\|\psi\|_p = \|\phi\|_q = 1$ with $1/p + 1/q = 1$.[4] In the context of quantum circuits $p = q = 2$ is the natural choice; however, we allow general ℓ^p-norms because the case $p = 1$, $q = \infty$ is also important and because the general case may be of independent interest. Here, as in the more general case that will follow, the key is to find a probability distribution $R(i)$ that will be suitable for application of corollary 3.2. It is needed that

$$b_{\max} = \max_i \left\{ \frac{|V(i)|}{R(i)} \right\} = \max_i \left\{ \frac{|\psi_i^* \phi_i|}{R(i)} \right\} \tag{3.20}$$

is not large. There are two obvious choices for the probability distribution: $P(i) = |\psi_i|^p$ and $Q(i) = |\phi_i|^q$. Unfortunately, neither of these will guarantee a small $b_{\max}$. However, for each i at least one of the distributions $P(i)$ or $Q(i)$ will work well. The solution is to take a convex combination of these two distributions,

$$R(i) = \frac{1}{p} P(i) + \frac{1}{q} Q(i). \tag{3.21}$$

[4] The ℓ^p-norm, $\|\cdot\|_p$, is defined as $\|\psi\|_p = \left(\sum_i |\psi_i|^p\right)^{1/p}$ when $1 \le p < \infty$ and $\|\psi\|_p = \max_i |\psi_i|$ when $p = \infty$. When $1/p + 1/q = 1$, the norms $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual to each other.

The algorithm that follows is an adaptation of one that appears in [VdN11] (they used $p = q = 2$ and a slightly different technique). We present it as a formal theorem, in order to demonstrate how to carefully track the algorithm's time complexity.

Example 3.4. *Let $1 \leq p \leq \infty$ and $1/p + 1/q = 1$. Let $|\psi\rangle$ and $|\phi\rangle$ be vectors with $\||\psi\rangle\|_p = \||\phi\rangle\|_q = 1$. Suppose that it is possible to sample from the probability distributions $P(i) = |\psi_i|^p$ and $Q(i) = |\phi_i|^q$, and to compute entries ψ_i and ϕ_i, in average time $\mathrm{O}(f)$ for some f. It is possible, with probability less than $\delta > 0$ of exceeding the error bound, to estimate $\langle\psi|\phi\rangle$ to within additive error $\epsilon > 0$ in average time $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}f)$.*

Proof. Let $V(i) = \psi_i^* \phi_i$ and $R(i) = P(i)/p + Q(i)/q$. To apply corollary 3.2 we need to bound $b_{\max} = \max_i\{|V(i)|/R(i)\}$. Making use of the (weighted) inequality of arithmetic and geometric means,[5]

$$b_{\max} = \max_i\{|V(i)|/R(i)\} \tag{3.22}$$

$$= \max_i\{|\psi_i^*\phi_i|/[P(i)/p + Q(i)/q]\} \tag{3.23}$$

$$\leq \max_i\{|\psi_i^*\phi_i|/[P(i)^{1/p}Q(i)^{1/q}]\} \tag{3.24}$$

$$= 1. \tag{3.25}$$

By corollary 3.2, $\langle\psi|\phi\rangle = \sum_i V(i)$ can be estimated at the cost of drawing $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2})$ samples i according to $R(i)$ and computing the corresponding $V(i)/R(i)$ values. Sampling from $R(i)$ can be accomplished as follows: flip a biased coin that lands heads with probability $1/p$. If it lands heads then draw i from $P(i)$, otherwise draw i from $Q(i)$. By assumption this takes average time $\mathrm{O}(f)$. Next, $V(i)/R(i)$ can be computed directly from ψ_i and ϕ_i, each of which can in turn be computed in average time $\mathrm{O}(f)$. The $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2})$ samples (as well as their mean) can therefore be computed in average time $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}f)$. □

3.4.3 Nearly stochastic matrices

We now move to a more general case, estimation of $\langle\psi|A^{(1)}\cdots A^{(S)}|\phi\rangle$. For the sake of simplicity, suppose that there are only two operators (i.e. $S = 2$) so that the goal is to estimate $\langle\psi|AB|\phi\rangle$. This can be written as a sum over paths as in (3.2),

$$\langle\psi|AB|\phi\rangle = \sum_{ijk} \psi_i^* A_{ij} B_{jk} \phi_k. \tag{3.26}$$

To apply corollary 3.2 to this problem, set $\pi = (i, j, k)$ and $V(i, j, k) = \psi_i^* A_{ij} B_{jk} \phi_k$. For efficient simulation it suffices to find a probability distribution $P(i, j, k)$ from which we can efficiently draw samples using a classical computer, for which $V(i, j, k)/P(i, j, k)$ can be efficiently computed, and for which

$$b_{\max} = \max_{ijk}\left\{\frac{|\psi_i^* A_{ij} B_{jk} \phi_k|}{P(i, j, k)}\right\} \tag{3.27}$$

is small enough that the estimation will converge reasonably fast. As discussed in the previous section, a tempting choice for the probability distribution is given by (3.10), however it is not clear how one would efficiently draw samples from this since doing so apparently requires an understanding of how $\langle\psi|$, A, B, and $|\phi\rangle$ interact with each other. To avoid this problem we define $P(i, j, k)$ in terms of a time-inhomogeneous Markov chain,

$$P(i, j, k) = P_\psi(i) P_A(j|i) P_B(k|j), \tag{3.28}$$

[5] The weighted inequality of arithmetic and geometric means is a generalization of the more familiar inequality $x/2 + y/2 \geq \sqrt{xy}$. If $1 \leq p \leq \infty$ and $1/p + 1/q = 1$ then $x/p + y/q \geq x^{1/p} y^{1/q}$.

with each transition depending on only one of the components of $\langle\psi|AB|\phi\rangle$. Plugging this into (3.27) gives

$$b_{\max} = \max_{ijk}\left\{\frac{|\psi_i^* A_{ij} B_{jk}\phi_k|}{P_\psi(i)P_A(j|i)P_B(k|j)}\right\} \tag{3.29}$$

$$= \max_{ijk}\left\{\frac{|\psi_i^*|}{P_\psi(i)}\cdot\frac{|A_{ij}|}{P_A(j|i)}\cdot\frac{|B_{jk}|}{P_B(k|j)}\cdot|\phi_k|\right\} \tag{3.30}$$

$$\leq \max_i\left\{\frac{|\psi_i^*|}{P_\psi(i)}\right\}\max_{ij}\left\{\frac{|A_{ij}|}{P_A(j|i)}\right\}\max_{jk}\left\{\frac{|B_{jk}|}{P_B(k|j)}\right\}\max_k\left\{|\phi_k|\right\}. \tag{3.31}$$

The goal is then to find $P_\psi(i)$, $P_A(j|i)$, and $P_B(k|j)$ that minimize the terms of (3.31). Consider first the case where $\langle\psi|$ is a probability distribution, the matrices A and B are right-stochastic matrices,[6] and $|\phi\rangle$ has small entries (say, $\|\phi\|_\infty \leq 1$). We can set $P_\psi(i) = \psi_i$, $P_A(j|i) = A_{ij}$, and $P_B(k|j) = B_{jk}$, with the result that each factor in (3.31) is bounded by 1. If $|\phi\rangle$ is not a probability distribution, we can turn it into one by defining $P_\psi(i) = |\psi_i| / \|\psi\|_1$, similarly if A is not a right-stochastic matrix we can set $P_A(j|i) = |A_{ij}| / \sum_{j'} |A_{ij'}|$ (and likewise for B). Then (3.31) becomes

$$b_{\max} \leq \|\psi\|_1 \max_i\left\{\sum_{j'}|A_{ij'}|\right\}\max_j\left\{\sum_{k'}|B_{jk'}|\right\}\|\phi\|_\infty \tag{3.32}$$

$$= \|\psi\|_1\|\bar{A}\|_\infty\|\bar{B}\|_\infty\|\phi\|_\infty. \tag{3.33}$$

Here, as in the rest of the paper, we use the induced norm for operators: $\|A\|_p = \max_{\boldsymbol{u}}\|A\boldsymbol{u}\|_p/\|\boldsymbol{u}\|_p$ (we do not use the entrywise or Schatten norms). Under this notation, $\|M\|_2$ is the largest singular value of M, $\|M\|_1$ is the maximum absolute column sum, and $\|M\|_\infty$ is the maximum absolute row sum. By corollary 3.2, the value of $\langle\psi|AB|\phi\rangle$ can be estimated by drawing

$$\mathrm{O}\left(\log(\delta^{-1})\epsilon^{-2}b_{\max}^2\right) \leq \mathrm{O}\left(\log(\delta^{-1})\epsilon^{-2}\|\psi\|_1^2\|\bar{A}\|_\infty^2\|\bar{B}\|_\infty^2\|\phi\|_\infty^2\right) \tag{3.34}$$

samples (i,j,k) from $P(i,j,k)$ and averaging the corresponding $V(i,j,k)/P(i,j,k)$.

3.4.4 General p, q

In the case of quantum circuits, it is the ℓ^2-norm that is relevant. Instead of $b_{\max} \leq \|\psi\|_1\|\bar{A}\|_\infty\|\bar{B}\|_\infty\|\phi\|_\infty$ from the previous example, we want $b_{\max} \leq \|\psi\|_2\|\bar{A}\|_2\|\bar{B}\|_2\|\phi\|_2$. For the sake of generality, we allow arbitrary p, q satisfying $1/p + 1/q = 1$. The goal is to find a probability distribution that yields $b_{\max} \leq \|\psi\|_p\|\bar{A}\|_q\|\bar{B}\|_q\|\phi\|_q$. As in section 3.4.2, the way to proceed is by taking a convex combination of two probability distributions, $R(i,j,k) = P(i,j,k)/p + Q(i,j,k)/q$. Here $P(i,j,k)$ will be a time-inhomogeneous Markov chain proceeding in the $i \to j \to k$ direction and $Q(i,j,k)$ a different Markov chain proceeding in the $k \to j \to i$ direction. Again the inequality of arithmetic and geometric means plays a crucial role, giving

$$R(i,j,k) = P(i,j,k)/p + Q(i,j,k)/q \tag{3.35}$$

$$\geq P(i,j,k)^{1/p}Q(i,j,k)^{1/q} \tag{3.36}$$

$$= \left[P_\psi(i)P_A(j|i)P_B(k|j)\right]^{1/p}\left[Q_A(i|j)Q_B(j|k)Q_\phi(k)\right]^{1/q}. \tag{3.37}$$

[6] A right-stochastic matrix is a nonnegative matrix with each row summing to 1, a left-stochastic matrix has each column summing to 1. We do not require stochastic matrices to be square.

With this we have

$$b_{\max} = \max_{ijk}\left\{\frac{|\psi_i^* A_{ij} B_{jk} \phi_k|}{R(i,j,k)}\right\} \tag{3.38}$$

$$\leq \max_{ijk}\left\{\frac{|\psi_i^* A_{ij} B_{jk} \phi_k|}{P(i,j,k)^{1/p} Q(i,j,k)^{1/q}}\right\} \tag{3.39}$$

$$= \max_{ijk}\left\{\frac{|\psi_i^*|}{P_\psi(i)^{1/p}} \cdot \frac{|A_{ij}|}{P_A(j|i)^{1/p} Q_A(i|j)^{1/q}} \cdot \frac{|B_{jk}|}{P_B(k|j)^{1/p} Q_B(j|k)^{1/q}} \cdot \frac{|\phi_k|}{Q_\phi(k)^{1/q}}\right\} \tag{3.40}$$

$$\leq \max_i\left\{\frac{|\psi_i^*|}{P_\psi(i)^{\frac{1}{p}}}\right\} \max_{ij}\left\{\frac{|A_{ij}|}{P_A(j|i)^{\frac{1}{p}} Q_A(i|j)^{\frac{1}{q}}}\right\} \max_{jk}\left\{\frac{|B_{jk}|}{P_B(k|j)^{\frac{1}{p}} Q_B(j|k)^{\frac{1}{q}}}\right\} \max_k\left\{\frac{|\phi_k|}{Q_\phi(k)^{\frac{1}{q}}}\right\} \tag{3.41}$$

$$= b_\psi b_A b_B b_\phi, \tag{3.42}$$

where b_ψ, b_A, b_B, and b_ϕ label the four factors of (3.41). By corollary 3.2, the number of samples needed in order to estimate $\langle\psi|AB|\phi\rangle$ is $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2} b_\psi^2 b_A^2 b_B^2 b_\phi^2)$. The quantities b_ψ, b_A, b_B, and b_ϕ are therefore identified as being the simulation cost due to each of the components of $\langle\psi|AB|\phi\rangle$. We show in appendix 3.A (theorem 3.22) that for any choice of probability distribution $b_A \geq \|\bar{A}\|_q$ and that there are optimal probability distributions achieving $b_A = \|\bar{A}\|_q$ (and similarly for B, ψ, and ϕ). Using these gives

$$b_{\max} \leq \|\psi\|_p \|\bar{A}\|_q \|\bar{B}\|_q \|\phi\|_q. \tag{3.43}$$

Whether these optimal probability distributions can be efficiently sampled from is a matter that needs to be considered on a case by case basis, however we show in section 3.6 that this is indeed the case for a wide range of matrices, both unitary and Hermitian. Additionally, in terms of query complexity rather than time complexity these efficient sampling requirements can for the most part be ignored, as we will discuss further in section 3.5.3.

3.4.5 Dyads and density operators

It is possible to further generalize to expressions of the form $\mathrm{Tr}\{AB\sigma\}$. The special case $\langle\psi|AB|\phi\rangle$ is obtained by setting $\sigma = |\phi\rangle\langle\psi|$. The above derivation is easily adapted by writing σ_{ki}, $P_\sigma(i)$, and $Q_\sigma(k)$ instead of $\phi_k \psi_i^*$, $P_\psi(i)$ and $Q_\phi(k)$. With these substitutions, (3.38)-(3.42) become

$$b_{\max} = \max_{ijk}\left\{\frac{|A_{ij} B_{jk} \sigma_{ki}|}{R(i,j,k)}\right\} \tag{3.44}$$

$$\leq \max_{ijk}\left\{\frac{|A_{ij}|}{P_A(j|i)^{1/p} Q_A(i|j)^{1/q}} \cdot \frac{|B_{jk}|}{P_B(k|j)^{1/p} Q_B(j|k)^{1/q}} \cdot \frac{|\sigma_{ki}|}{P_\sigma(i)^{1/p} Q_\sigma(k)^{1/q}}\right\} \tag{3.45}$$

$$\leq \max_{ij}\left\{\frac{|A_{ij}|}{P_A(j|i)^{1/p} Q_A(i|j)^{1/q}}\right\} \max_{jk}\left\{\frac{|B_{jk}|}{P_B(k|j)^{1/p} Q_B(j|k)^{1/q}}\right\} \max_{ki}\left\{\frac{|\sigma_{ki}|}{P_\sigma(i)^{1/p} Q_\sigma(k)^{1/q}}\right\} \tag{3.46}$$

$$= b_A b_B b_\sigma. \tag{3.47}$$

The b_σ factor differs from the other two in that the probability distributions are not conditional. This stems from the fact that σ represents the starting point of the Markov chains. If $\sigma = |\phi\rangle\langle\psi|$ then taking the probability distributions $P_\sigma(i) = |\psi_i|^p / \|\psi\|_p$ and $Q_\sigma(k) = |\phi_k|^q / \|\phi\|_q$ gives $b_\sigma = \|\psi\|_p \|\phi\|_q$ as in (3.43). If $p = q = 2$ and if σ is a density operator (positive semidefinite and trace 1) then taking the probability distributions $P_\sigma(i) = Q_\sigma(i) = \sigma_{ii}$ gives $b_\sigma = 1$ due to the inequality $|\sigma_{ki}| \leq \sqrt{\sigma_{kk}\sigma_{ii}}$, which is satisfied by positive semidefinite matrices.

3.4.6 Interference producing capacity

In section 3.3.2 we interpreted the lowest possible $b_{\max}$ value, obtained by using the holistic probability distribution (3.10), as being the interference of a quantum circuit. Although this probability distribution achieves the lowest $b_{\max}$, there is no clear way to draw samples efficiently and for this reason the Markov chain technique of this section was developed. The result was a strategy that depends only on properties of the individual operators rather than on the expression as a whole. The $b_{\max}$ value for this strategy is upper bounded by (3.47).

Consider now the minimum possible value of one of the factors in (3.47), for instance b_A. In appendix 3.A (theorem 3.22) we will show that the best possible choice of $P_A(j|i)$ and $Q_A(i|j)$ yields $b_A = \|\bar{A}\|_q$. In the case of quantum circuits the relevant norm is $p = q = 2$, so this becomes[7]

$$b_A = \|\bar{A}\|_2. \tag{3.48}$$

This can be interpreted in terms of interference: it is the largest possible contribution A can make to the interference $\mathcal{I}$ of definition 3.3. Specifically, since $\|\cdot\|_2$ gives the maximum singular value of its argument, we have

$$\mathcal{I}(A^{(1)}, \ldots, A^{(S)}, |\phi\rangle\langle\psi|) \leq \|\bar{A}^{(1)}\|_2 \cdots \|\bar{A}^{(S)}\|_2 \|\phi\|_2 \|\psi\|_2. \tag{3.49}$$

Furthermore, for any operator A we have

$$\max_{\|\psi\|_2 = \|\phi\|_2 = 1} \mathcal{I}(A, |\phi\rangle\langle\psi|) = \|\bar{A}\|_2. \tag{3.50}$$

For this reason, we interpret $\|\bar{A}\|_2$ as being the interference producing capacity of A.[8]

Definition 3.5. *The interference producing capacity of a matrix A is*

$$\mathcal{I}_{\max}(A) = \|\bar{A}\|_2. \tag{3.51}$$

This definition, like definition 3.3, is basis dependent. Here the basis dependence arises from the entrywise absolute value. Unless otherwise specified, we will work in the computational basis. In the next sections we will show the product of the $\mathcal{I}_{\max}$ values for the operations and final measurement of a circuit to be a necessary resource for quantum speedup: if this quantity is low then a circuit can be classically simulated. The same claim applies also for other bases, and even for more exotic representations (as we will show in section 3.7.1). The situation is not so much different from, for instance, Gottesman–Knill theorem which claims that stabilizer circuits may be efficiently simulated [Got98]. Although a circuit may at first not appear to be a stabilizer circuit it may be so after a change of basis (i.e. after conjugating the initial state, all unitary operations, and all measurements by some unitary).

The $\mathcal{I}_{\max}$ value for various operators is listed in table 3.1. As was shown informally in this section, and more formally in the next section, it is possible to efficiently simulate quantum circuits when the product of the $\mathcal{I}_{\max}$ values of all operators is not large. So, one may interpret a small $\mathcal{I}_{\max}$ value to mean that a unitary operator contributes only minimally to quantum speedup. On the high end of the table are the Fourier and Hadamard transforms, having the maximum possible value of $\mathcal{I}_{\max}$; these are difficult for us to simulate (at least in the computational basis). On the low end are the Pauli and the permutation matrices, having $\mathcal{I}_{\max} = 1$; these contribute nothing to quantum speedup (relative to our simulation scheme). Among unitaries, the only operators with $\mathcal{I}_{\max} = 1$ are permutations with phases, $U = \sum_j e^{i\theta_j} |\sigma(j)\rangle\langle j|$.

[7] We focus here on the case $p = 2$ of relevance to quantum circuits, although the entire subsection could easily be generalized to $p \neq 2$.

[8] Our measure of interference is different from, and seemingly unrelated to, the one defined in [BG06], which in the case of unitary matrices reduces to $N - \sum_{ij} |U_{ij}|^4$.

Matrix	$\mathcal{I}_{\max}$
Fourier or Hadamard transform on n qubits	$2^{n/2}$
Arbitrary gate on n qudits	no more than $d^{n/2}$
Haar wavelet transform on n qubits	$\sqrt{1+n}$
k-sparse unitary	no more than $\sqrt{k}$
Grover reflection	$\mathcal{I}_{\max} \to 3$ as $n \to \infty$
Permutation in computational basis	1
Pauli matrices	1
Rank one projector	1

Table 3.1: The $\mathcal{I}_{\max}$ value for various matrices. Operators with larger $\mathcal{I}_{\max}$ value are harder to simulate using our technique. Proofs for the nontrivial cases are presented in appendix 3.C.

3.5 EPS and EHT operators

3.5.1 Definitions

We will now present two definitions codifying the requirements operators must meet in order that products of the form $\mathrm{Tr}\{A^{(1)} \cdots A^{(S)} \sigma\}$ can be estimated using the techniques of the previous section. In the previous section, using a pair of Markov chains yielded a simulation strategy in which each component of $\mathrm{Tr}(AB\sigma)$ can be treated independently, with A, B, and σ contributing costs b_A, b_B, and b_σ to the total number of samples needed as per (3.47). Each sample requires drawing a random path according to the distribution $R(i,j,k)$ and then computing $V(i,j,k)/R(i,j,k)$. Drawing the random path requires considering only one operator at a time since $R(i,j,k)$ is defined in terms of Markov chains. Similarly, computing $V(i,j,k)/R(i,j,k)$ can be done considering one operator at a time since

$$\frac{V(i,j,k)}{R(i,j,k)} = \frac{A_{ij} B_{jk} \sigma_{ki}}{P(i,j,k)/p + Q(i,j,k)/q} \tag{3.52}$$

$$= \left\{ \frac{1}{p} \frac{P(i,j,k)}{A_{ij} B_{jk} \sigma_{ki}} + \frac{1}{q} \frac{Q(i,j,k)}{A_{ij} B_{jk} \sigma_{ki}} \right\}^{-1} \tag{3.53}$$

$$= \left\{ \frac{1}{p} \frac{P_A(j|i)}{A_{ij}} \frac{P_B(k|j)}{B_{jk}} \frac{P_\sigma(i)}{\sigma_{ki}} + \frac{1}{q} \frac{Q_A(i|j)}{A_{ij}} \frac{Q_B(j|k)}{B_{jk}} \frac{Q_\sigma(k)}{\sigma_{ki}} \right\}^{-1}. \tag{3.54}$$

Focusing on a single component, say A, conditions for efficient simulation can be identified (note that σ requires slightly different conditions, which we deal with later). First, the quantity b_A of (3.47) should be small in order that the number of samples required be small. Second, it must be possible to efficiently sample from the probability distributions $P_A(j|i)$ and $Q_A(i|j)$ and to compute the contributions due to A in (3.54), namely $P_A(j|i)/A_{ij}$ and $Q_A(i|j)/A_{ij}$. We express these conditions as a definition. However, it will be useful to generalize by allowing an extra index k in the definition below (not related to the k that appears above). If k takes only a single value (say, $k = 0$) the definition below exactly encompasses the conditions outlined above. The extra freedom granted by k will allow, as we will show shortly, treatment of sums, products, and exponentials of matrices (theorem 3.10). In the case $p = 1$, $q = \infty$ it was the matrices resembling stochastic matrices that could be efficiently simulated. For this reason, for general p, q we give the name *efficient pseudo-stochastic* (EPS) to matrices that we can efficiently simulate.

Definition 3.6 (EPS)**.** *Let $1 \le p \le \infty$, $1/p + 1/q = 1$, and $b < \infty$. An $M \times N$ matrix A is $\mathrm{EPS}_p(b, f)$ if there is a finite or countable set K, values $\alpha_{mnk} \in \mathbb{C}$, and conditional probability distributions $P(n,k|m)$ and $Q(m,k|n)$ with $m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, and $k \in K$, satisfying the following conditions:*

(a) $\sum_{k\in K} \alpha_{mnk} = A_{mn}$.[9]

(b)

$$\max_{mnk} \left\{ \frac{|\alpha_{mnk}|}{P(n,k|m)^{1/p} Q(m,k|n)^{1/q}} \right\} \leq b, \tag{3.55}$$

with the convention that $0/0 = 0$.

(c) Given any m*, it is possible in average time* $\mathrm{O}(f)$ *on a classical computer to sample* n, k *from the probability distribution* $P(n,k|m)$ *and then compute* $\alpha_{mnk}/P(n,k|m)$ *and* $\alpha_{mnk}/Q(m,k|n)$.

(d) Given any n*, it is possible in average time* $\mathrm{O}(f)$ *on a classical computer to sample* m, k *from the probability distribution* $Q(m,k|n)$ *and then compute* $\alpha_{mnk}/P(n,k|m)$ *and* $\alpha_{mnk}/Q(m,k|n)$.

This definition is related to interference producing capacity in the following way. It is always possible to satisfy conditions (a) and (b) with $b = \|\bar{A}\|_q$, and it is impossible to do better. This is proved in appendix 3.A. So, for the case $p = q = 2$ the optimal value of b is equal to the interference producing capacity of A. Since b (multiplied for all operators in a circuit) determines how many samples will be required for our simulation technique, this connects interference producing capacity to difficulty of simulation.

Although conditions (a) and (b) can always be satisfied with $b = \|\bar{A}\|_q$ for some α_{mnk}, $P(n,k|m)$, and $Q(m,k|n)$, it could be the case that these do not satisfy (c) and (d). In other words, it may be time consuming to sample from these probability distributions. An example would be a permutation matrix $A|x\rangle = |g(x)\rangle$. Such a matrix has $\|\bar{A}\|_q = 1$, so it has no interference producing capacity. Nevertheless, it would be difficult to simulate if the function g were difficult to calculate. In some sense (c) and (d) constitute a requirement that the matrix A be well understood from a computational perspective. In practice, (c) and (d) have not presented an obstacle for any of the operators that we have considered. If one is concerned with query complexity rather than time complexity then (c) and (d) can mostly be ignored. This will be explored in section 3.5.3.

There is a subtlety in conditions (c) and (d) that deserves discussion. It is required that the operations be carried out in *average* time $\mathrm{O}(f)$. It is allowed that $\alpha_{mnk}/P(n,k|m)$ and $\alpha_{mnk}/Q(m,k|n)$ be difficult to compute for some m, n, k triples as long as those occur rarely when sampling from $P(n,k|m)$ or $Q(m,k|n)$. In our implementation of exponentials of operators (theorem 3.10(c)) the time required is proportional to k, and so is unbounded since $k \in \{0, 1, \dots\}$, however $P(n,k|m)$ and $Q(m,k|n)$ decay exponentially in k so the average time is small.

We now present a definition that embodies the conditions σ must satisfy in order to yield an efficient simulation. Looking to (3.46) and (3.54), the difference between the factors relating to σ and those relating to A are that the latter involve conditional probability distributions. This stems from the fact that the Markov chains begin at σ and so have no index to condition upon. With this difference in mind, we provide a definition analogous to definition 3.6 but with non-conditional probability distributions. Since the Markov chains begin and end at σ, we name the suitable matrices *efficient head/tail* (EHT) matrices.

Definition 3.7 (EHT)**.** *Let* $1 \leq p \leq \infty$*,* $1/p + 1/q = 1$*, and* $b < \infty$*. An* $M \times N$ *matrix* σ *is* $\mathrm{EHT}_p(b, f)$ *if there is a finite or countable set* K*, values* $\alpha_{mnk} \in \mathbb{C}$*, and probability distributions* $P(n,k)$ *and* $Q(m,k)$ *with* $m \in \{1, \dots, M\}$*,* $n \in \{1, \dots, N\}$*, and* $k \in K$*, satisfying the following conditions:*

(a) $\sum_{k\in K} \alpha_{mnk} = \sigma_{mn}$.

(b)

$$\max_{mnk} \left\{ \frac{|\alpha_{mnk}|}{P(n,k)^{1/p} Q(m,k)^{1/q}} \right\} \leq b, \tag{3.56}$$

[9] We show in appendix 3.B (lemma 3.30) that this series converges absolutely, so there is no ambiguity regarding the way that an infinite K is enumerated.

with the convention that $0/0 = 0$.

(c) It is possible in average time $\mathrm{O}(f)$ *on a classical computer to sample* n, k *from the probability distribution* $P(n,k)$ *and then, given any* $m \in \{1,\dots,M\}$ *to compute* $\alpha_{mnk}/P(n,k)$ *and* $\alpha_{mnk}/Q(m,k)$.

(d) It is possible in average time $\mathrm{O}(f)$ *on a classical computer to sample* m, k *from the probability distribution* $Q(m,k)$ *and then, given any* $n \in \{1,\dots,N\}$ *to compute* $\alpha_{mnk}/P(n,k)$ *and* $\alpha_{mnk}/Q(m,k)$.

This definition does not relate to interference. For the case of quantum circuits we can assume σ to be a density operator. In section 3.5.4 we show that for density operators it is always possible to achieve $b = 1$ in the above definition as long as one can simulate measurements in the computational basis and compute individual matrix entries in average time $\mathrm{O}(f)$.

The definition of EHT is more strict than that of EPS: any EHT operator can be seen to also be EPS by using the probability distributions $P(n,k|m) = P(n,k)$ and $Q(m,k|n) = Q(m,k)$. Therefore, since it is not possible to have $b < \|\bar{A}\|_q$ for EPS operators, it is also not possible to have $b < \|\bar{\sigma}\|_q$ for EHT operators. As mentioned above, in the case of EPS it is always possible to satisfy conditions (a) and (b) with $b = \|\bar{A}\|_q$, however since EHT is more strict there are operators σ for which it is not possible to have $b = \|\bar{\sigma}\|_q$. Theorem 3.22(d) in appendix 3.A gives that $b = \|\bar{\sigma}\|_{\mathrm{Tr}}$ is possible when $p = q = 2$ where $\|\cdot\|_{\mathrm{Tr}}$ is the trace norm (and a generalization is provided for $p \neq 2$).

In section 3.6 we will consider the case $p = q = 2$, which is the norm relevant to quantum circuits, and give several examples of states that are $\mathrm{EHT}_2(b,f)$ and operators that are $\mathrm{EPS}_2(b,f)$ where b is small and f is polynomial in the number of qubits (or polylog in the dimension of the system). Expectation values of circuits built from such states and operators can be efficiently simulated. Specifically, we have the following theorem, the central theorem of this paper, whose proof will be deferred until after lemma 3.11.

Theorem 3.8 (Efficient simulation)**.** *Let* σ *be* $\mathrm{EHT}_p(b_\sigma, f_\sigma)$ *and for* $t \in \{1,\dots,S\}$ *let* $A^{(t)}$ *be* $\mathrm{EPS}_p(b_t, f_t)$. *Then, with probability less than* $\delta > 0$ *of exceeding the error bound,* $Tr\{A^{(1)} \cdots A^{(S)} \sigma\}$ *can be estimated to within additive error* $\epsilon > 0$ *in average time* $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2} b^2 f)$ *where* $b = b_\sigma \prod_t b_t$ *and* $f = f_\sigma + \sum_t f_t$.

3.5.2 Operations that preserve EPS/EHT properties

We now discuss mathematical operations that preserve the EPS and EHT properties. These include scaling, transpose, adjoint, multiplication, addition, and exponentiation (theorems 3.9 and 3.10). The first three follow immediately from the definitions, so the following theorem is presented without proof.

Theorem 3.9. *Let* A *be* $\mathrm{EPS}_p(b,f)$ *and* σ *be* $\mathrm{EHT}_p(b,f)$. *Let* $s \in \mathbb{C}$ *be a scalar. Then*

(a) σ *is* $\mathrm{EPS}_p(b,f)$.

(b) sA *is* $\mathrm{EPS}_p(|s|\, b, f)$.

(c) $s\sigma$ *is* $\mathrm{EHT}_p(|s|\, b, f)$.

(d) $A^\top$ *and* $A^\dagger$ *are* $\mathrm{EPS}_p(b,f)$.

(e) $\sigma^\top$ *and* $\sigma^\dagger$ *are* $\mathrm{EHT}_p(b,f)$.

The presence of the k index in definition 3.6 allows treatment of sums and products of operators. Consider for instance the product AB. The two factors of (3.45) relating to A and B can be combined

to match the conditions of definition 3.6 as follows. Begin by relabeling the indices of (3.45) from i, j, k to m, k, n and proceed as follows,

$$b_{\max} \leq \max_{mnk} \left\{ \frac{|\sigma_{nm}|}{P_\sigma(m)^{1/p} Q_\sigma(n)^{1/q}} \cdot \frac{|A_{mk}|}{P_A(k|m)^{1/p} Q_A(m|k)^{1/q}} \cdot \frac{|B_{kn}|}{P_B(n|k)^{1/p} Q_B(k|n)^{1/q}} \right\} \tag{3.57}$$

$$\leq \max_{mn} \left\{ \frac{|\sigma_{nm}|}{P_\sigma(m)^{1/p} Q_\sigma(n)^{1/q}} \right\} \cdot \max_{mnk} \left\{ \frac{|A_{mk}|}{P_A(k|m)^{1/p} Q_A(m|k)^{1/q}} \cdot \frac{|B_{kn}|}{P_B(n|k)^{1/p} Q_B(k|n)^{1/q}} \right\} \tag{3.58}$$

$$\leq \max_{mn} \left\{ \frac{|\sigma_{nm}|}{P_\sigma(m)^{1/p} Q_\sigma(n)^{1/q}} \right\} \cdot \max_{mnk} \left\{ \frac{|A_{mk} B_{kn}|}{[P_A(k|m) P_B(n|k)]^{1/p} [Q_A(m|k) Q_B(k|n)]^{1/q}} \right\} \tag{3.59}$$

$$= b_\sigma b_{AB}. \tag{3.60}$$

Defining $P_{AB}(n, k|m) = P_A(k|m) P_B(n|k)$, $Q_{AB}(m, k|n) = Q_B(k|n) Q_A(m|k)$, and $\alpha_{mnk} = A_{mk} B_{kn}$, the b_{AB} factor reduces to

$$b_{AB} = \max_{mnk} \left\{ \frac{|\alpha_{mnk}|}{P_{AB}(n, k|m)^{1/p} Q_{AB}(m, k|n)^{1/q}} \right\}. \tag{3.61}$$

This resembles the factors involving A or B that appear in (3.46) but with the addition of an extra index k appearing in both the numerator and in the probability distributions. Allowing such an extra index enables treatment of AB in the same manner as the individual factors A and B. This is formalized by theorem 3.10(b) below, which states that the product of EPS matrices is EPS. In the general case this procedure is slightly complicated by the fact that A and B may in turn have their own extra indices k' and k'', which must be inherited by the product AB.

Sums are handled in a similar way. An expression such as $\mathrm{Tr}((A + B)\sigma)$ is estimated by using A for a fraction of the samples and B for the remainder. This works since $\mathrm{Tr}((A + B)\sigma)$ is twice the average of $\mathrm{Tr}(A\sigma)$ and $\mathrm{Tr}(B\sigma)$. The k index is used to randomly choose between A or B for each sample. Exponentials are treated by applying these sum and product rules to $e^A = \sum_{j=0}^{\infty} A^j / j!$.

Theorem 3.10 (Operations on EPS)**.** *Let A be a matrix that is $\mathrm{EPS}_p(b_A, f_A)$ and let B be a matrix that is $\mathrm{EPS}_p(b_B, f_B)$. Then, assuming in each case that A and B have a compatible number of rows and columns, the following hold.*

(a) $A + B$ is $\mathrm{EPS}_p(b_A + b_B, \max\{f_A, f_B\})$.

(b) AB is $\mathrm{EPS}_p(b_A b_B, f_A + f_B)$.

(c) e^A is $\mathrm{EPS}_p(e^b, bf)$.

Proof. The proofs are in appendix 3.B. Rule (a) is a special case of theorem 3.31, which treats finite or infinite linear combinations. □

Since the value b in definition 3.6 (with $p = q = 2$) is lower bounded by interference producing capacity $\mathcal{I}_{\max}$, theorem 3.10 has the following interpretation. By (a), $\mathcal{I}_{\max}$ is convex. By (b), it is sub-multiplicative. By (c), the interference producing capacity of a Hamiltonian evolution e^{iHt} is at most exponential in $t\mathcal{I}_{\max}(H)$.

We now prove theorem 3.8, regarding estimation of $\mathrm{Tr}\{A^{(1)} \cdots A^{(S)} \sigma\}$. While this can be proved directly using Markov chains, as was done in section 3.4, this would be notationally tedious. It is much easier to first repeatedly apply the product rule, theorem 3.10(b), to show that $A = A^{(1)} \cdots A^{(S)}$ is $\mathrm{EPS}_p(\prod_t b_t, \sum_t f_t)$. It then suffices to show that $\mathrm{Tr}(A\sigma)$ can be estimated. Although this may seem like a slightly non-constructive proof, this strategy arose due to object-oriented techniques (C++) used during actual implementation of the algorithm. Unrolling the proof of the product theorem, as well as the proof of the theorem that follows, gives an argument very similar to that presented in section 3.4.

Lemma 3.11. *Let σ be an $N \times M$ matrix that is $\mathrm{EHT}_p(b_\sigma, f_\sigma)$. Let A be an $M \times N$ matrix that is $\mathrm{EPS}_p(b_A, f_A)$. It is possible to estimate $Tr(A\sigma)$ to within additive error $\epsilon > 0$, with probability less than $\delta > 0$ of exceeding the error bound, in average time $\mathrm{O}[\log(\delta^{-1})\epsilon^{-2}b_\sigma^2 b_A^2(f_\sigma + f_A)]$.*

Proof. The proof is in appendix 3.B, and follows along the lines of the techniques developed in section 3.4. □

Proof of theorem 3.8. By iterated application of theorem 3.10(b), the product $A = A^{(1)} \cdots A^{(S)}$ is $\mathrm{EPS}_p(\prod_t b_t, \sum_t f_t)$. By lemma 3.11 the value of $\mathrm{Tr}(A\sigma)$ can be estimated in time $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b^2 f)$ where $b = b_\sigma \prod_t b_t$ and $f = f_\sigma + \sum_t f_t$. □

3.5.3 Query complexity

The simulation algorithm of this paper involves sampling a number of paths via Markov chains, each path evaluation in turn requiring certain operations to be performed. Definitions 3.6 and 3.7 each consist of two pairs of conditions, (a) and (b) relating to the number of paths that need to be evaluated (quantified by b), and (c) and (d) concerning tasks that need to be performed for each path (quantified by f). In appendix 3.A we show (theorem 3.22) that there are always α_{mnk}, $P(n, k|m)$, and $Q(m, k|n)$ satisfying conditions (a) and (b) with $b = \|\bar{A}\|_q$ (and in fact smaller b is not possible). However, these probability distributions may not satisfy (c) and (d), which require that the distributions can be sampled from efficiently. It is difficult to make any general statement regarding satisfaction of (c) and (d), since time complexity of computation is in general a difficult problem; satisfaction of these two conditions needs to be considered on a case-by-case basis. However, when considering query complexity rather than time complexity, (c) and (d) can for the most part be ignored as we shall now explain. Note that communication complexity (discussed in section 3.7.2) offers another context in which (c) and (d) can be ignored, since there too computation time is free.

Consider the situation where an algorithm is required to answer some question about an oracle, which is to be thought of as a black box provided to the algorithm (Grover's algorithm is a prominent example). For a classical (i.e. non-quantum) algorithm the oracle can be any function between two finite sets, say $g : X \to Y$. It will be convenient to consider sets of integers, $X = \{0, 1, \ldots, |X| - 1\}$ and $Y = \{0, 1, \ldots, |Y| - 1\}$. The algorithm can query the oracle by providing it a value $x \in X$, and the oracle responds with $g(x)$. This is the only allowed way to gain information about g. The query complexity of the algorithm is defined to be the number of times it queries the oracle. In particular, the query complexity is not affected by the amount of time spent performing computations between queries; computation, even lengthy computation, is not charged for.

Quantum circuits are provided access to an oracle in the form of a unitary operator[10]

$$\mathcal{O}_g = \sum_{x \in X, y \in Y} |x\rangle \langle x| \otimes |y + g(x)\rangle \langle y| \tag{3.62}$$

where $|x\rangle \otimes |y\rangle \in \mathbb{C}^{|X|} \otimes \mathbb{C}^{|Y|}$ are computational basis vectors and where the addition $y + g(x)$ is modulo $|Y|$. The query complexity of a quantum circuit is defined to be the number of times $\mathcal{O}_g$ appears in the circuit. For example, Grover's algorithm has query complexity $\mathrm{O}(\sqrt{N})$.

Computational complexity classes can be analyzed by comparing how two classes perform when given access to equivalent oracles. For example, oracles have been constructed relative to which quantum computers perform exponentially more efficiently than classical computers (e.g. Simon's problem [Sim94]), whereas proving that quantum computers are faster than classical computers in the absence of an oracle is an extremely difficult open problem.

Considering query complexity rather than time complexity simplifies the analysis of the present paper. Suppose we wish to simulate a quantum circuit containing at least one instance of an oracle

[10] Sometimes an alternate definition $\mathcal{O}'_g = \sum_{x \in X, y \in Y} e^{2\pi i g(x) y/|Y|} |x\rangle \langle x| \otimes |y\rangle \langle y|$ is used. All claims apply to this definition as well, requiring only a modification of (3.63)-(3.66).

$\mathcal{O}_g$ (e.g. Grover's algorithm) on a classical computer that also has oracle access to g. Simulation of the quantum circuit on the classical computer will require making queries to g and we can ask how many queries are needed, ignoring the amount of computational time used. We do this by modifying conditions (c) and (d) of definitions 3.6 and 3.7 to require that the sampling and computation tasks be completed using $\mathrm{O}(f)$ queries to g, rather than requiring $\mathrm{O}(f)$ time (time now being a resource that is not charged for). We will refer to such modified definitions by invoking the phrase "in terms of query complexity".

We will now show that in terms of query complexity, $\mathcal{O}_g$ is $\mathrm{EPS}_p(1,1)$. Since this unitary operates on two subsystems, $\mathbb{C}^{|X|} \otimes \mathbb{C}^{|Y|}$, the indices m and n in definition 3.6 are tuple valued. We write $m = (x, y) \in X \times Y$ and $n = (x', y') \in X \times Y$. Take K to be the singleton set $\{0\}$ and define

$$\begin{aligned} \alpha_{(x,y)(x',y')k} &:= P((x', y'), k|(x, y)) && (3.63)\\ &:= Q((x, y), k|(x', y')) && (3.64)\\ &:= \langle xy|\mathcal{O}_g|x'y'\rangle && (3.65)\\ &= \delta(x, x')\delta(y + g(x), y') && (3.66) \end{aligned}$$

where δ is the Kronecker delta. It is easy to see that these satisfy conditions (a) and (b) of definition 3.6 with $b = 1$. Sampling from these probability distributions and computing the values of any of these quantities can be done with a single query of g (note that the conditional probability distributions are deterministic), therefore conditions (c) and (d) are satisfied with $f = 1$.

On the other hand, for matrices that are not defined in terms of the oracle g, such as the $I - 2|+\rangle\langle+|$ reflection operators in Grover's algorithm, the operations required by conditions (c) and (d) can be carried out using zero queries. Therefore conditions (c) and (d) can be completely ignored, and we can take $f = 0$. We are then free to focus on determining the probability distributions giving the smallest possible value of b in conditions (a) and (b) without regard to whether these can be efficiently sampled from (since we are charging for queries only and time is free). It is desirable to make b as small as possible, since this determines the number of paths that need to be sampled. The number of paths sampled matters, because each will require evaluating the entire Markov chain, which involves every operator. At least one of these operators involves the oracle, so at least one query needs to be made for each path that is sampled. The total number of oracle queries will be the number of paths sampled times the number of queries per path. In appendix 3.A we show (theorem 3.22) the existence of probability distributions which satisfy conditions (a) and (b) with $b = \|\bar{A}\|_q$. So in terms of query complexity, any matrix A not defined in terms of an oracle is $\mathrm{EPS}_p(\|\bar{A}\|_q, 0)$. In the case $p = q = 2$ of relevance to quantum circuits, we have $\|\bar{A}\|_2 = \mathcal{I}_{\max}(A)$, the interference producing capacity of A. Theorem 3.22 also shows that any σ not defined in terms of an oracle is $\mathrm{EHT}_2(\|\sigma\|_{\mathrm{Tr}}, 0)$ where $\|\cdot\|_{\mathrm{Tr}}$ is the trace norm (a generalization is provided for $p \neq 2$).

3.5.4 Sufficient conditions for EPS/EHT

We now present theorems that can be used to show that specific operators are EPS or EHT. As stated above, if one is only interested in query complexity then any matrix A not depending on an oracle is guaranteed to be $\mathrm{EPS}_p(\|\bar{A}\|_q, 0)$. However, in terms of time complexity it is possible that the probability distributions that achieve $b = \|\bar{A}\|_q$ cannot be sampled from efficiently (giving large f). For this reason it is worthwhile to introduce probability distributions that are more likely to be efficiently sampled, and which in some cases still achieve a small b. In the theorem below each row and column of A is treated as a probability distribution, correcting for phases and normalization. This works well when the absolute row and column sums of A are small.

Theorem 3.12. *Let $1 \leq p \leq \infty$ and $1/p + 1/q = 1$. Let A be an $M \times N$ matrix. Define the probability distributions*

$$P(n|m) = \frac{|A_{mn}|}{\sum_{n'} |A_{mn'}|}, \quad Q(m|n) = \frac{|A_{mn}|}{\sum_{m'} |A_{m'n}|}. \tag{3.67}$$

Suppose that it is possible in average time $\mathrm{O}(f)$ *on a classical computer to perform the following operations.*

(a) Given m*, sample* n *from the probability distribution* $P(n|m)$*.*

(b) Given n*, sample* m *from the probability distribution* $Q(m|n)$*.*

(c) Given m, n*, compute* A_{mn}*,* $\sum_{n'} |A_{mn'}|$*, and* $\sum_{m'} |A_{m'n}|$*.*

Then A *is* $\mathrm{EPS}_p(b, f)$ *with* $b = \|A\|_\infty^{1/p} \|A\|_1^{1/q}$*. Note that* b *is the weighted geometric mean of the maximum row and column sums of* A*.*

Proof. This follows directly from plugging the probability distributions (3.67) into definition 3.6, with $K = \{0\}$ (i.e. not making use of the index k). Note that $\|A\|_\infty$ is the maximum absolute row sum and $\|A\|_1$ is the maximum absolute column sum of A. □

Finally, we present theorems that cover the two most important examples of EHT operators: dyads and density operators.

Theorem 3.13 (Dyads are EHT)**.** *Let* $|\phi\rangle$ *and* $\langle\psi|$ *be vectors such that the probability distributions* $P(n) = |\psi_n|^p / \|\psi\|_p^p$ *and* $Q(m) = |\phi_m|^q / \|\phi\|_q^q$ *can be sampled from, and the corresponding* ψ_n *and* ϕ_m *can be computed, in average time* $\mathrm{O}(f)$*. Then the dyad* $|\phi\rangle\langle\psi|$ *is* $\mathrm{EHT}_p(\|\psi\|_p \|\phi\|_q, f)$*.*

Proof. This can be seen immediately by plugging the given probability distributions into definition 3.7, with $K = \{0\}$ (i.e. without making use of index k). This is the best possible value of b, which can be seen by applying theorem 3.22(a) and using $\|(|\phi\rangle\langle\psi|)\|_q = \|\psi\|_p \|\phi\|_q$. □

Corollary 3.14 (Estimate matrix entries)**.** *Let* A *be* $\mathrm{EPS}_p(b, f)$*. Then, given any indices* i, j*, the value of the matrix entry* A_{ij} *can be estimated to within additive error* $\epsilon > 0$*, with probability less than* $\delta > 0$ *of exceeding the error bound, in average time* $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b^2 f)$*.*

Proof. By theorem 3.13 the dyad of computational basis vectors $|j\rangle\langle i|$ is $\mathrm{EHT}_p(1, \log(N))$. Note: $f \geq \log(N)$ in all cases (unless one is dealing with query complexity) since it takes $\mathrm{O}(\log(N))$ time to even write down the indices i and j, which are $\log(N)$ bits long. By lemma 3.11, $A_{ij} = \mathrm{Tr}(A|j\rangle\langle i|)$ can be estimated in time $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b^2[f + \log(N)]) = \mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b^2 f)$. □

Theorem 3.15 (Density operators are EHT)**.** *Let* σ *be a density operator. Suppose that it is possible to sample from the probability distribution* $P(n) = \sigma_{nn}$ *in average time* $\mathrm{O}(f)$ *and, given* i, j*, to compute* σ_{ij} *in average time* $\mathrm{O}(f)$*. Then* σ *is* $\mathrm{EHT}_2(1, f)$*.*

Proof. This follows from plugging the probability distributions $P(n) = \sigma_{nn}$ and $Q(m) = \sigma_{mm}$ into definition 3.7 and using the inequality $|\sigma_{mn}| \leq \sqrt{\sigma_{mm}\sigma_{nn}}$, which is satisfied by positive semidefinite matrices. □

3.6 Simulation of quantum circuits

3.6.1 Efficiently simulated states and operators

In this section we take up the case $p = q = 2$, which is relevant to quantum circuits, and list several examples of $\mathrm{EHT}_2(b, f)$ states and $\mathrm{EPS}_2(b, f)$ operators where b is small and $f \leq \mathrm{polylog}(N)$ where N is the dimension of the system (i.e. $N = 2^n$ where n is the number of qubits). By theorem 3.8, circuits made of such states and operators can be efficiently simulated. For example, the circuit depicted in fig. 3.1 can be simulated in $\mathrm{polylog}(N)$ time. After providing several examples of such states and operators, we discuss a few circuits that cannot be efficiently simulated using our technique.

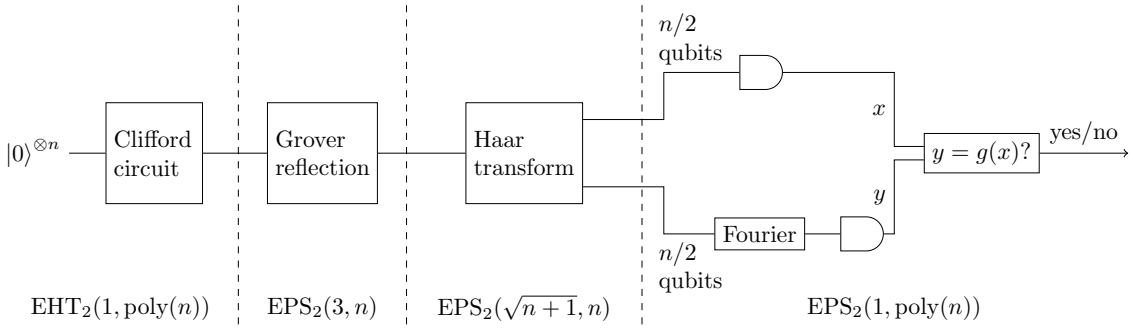


Figure 3.1: An example of the type of circuit that can be simulated in $\mathrm{poly}(n)$ time using the techniques of this paper. The circuit is divided into four sections: the first section is considered to be the initial state, the middle two sections are unitary matrices, and the last section is a projector. The block labeled $y = g(x)$ represents a classical computation step that outputs "yes" if the first and second measurement operations result in values that are related by an arbitrary (but $\mathrm{poly}(n)$ time computable) function g.

The initial states we are able to efficiently simulate include the *computationally tractable* (CT) states of [VdN11]. We reproduce the definition here.[11]

Definition 3.16. *A normalized state $|\psi\rangle$ of dimension N is called computationally tractable (CT) if the following conditions hold:*

(a) It is possible to sample in $\mathrm{polylog}(N)$ *time with classical means from the probability distribution* $P(i) = |\psi_i|^2$.

(b) Upon input of any $i \in \{0, \ldots, N-1\}$, the coefficient ψ_i can be computed in $\mathrm{polylog}(N)$ *time on a classical computer.*

It follows immediately from theorem 3.13 that if $|\psi\rangle$ is a CT state then $\rho = |\psi\rangle\langle\psi|$ is $\mathrm{EHT}_2(1, \mathrm{polylog}(N))$. For convenience we present here a brief list of examples of such states from [VdN11] and refer the reader to their paper for details:

- Product states of qubits (we allow also qudits).
- Stabilizer states.
- States of the form $|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} e^{i\theta(x)} |x\rangle$ where $e^{i\theta(x)}$ for a given x can be computed in $\mathrm{polylog}(N)$ time.
- Matrix product states of polynomial bond dimension.
- States obtained by applying a polynomial sized nearest-neighbor matchgate circuit to a computational basis state.
- States obtained by applying the quantum Fourier transform to a product state.
- The output of quantum circuits with logarithmically scaling tree-width acting on product input states.

[11] Their definition referred to qubits. We generalize slightly to the abstract case where the decomposition into subsystems is not defined, only the total dimension of the space matters.

We present a list of examples of $\mathrm{EPS}_2(b, f)$ operators with b small and $f \leq \mathrm{polylog}(N)$. All proofs are in appendix 3.C.

- If A is $\mathrm{EPS}_p(b, f)$ then $I \otimes \cdots \otimes I \otimes A \otimes I \otimes \cdots \otimes I$ is $\mathrm{EPS}_p(b, \max\{f, \log^2(N)\})$ (corollary 3.37). In other words, EPS operations on subsystems are EPS. The $\log^2(N)$ is due to the amount of time needed to convert indices of $I \otimes \cdots \otimes I \otimes A \otimes I \otimes \cdots \otimes I$ to indices of A.

- Any operator A on a constant number of qubits or qudits is $\mathrm{EPS}_2(\mathcal{I}_{\max}(A), 1)$ where $\mathcal{I}_{\max}(A) = \|\bar{A}\|_2$ is the interference producing capacity of A. In other words, the simulation cost due to such an operator is equal to the fourth power of its interference producing capacity (because of the b_t^4 term in (3.69)).

- If A is an $M \times M$ matrix with maximum singular value bounded by 1 (e.g. a unitary, projector, or POVM element) then $\mathcal{I}_{\max}(A) \leq \sqrt{M}$. This inequality is saturated when A is a unitary with rows forming a basis mutually unbiased to the computational basis (e.g. a Hadamard or Fourier transform).

- In terms of query complexity rather than time complexity, any operator A not depending on an oracle is $\mathrm{EPS}_2(\mathcal{I}_{\max}(A), 0)$ by theorem 3.22. Oracles themselves are $\mathrm{EPS}_2(1, 1)$.

- Efficiently computable sparse matrices as defined in [VdN11] are $\mathrm{EPS}_p(\mathrm{polylog}(N), \mathrm{polylog}(N))$ (theorem 3.33). These include:

  - Permutation matrices are $\mathrm{EPS}_p(1, f)$ as long as the permutation and its inverse can be computed in time $\mathrm{O}(f)$.
  - Diagonal unitary matrices are $\mathrm{EPS}_p(1, f)$ as long as the phases can be computed in time $\mathrm{O}(f)$.
  - Pauli matrices are $\mathrm{EPS}_p(1, 1)$.

- Grover reflections $I - 2(|+\rangle \langle +|)^{\otimes n}$ are $\mathrm{EPS}_2(3, n)$ (theorem 3.38).

- The Haar wavelet transform on n qubits (definition 3.39) is $\mathrm{EPS}_2(\sqrt{n+1}, n)$ (theorem 3.40).

- One dimensional projectors onto CT states are $\mathrm{EPS}_2(1, \mathrm{polylog}(N))$ since CT dyads are $\mathrm{EHT}_2(1, \mathrm{polylog}(N))$ and EHT operators are EPS (theorem 3.9).

- Rank r projectors onto spaces defined by CT states are $\mathrm{EPS}_2(r, \mathrm{polylog}(N))$ (by applying the sum rule theorem 3.10(a) to the previous item).

- Block diagonal matrices where each block is $\mathrm{EPS}_p(b, f)$ are $\mathrm{EPS}_p(b, f)$, as long as matrix indices can be converted to/from block indices in time $\mathrm{O}(f)$ (theorem 3.34).

- As a special case of block diagonal matrices, projectors of the form $\sum_x |x\rangle \langle x| \otimes |\phi_x\rangle \langle \phi_x|$, where the $|x\rangle$ are computational basis states and each $|\phi_x\rangle$ is a CT state, are $\mathrm{EPS}_2(1, \mathrm{polylog}(N))$. Example: given an even number of qubits, measure half of the qubits in the computational basis to get x, measure the other half in the Fourier basis to get y, return true if $y = g(x)$ for some function g computable in $\mathrm{polylog}(N)$ time (corollary 3.36). In this example, $|\phi_x\rangle = F\,|g(x)\rangle$. The measurement depicted in fig. 3.1 is of this form.

3.6.2 Simulation techniques

As a matter of convenience, we present a theorem that is essentially a direct corollary of theorem 3.8, but written in the language of quantum circuits.

Theorem 3.17. *Consider a quantum circuit using states of dimension N (i.e. $\log_2(N)$ qubits or $\log_d(N)$ qudits). Let $|\psi\rangle$ be a computationally tractable (CT) state. For $t \in \{1, \ldots, T\}$ let $U^{(t)}$ be an $\mathrm{EPS}_2(b_t, \mathrm{polylog}(N))$ unitary and let M be an $\mathrm{EPS}_2(b_M, \mathrm{polylog}(N))$ Hermitian observable. It is possible, with probability less than $\delta > 0$ of exceeding the error bound, to estimate*

$$\left\langle \psi \middle| U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)} \middle| \psi \right\rangle \tag{3.68}$$

to within additive error $\epsilon > 0$ in average time

$$\mathrm{O}\left(T \log(\delta^{-1}) \epsilon^{-2} \mathrm{polylog}(N) b_M^2 \prod_{t=1}^{T} b_t^4 \right). \tag{3.69}$$

In particular, if b_M, $\prod_t b_t$, and T are $\mathrm{polylog}(N)$, and if δ and ϵ are constant, then the simulation time is $\mathrm{polylog}(N)$ on average.

Note that in (3.69) each unitary $U^{(t)}$ incurs a cost of b_t^4 rather than b_t^2 since it appears twice in (3.68). If M is a rank one projector onto a CT state, $M = |\phi\rangle\langle\phi|$, then it is much more efficient to compute (3.68) as the absolute square of

$$\mathrm{Tr}\{|\psi\rangle\langle\phi| U^{(T)} \cdots U^{(1)}\}. \tag{3.70}$$

Since $|\psi\rangle\langle\phi|$ is $\mathrm{EHT}_2(1, \mathrm{polylog}(N))$, and since each unitary only occurs once, theorem 3.8 gives that this expression can be estimated in average time

$$\mathrm{O}\left(T \log(\delta^{-1}) \epsilon^{-2} \mathrm{polylog}(N) \prod_{t=1}^{T} b_t^2 \right), \tag{3.71}$$

which is much better than (3.69). If M is a low rank projector, the same trick can be used by decomposing M as the sum of rank one projectors and computing each resulting term individually. The complexity of such a technique will scale proportional to the rank of M.

Theorem 3.17 is just an application of theorem 3.8 with $p = q = 2$. One may wonder whether other values of p, q would lead to a lower simulation cost. Ignore for the moment the efficient sampling conditions (c) and (d) of definition 3.6 and definition 3.7. When estimating (3.68), the optimal probability distributions give (by theorem 3.22)

$$\begin{aligned} b &:= b_\psi b_{U^{(1)}} \cdots b_{U^{(T)}} b_M b_{U^{(T)}} \cdots b_{U^{(1)}} b_\psi && (3.72)\\ &= \|\psi\|_p \|\bar{U}^{(1)\dagger}\|_q \cdots \|\bar{U}^{(T)\dagger}\|_q \|\bar{M}\|_q \|\bar{U}^{(T)}\|_q \cdots \|\bar{U}^{(1)}\|_q \|\psi\|_q. && (3.73) \end{aligned}$$

This achieves its minimum value at $p = q = 2$, since

$$\begin{aligned} b &= \|\psi\|_p \|\psi\|_q \|\bar{U}^{(1)}\|_p \|\bar{U}^{(1)}\|_q \cdots \|\bar{U}^{(T)}\|_p \|\bar{U}^{(T)}\|_q \left(\|\bar{M}\|_p \|\bar{M}\|_q\right)^{1/2} && \text{(using } \|A^\dagger\|_q = \|A\|_p) && (3.74)\\ &\geq \langle\psi|\psi\rangle \|\bar{U}^{(1)}\|_p \|\bar{U}^{(1)}\|_q \cdots \|\bar{U}^{(T)}\|_p \|\bar{U}^{(T)}\|_q \left(\|\bar{M}\|_p \|\bar{M}\|_q\right)^{1/2} && \text{(Hölder's inequality)} && (3.75)\\ &\geq \langle\psi|\psi\rangle \|\bar{U}^{(1)}\|_2^2 \cdots \|\bar{U}^{(T)}\|_2^2 \|\bar{M}\|_2 && \text{(Riesz–Thorin theorem)} && (3.76)\\ &= \|\psi\|_2 \|\bar{U}^{(1)\dagger}\|_2 \cdots \|\bar{U}^{(T)\dagger}\|_2 \|\bar{M}\|_2 \|\bar{U}^{(T)}\|_2 \cdots \|\bar{U}^{(1)}\|_2 \|\psi\|_2. && (3.77) \end{aligned}$$

On the other hand, when estimating an expression of the form (3.70), each unitary is no longer repeated twice and Riesz–Thorin cannot be applied. In this case the minimum value of b does not necessarily occur at $p = 2$.

Certain algorithms, such as Shor's algorithm, consist of a quantum circuit terminating in a many-outcome measurement (e.g. measurement in the computational basis of several different qubits) which

is then post-processed by a classical computer to produce a final result. This does not immediately fit into our scheme of estimating expectation values. However, in the case where the final result is a two-outcome yes/no answer (e.g. "does N have a prime factor in the range $[a, b]$"), the final measurement and classical post-processing can be combined into a single collective projector or POVM element as follows. Suppose the final state is measured using a POVM $\{F_i\}$. A classical post-processing step then inspects the measurement outcome i and returns "yes" or "no". Denote by R the set of measurement outcomes that will result in "yes". The classical post-processing can be absorbed into the measurement, resulting in the POVM element $F' = \sum_{i \in R} F_i$. The expectation value of F' gives the probability that a measurement of $\{F_i\}$ would yield "yes" after post-processing.

In some cases F' may be efficiently simulated, a (somewhat contrived) example being the final stage of the circuit of fig. 3.1. Note that this example involves a Fourier transform, which by itself cannot be efficiently simulated by our technique since it has large interference producing capacity. However, when the Fourier transform is followed by the particular classical post-processing depicted in fig. 3.1, the resulting composite operator *can* be efficiently simulated (corollary 3.36). Shor's algorithm also has a Fourier transform followed by classical post-processing, however in that case the composite operator (Fourier transform followed by post-processing) has large interference producing capacity and so *cannot* be efficiently simulated (by our algorithm).

3.6.3 Circuits that our technique can't efficiently simulate

Many examples of efficiently simulatable circuits can be constructed, but it is probably more enlightening to instead discuss examples of circuits that cannot be efficiently simulated using our technique. Since the efficiency of our technique depends upon choice of basis and on choice of representation (see section 3.7.1), a circuit which our technique cannot simulate efficiently in one basis may be efficiently simulatable in another basis. In this section we choose to focus only on the computational basis. That being said, most of the examples in this section have been proved (relative to an oracle) to have no efficient classical solution.

We cannot efficiently simulate Shor's algorithm. The reason for this is that the Fourier transform has high interference producing capacity: the Fourier transform F on n qubits has $\mathcal{I}_{\max}(F) = 2^{n/2}$. Replacing the Fourier transform by the Haar wavelet transform (fig. 3.2) yields a circuit that can be efficiently simulated, since the Haar transform has low interference producing capacity, $\mathcal{I}_{\max}(G_n) = \sqrt{n+1}$. Note that this circuit no longer factors numbers (and probably does nothing at all useful). The Fourier and Haar transforms play similar roles in classical signal processing, with the latter providing spatially localized rather than global information for the high frequency components. The fact that replacing the Fourier transform enables efficient classical simulation points to the Fourier transform as being the source of the quantum speedup in Shor's algorithm (for a contrasting point of view, see [ALM07, YS07]).

Deutsch–Jozsa provides an oracle relative to which deterministic quantum computation is more powerful than deterministic classical computation. Our algorithm can efficiently simulate the Deutsch–Jozsa algorithm, but not deterministically.[12] The Deutsch–Jozsa algorithm consists of an initial CT state $|+\rangle^{\otimes n} \otimes |-\rangle$, acted upon by an oracle $\sum_{xy} |x\rangle\langle x| \otimes |y + g(x)\rangle\langle y|$, followed by a rank-one projective measurement onto the state $|+\rangle^{\otimes n} \otimes |-\rangle$. The initial state is $\mathrm{EHT}_2(1, n)$ and the operators are $\mathrm{EPS}_2(1, n)$, so we can efficiently simulate this algorithm. However, the simulation will always have a small chance of error due to the δ in theorem 3.17.

Our simulation algorithm performs very poorly when applied to Grover's algorithm. Each iteration of Grover's algorithm consists of an oracle query followed by a Grover reflection. These operations have low interference producing capacity: 1 for the oracle and just under 3 for the Grover reflection. However, our algorithm is exponentially slow in the circuit length, due to the $\prod_t b_t^4$ factor in (3.69). Since the Grover reflection is used $\Theta(\sqrt{N})$ times, the simulation would run in time $\exp(\Theta(\sqrt{N}))$.

[12] This was discussed in [VdN11], which our paper extends. However, we mention it here for completeness.

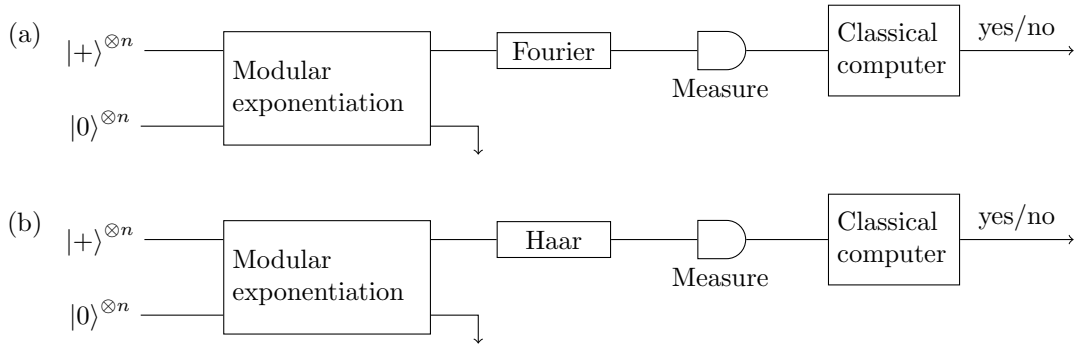


Figure 3.2: (a) A depiction of the decisional version of Shor's algorithm, which outputs "yes" if there is a prime factor within some given range. (b) The Haar wavelet transform (definition 3.39) plays a similar role as the Fourier transform in classical signal processing. However, substituting the Haar transform for the Fourier transform in Shor's algorithm yields a circuit that can be efficiently simulated on a classical computer. Note that the resulting circuit won't factor numbers, and in fact probably has no practical use.

Even though each iteration of Grover's algorithm produces small interference, the total interference of the whole circuit, by definition 3.3, is $\exp(\Theta(\sqrt{N}))$.

In [CCD+03] a quantum random walk is presented that provides an exponential speedup over any possible classical algorithm for the graph traversal problem. The walk is carried out by evolving the initial state with a Hamiltonian that is defined in terms of an oracle. We cannot efficiently simulate this algorithm for the same reason that we cannot efficiently simulate Grover: the runtime of the quantum algorithm increases with the problem size, and our simulation must pay an exponentially large penalty for this due to the $\prod_t b_t^4$ factor in (3.69). On the other hand, short time/low energy Hamiltonian evolutions can be efficiently simulated by our technique. In particular, theorem 3.10(c) gives that if H is $\mathrm{EPS}_p(b, f)$ then e^{iHt} is $\mathrm{EPS}_p(e^{bt}, btf)$. In terms of query complexity the Hamiltonian in the algorithm of [CCD+03] is $\mathrm{EPS}_2(\mathrm{O}(1), 1)$, so we could feasibly simulate e^{iHt} for small t. However, their algorithm has $t = \Theta(n^4)$, so our simulation would have query complexity $e^{\Theta(n^4)}$, making it unfeasibly slow.

3.7 Applications and discussion

3.7.1 Wigner representation

An $N \times N$ matrix can also be viewed as an N^2 dimensional vector, so we can write for instance $\langle \boldsymbol{M}|\boldsymbol{\rho}\rangle$ in place of $\mathrm{Tr}\{M\rho\}$. Superoperators become $N^2 \times N^2$ matrices in this representation, and we can write $\langle \boldsymbol{M}|\boldsymbol{VU}|\boldsymbol{\rho}\rangle = \mathrm{Tr}\{MVU\rho U^\dagger V^\dagger\}$. Simulating a quantum circuit using this representation offers an alternative to the customary representation that was the focus of section 3.6.

Any basis can be used (even ones that are not orthonormal), although some choices of basis may yield more efficient simulation. One notable choice is given by the discrete Wigner representation, which is only defined for qudits of odd dimension. We will not describe the details here but refer the reader to [VFGE12, ME12] in which it is shown that in the discrete Wigner representation stabilizer states become probability distributions and Clifford operations become permutation matrices.

It was shown independently in [VWFE13, ME12] that when operations in the Wigner representation are given by nonnegative matrices, such matrices are stochastic and therefore can be efficiently simulated. Our algorithm, taking $p = \infty$ and $q = 1$, extends this result by also allowing states

and operations in which the Wigner representation contains a small quantity of negative values, although ours is weaker in that it only computes expectation values rather than allowing sampling of a many-outcome measurement. With $q = 1$ rather than $q = 2$, the difficulty of simulating an operation is given not by $\mathcal{I}_{\max}(A) = \|\bar{A}\|_2$ but rather by $\|\bar{A}\|_1 = \|A\|_1$, the maximum absolute column sum. In cases where the matrix in the Wigner representation is nonnegative, the matrix will be left-stochastic and $\|A\|_1 = 1$, such matrices will not increase the number of samples needed. If there are some negative values then $\|A\|_1$ will be larger.

After the present work was completed, the quantity $\log\|\boldsymbol{\rho}\|_1$ was investigated in [VMGE14]. This quantity was termed "mana" and was shown to be monotone under Clifford operations, and to be monotone on average under stabilizer measurements, thus providing bounds on magic state distillation by Clifford circuits. Given the results of the present paper, it should perhaps make sense to extend the concept of mana also to quantum operations, defining their mana to be $\log\|A\|_1$. Then Clifford operations have zero mana and in general the following monotonicity relation is satisfied:

$$\log\|A\boldsymbol{\rho}\|_1 \leq \log(\|A\|_1\|\boldsymbol{\rho}\|_1) = \log\|A\|_1 + \log\|\boldsymbol{\rho}\|_1. \tag{3.78}$$

So $\log\|A\|_1$, which is the Wigner representation analogue of the log of interference producing capacity, bounds the amount by which the operator A may increase the mana of a state. For each A there will be some $\boldsymbol{\rho}$ that saturates this inequality (by the definition of operator norm), but it is not clear whether this would correspond to a physical state.

Stated in this language, theorem 3.8, applied in the Wigner representation, gives that quantum circuits may be efficiently simulated classically in time polynomial in $\|\boldsymbol{M}\|_\infty$ (where $\boldsymbol{M}$ is the final measurement) and exponential in the sum of the mana of the initial state and the mana of each operation. Specifically, write $\langle \boldsymbol{M}|\boldsymbol{V}\boldsymbol{U}|\boldsymbol{\rho}\rangle = \mathrm{Tr}\{|\boldsymbol{\rho}\rangle\langle \boldsymbol{M}|\,\boldsymbol{V}\boldsymbol{U}\}$. Then, ignoring for the moment conditions (c)-(d) of definition 3.6 and (c)-(d) of definition 3.7, we have (by theorem 3.22) that $|\boldsymbol{\rho}\rangle\langle \boldsymbol{M}|$ is $\mathrm{EHT}_\infty(\|\boldsymbol{\rho}\|_1\|\boldsymbol{M}\|_\infty, f)$ and $\boldsymbol{U}$ is $\mathrm{EPS}_\infty(\|\boldsymbol{U}\|_1, f)$ (similarly for $\boldsymbol{V}$). So by theorem 3.8 this can be simulated in time

$$\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}\|\boldsymbol{M}\|_\infty\|\boldsymbol{U}\|_1\|\boldsymbol{V}\|_1\|\boldsymbol{\rho}\|_1 f). \tag{3.79}$$

This complements the result of [VMGE14] which showed mana to be a necessary resource for magic state distillation but did not show that circuits of low total mana have no quantum speedup (although the zero mana case was treated in [VWFE13, ME12]).

3.7.2 Communication complexity

Consider a scenario in which two parties, Alice and Bob, are to cooperatively evaluate a boolean function. Specifically, suppose that Alice receives input x, Bob receives input y, and they are to evaluate $g(x, y)$ where the function $g : X \times Y \to \{0, 1\}$ is known to the two parties ahead of time. They must provide the correct answer with probability at least $2/3$. For non-trivial functions this will require communication, which can be either quantum or classical. The communication complexity of g is the number of bits of communication required by the optimal protocol, with no regard for the amount of time Alice and Bob spend on local computations. For some problems quantum communication is exponentially more efficient than classical communication [RK11].

Consider a quantum communication protocol as depicted by fig. 3.3. The initial state, denoted $|\psi\rangle$, is a pure (but possibly entangled) state on three subsystems $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. Subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$ are owned by Alice and Bob respectively, and subsystem $\mathcal{H}_C$ is passed between Alice and Bob through a noiseless quantum channel for each round of communication. Alice begins by performing a unitary operation $A^{(1,x)}$, which can depend on her input x, on subsystems $\mathcal{H}_A \otimes \mathcal{H}_C$. She then sends the $\mathcal{H}_C$ subsystem to Bob, who performs a unitary operation $B^{(2,y)}$, which can depend on his input y, on subsystems $\mathcal{H}_B \otimes \mathcal{H}_C$. Bob sends $\mathcal{H}_C$ back to Alice who then performs $A^{(3,x)}$ and so on. Finally, the last party (say, Bob) performs a two outcome projective (or POVM) measurement

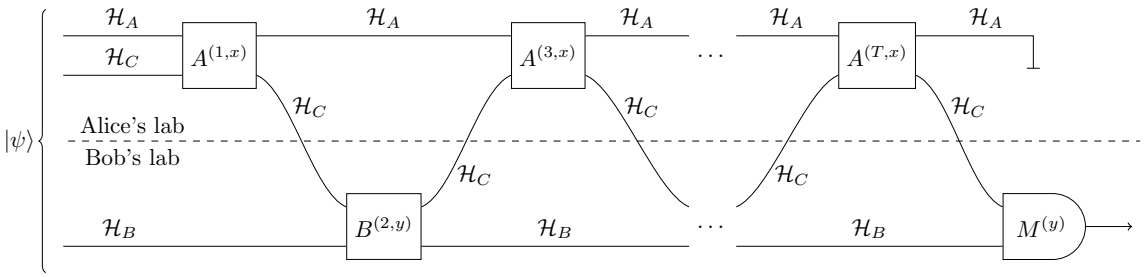


Figure 3.3: A quantum communication protocol. The expectation value of the final measurement is given by (3.80).

$\{M^{(y)}, I - M^{(y)}\}$, which can depend on y, on subsystems $\mathcal{H}_B \otimes \mathcal{H}_C$ and reports the outcome. The expectation value of the final measurement is given by

$$\left\langle \psi \middle| A^{(1,x)\dagger} B^{(2,y)\dagger} A^{(3,x)\dagger} \cdots A^{(T,x)\dagger} M^{(y)} A^{(T,x)} \cdots A^{(3,x)} B^{(2,y)} A^{(1,x)} \middle| \psi \right\rangle \tag{3.80}$$

and must be $\leq 1/3$ if $g(x,y) = 0$ and $\geq 2/3$ if $g(x,y) = 1$. The communication complexity of the protocol is the number of qubits transmitted, $T \log(\dim(\mathcal{H}_C))$ where T is the number of rounds of communication. The dimensionality of the subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$ is not taken into consideration.

The algorithm of this paper can be adapted to provide classical communication simulations of quantum communication protocols, in the case where the quantum protocols are built using operators having low interference producing capacity, and making a certain assumption regarding the initial state $|\psi\rangle$. Since the expectation value of the final measurement in the quantum protocol will be either $\leq 1/3$ or $\geq 2/3$, a classical simulation of the quantum protocol can with probability $\geq 2/3$ determine $g(x,y)$ if it can, with chance of error $\delta \leq 1/3$, estimate the expectation value of the quantum protocol to within additive error $\epsilon < 1/6$. This is exactly the type of estimation provided by the algorithm of this paper, we need only adapt it to the communication scenario.

The algorithm presented in section 3.4.4 involves computing $\mathrm{O}(b_{\max}^2)$ path samples,[13] each of which require evaluation of a left-to-right or a right-to-left Markov chain. Crucially, each transition operator in these chains is defined solely in terms of a single operator of (3.80). Therefore, each transition can be computed by Alice alone (for the $A^{(t,x)}$ operators) or by Bob alone (for the $B^{(t,y)}$ and $M^{(y)}$ operators). The state space of the Markov chains consists of indices corresponding to computational basis states of $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, so the indices can be thought of as triples (i_A, i_B, i_C) of indices over $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_C$. Since Alice's operators $A^{(t,x)}$ act only on subsystems $\mathcal{H}_A \otimes \mathcal{H}_C$, the corresponding transition operators in the Markov chain involve only indices i_A and i_C. Similarly, Bob's transition operators involve only i_B and i_C. Therefore Alice and Bob need to communicate only the index i_C for each transition of the Markov chain.

Also needed is selection of the initial index according to the probability distribution $P(i_A, i_B, i_C) = |\langle i_A, i_B, i_C|\psi\rangle|^2$ (with Alice getting (i_A, i_C) and Bob getting i_B), as well as evaluation of $\langle i_A, i_B, i_C|\psi\rangle$ for a given (i_A, i_B, i_C) triple (where Alice knows (i_A, i_C) and Bob knows i_B). If the initial state is a product state, $|\psi\rangle = |\psi_{AC}\rangle \otimes |\psi_B\rangle$, these tasks are easily accomplished using no communication. In fact, even if $|\psi\rangle$ is entangled between Alice and Bob these two tasks can in some cases be accomplished using only a small amount of communication. Alice and Bob both know $|\psi\rangle$ (since it does not depend on x or y), so they can individually sample from $P(i_A, i_B, i_C)$. If Alice and Bob are granted access to shared randomness (a.k.a. public coins), they can sample from $P(i_A, i_B, i_C)$ in a synchronous way (i.e. they both get the same outcome). Computation of $\langle i_A, i_B, i_C|\psi\rangle$ for a given (i_A, i_B, i_C)

[13] Specifically, $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2} b_{\max}^2)$ samples are needed. However, in order to achieve the goal of guessing $g(x,y)$ with probability $\geq 2/3$ it suffices to set constant $\delta < 1/3$ and $\epsilon < 1/6$.

triple, with (i_A, i_C) known to Alice and i_B known to Bob, is trickier and how much communication is needed depends on $|\psi\rangle$. For example, let $\mathcal{H}_A = \mathcal{H}_{A'} \otimes \mathcal{H}_{A''}$ and $\mathcal{H}_B = \mathcal{H}_{B'} \otimes \mathcal{H}_{B''}$ and consider an initial state of the form

$$|\psi\rangle = |\psi_{A'}\rangle \otimes |\psi_{B'}\rangle \otimes |\psi_C\rangle \otimes \sum_i \alpha_i |i\rangle_{A''} \otimes |i\rangle_{B''} \tag{3.81}$$

with $|i\rangle_{A''}$ and $|i\rangle_{B''}$ denoting computational basis vectors. This is the most common type of initial state for quantum protocols that make use of shared entanglement. Then

$$\langle i_A, i_B, i_C|\psi\rangle = \langle i_{A'}|\psi_{A'}\rangle \langle i_{B'}|\psi_{B'}\rangle \langle i_C|\psi_C\rangle \alpha_{i_{A''}} \delta(i_{A''}, i_{B''}) \tag{3.82}$$

where δ is the Kronecker delta. This can be computed using shared randomness and O(1) communication by making use of a bounded error protocol for testing equality of $i_{A''}$ and $i_{B''}$ (example 3.13 of [KN06]).

Since each unitary appears twice in (3.80), evaluation of the entire Markov chain is accomplished with twice as much communication as the classical protocol, or $2T\log(\dim(\mathcal{H}_C))$ bits. The algorithm also requires computing the amplitude associated with the path, as well as the probability of the path. However, this requires only transmission of $\mathrm{O}(T)$ scalar quantities from Alice to Bob, using $\mathrm{O}(T)$ bits of communication.[14] The total classical communication complexity of this simulation protocol is therefore $\mathrm{O}(b_{\max}^2 T \log[\dim(\mathcal{H}_C)])$, a factor $O(b_{\max}^2)$ greater than that of the quantum protocol. Using the optimal probability distributions defined in appendix 3.A, $b_{\max}$ is upper bounded by the product of the interference producing capacities of the operators in (3.80). The communication complexity of the classical simulation is then

$$\mathrm{O}\left(T \log[\dim(\mathcal{H}_C)] \max_{x,y} \left\{ \|\bar{A}^{(1,x)}\|_2^4 \cdot \|\bar{B}^{(2,y)}\|_2^4 \cdot \|\bar{A}^{(3,x)}\|_2^4 \cdots \|\bar{A}^{(T,x)}\|_2^4 \cdot \|\bar{M}^{(y)}\|_2^2 \right\}\right). \tag{3.83}$$

The consequence of this construction is that any quantum communication protocol exhibiting superpolynomial advantage in communication complexity over any classical protocol must have a superpolynomial value of $b_{\max}$ (i.e. the product of the interference producing capacities of the quantum operators must be high) or must make use of an initial state not of the form (3.81). There is, however, an interesting caveat to this claim. Due to the fact that each unitary, as well as the initial state, appears twice in (3.80), our classical simulation will require twice as many communication rounds as the quantum protocol.[15] Our technique therefore does not apply if one limits the number of rounds. For example, the quantum protocol for the Perm-Invariance problem described in [Mon11] has $b_{\max} = 1$ yet is exponentially more efficient than any one-round classical protocol.

There is a way to avoid the doubling of the number of rounds of communication, but at a price. Consider a one-round quantum protocol in which Alice sends a state $|\psi\rangle$ and Bob measures a projector (or POVM element) M. The expectation value is $\langle\psi|M|\psi\rangle = \mathrm{Tr}\{|\psi\rangle\langle\psi| M\}$. As described in the previous subsection, the state $|\psi\rangle\langle\psi|$ and operator M can be vectorized to give $\langle\boldsymbol{\rho}|\boldsymbol{M}\rangle = \mathrm{Tr}\{|\psi\rangle\langle\psi| M\}$. By taking $p = 1$ and $q = \infty$ instead of $p = q = 2$ our algorithm can estimate $\langle\boldsymbol{\rho}|\boldsymbol{M}\rangle$ using only a left-to-right Markov chain, thus requiring only a single round of communication, from Alice to Bob. However, since $p = 1$ and $q = \infty$, the number of bits communicated is $\mathrm{O}(\|\boldsymbol{\rho}\|_1^2 \|\boldsymbol{M}\|_\infty^2 n)$ with n being the number of qubits in $|\psi\rangle$. The reason we can't efficiently simulate the quantum protocol of [Mon11] using this technique is that $\|\boldsymbol{\rho}\|_1$ is exponentially large. Interestingly, [KNR95] provides a one-round protocol that can estimate $\langle\boldsymbol{\rho}|\boldsymbol{M}\rangle$ using $\mathrm{O}(\|\boldsymbol{\rho}\|_2^2 \|\boldsymbol{M}\|_2^2)$ bits of classical communication. However, this again fails to provide an efficient simulation since $\|\boldsymbol{M}\|_2$ is exponentially large.

[14] Actually a careful look shows that only O(1) communication is needed. Alice can locally multiply her transition probabilities and the amplitudes for her operators for the given path and report these $O(1)$ values to Bob who is then able to complete the computation.

[15] Note that independent evaluations of the Markov chain can be run in parallel, otherwise the number of rounds would scale as $\mathrm{O}(b_{\max}^2)$.

3.7.3 Continuity of $\mathcal{I}$ and $\mathcal{I}_{\max}$

Our measures $\mathcal{I}_{\max}$ of definition 3.5 (which we have related to quantum speedup) and $\mathcal{I}$ of definition 3.3 (which we have conjectured to be related to quantum speedup) are continuous as a function of the states and operators of a circuit. To our knowledge, this is the first continuous quantity that has been identified as being a necessary resource for quantum speedup, other resources such as Schmidt rank [Vid03] or tree width [MS08, Joz06] being discrete valued.

An argument was put forth in [Nes12] as to why most continuous quantities could not be considered as a necessary resource for quantum speedup. Although their argument focuses on functions of the state vector, such as entanglement entropy, rather than of the operators, it is still worthwhile to examine whether it is applicable to the present work. We paraphrase their argument here, modifying it slightly to fit the circuit paradigm that we have been using in this paper. Consider a quantum circuit with initial state $|0\rangle^{\otimes n}$, followed by several unitaries, terminated by a final measurement having expectation value v. Add a control to all of the operators in the circuit: $I \otimes |0\rangle\langle 0| + U \otimes |1\rangle\langle 1|$ in place of U for each unitary and similarly for the final measurement. All operators are controlled by an ancillary qubit initially in the state $\sqrt{1-\epsilon}\,|0\rangle + \sqrt{\epsilon}\,|1\rangle$. By repeating execution of the circuit $\mathrm{O}(\epsilon^{-2})$ times, the value of v can be recovered to high accuracy. However, by setting ϵ to a sufficiently low value, the state at all times during the computation will be arbitrarily close to $|0\rangle^{\otimes n+1}$, and thus will have arbitrarily low entanglement. The most commonly used entanglement measures take values that depend polynomially on ϵ, so entanglement can be made quite low without $\mathrm{O}(\epsilon^{-2})$ growing to an unfeasible magnitude. As a consequence, it is not possible to claim without qualification that entanglement is necessary for quantum speedup.

This construction has no effect on the interference producing capacity of the operators of the circuit since $\mathcal{I}_{\max}(I \otimes |0\rangle\langle 0| + U \otimes |1\rangle\langle 1|) = \mathcal{I}_{\max}(U)$. For this reason, our main result regarding $\mathcal{I}_{\max}$ as a necessary resource for quantum speedup is immune to the above argument. On the other hand, the interference measure $\mathcal{I}$ of definition 3.3, which is the subject of the conjectures of section 3.8, is immune to this argument for a different reason. The value of $\mathcal{I}$ can be exponentially high in the number of qubits or number of unitaries of a circuit. In order to make $\mathcal{I}$ small, ϵ would have to be exponentially small, in turn requiring an exponentially large number of repetitions of the circuit. So the construction of [Nes12] is not able to significantly lower the interference of a circuit without also losing the quantum speedup.

3.7.4 Connection to decoherence functional

There is a close connection between the interference $\mathcal{I}$ of definition 3.3 and the decoherence functional introduced by Gell-Mann and Hartle.[16] The latter represents an extension of the Born rule so as to be able to define probabilities for a sequence of events in a closed quantum system. Consider a *family of histories* corresponding to projection onto the computational basis at each step (i.e. after the initial state and after each unitary) of a quantum circuit $\mathrm{Tr}\{U^{(1)\dagger}\cdots U^{(T)\dagger} M U^{(T)}\cdots U^{(1)}\rho\}$. In this case the *decoherence functional* is defined as

$$\mathcal{D}(\boldsymbol{j};\boldsymbol{k}) = \mathrm{Tr}[MW(\boldsymbol{j})\rho W^{\dagger}(\boldsymbol{k})], \tag{3.84}$$

where ρ is the initial state, M is a projector, and

$$W(\boldsymbol{j}) = |j_T\rangle\langle j_T|\, U^T \cdots |j_2\rangle\langle j_2|\, U^{(2)}\, |j_1\rangle\langle j_1|\, U^{(1)}\, |j_0\rangle\langle j_0|\,. \tag{3.85}$$

It is convenient to think of $\mathcal{D}(\boldsymbol{j};\boldsymbol{k})$ as a matrix with rows labeled by $\boldsymbol{j}$ and columns by $\boldsymbol{k}$, and then it is not difficult to show that

$$\sum_{\boldsymbol{j}}\sum_{\boldsymbol{k}} \mathcal{D}(\boldsymbol{j};\boldsymbol{k}) = \mathrm{Tr}\{U^{(1)\dagger}\cdots U^{(T)\dagger} M U^{(T)}\cdots U^{(1)}\rho\}. \tag{3.86}$$

[16] See [GMH93]. Here we use the notation of Chs. 7, 8 and 10 of [Gri03], which is more convenient for our purposes because it employs the Schrödinger rather than the Heisenberg representation.

If the *consistency condition*

$$\mathcal{D}(\boldsymbol{j};\boldsymbol{k}) = 0 \text{ whenever } \boldsymbol{j} \neq \boldsymbol{k} \tag{3.87}$$

is satisfied, then each diagonal element $\mathcal{D}(\boldsymbol{j};\boldsymbol{j})$ can be interpreted (up to normalization) as the probability of the history corresponding to $\boldsymbol{j}$ occurring. The sum of these diagonal elements is then equal to the expectation value of the final observable, the right side of (3.86), since the off diagonal terms vanish.

It is straightforward to show that $\mathcal{I}$ of definition 3.3 is equal to

$$\mathcal{I}\left(U^{(1)\dagger},\ldots,U^{(T)\dagger},M,U^{(T)},\ldots,U^{(1)},\rho\right) = \sum_{\boldsymbol{j}}\sum_{\boldsymbol{k}}\left|\mathcal{D}(\boldsymbol{j};\boldsymbol{k})\right|. \tag{3.88}$$

When the consistency condition (3.87) is satisfied, this will be equal to $\sum_{\boldsymbol{j}}\mathcal{D}(\boldsymbol{j};\boldsymbol{j})$ (since the diagonal entries are always positive), which in turn is equal to the right hand side of (3.86). In general, (3.88) gives a measure of how badly the consistency condition is violated.

3.8 Conjectures

We have shown that quantum speedup requires circuit elements with a large interference producing capacity. In this section we formally state our conjecture that low interference (rather than low interference producing capacity) is sufficient to ensure efficient simulation of a quantum circuit. In general we are interested in circuits of arbitrary length, but for concreteness consider the task of estimating sums of the form

$$\left\langle\psi\middle|U^\dagger M U\middle|\psi\right\rangle = \sum_{ijkl} V(i,j,k,l), \tag{3.89}$$

$$V(i,j,k,l) = \psi_i^* U_{ij}^\dagger M_{jk} U_{kl}\psi_l. \tag{3.90}$$

As discussed in section 3.3, this sum can be estimated by considering a number of randomly chosen paths $\pi = (i,j,k,l)$. If these paths are chosen according to the optimal probability distribution $R_{\text{opt}}(\pi)$ of (3.10) then the number of samples required to estimate (3.89) to within error ϵ (with probability δ of exceeding this error bound) is $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}\mathcal{I}^2)$ where $\mathcal{I} = \left\langle\bar{\psi}\middle|\bar{U}^\dagger\bar{M}\bar{U}\middle|\bar{\psi}\right\rangle$ is the interference of the circuit as given by definition 3.3. The difficulty with this strategy is that we do not know how to efficiently sample paths according to the distribution $R_{\text{opt}}(\pi)$, or anything sufficiently close to it. In other words, we do not have a strategy for finding the most relevant paths. However, we conjecture that there is a way.

Loosely speaking, we conjecture a quantum circuit can be simulated in time $\text{poly}(\log(\delta^{-1})\epsilon^{-1}\mathcal{I})$ as long as the initial state and operators meet some computational tractability conditions, analogous to conditions (c) and (d) of definitions 3.6 and 3.7. Exactly what tractability conditions should be required is difficult to know ahead of time for the following reason. In sections 3.3 and 3.4 a simulation algorithm was developed, which required certain tasks to be performed involving the initial state and the operators of the circuit being simulated. The need to efficiently perform these tasks led directly to the definition of conditions (c) and (d). Now we conjecture a better algorithm, whose specific structure is not known ahead of time. Not knowing the specifics of this conjectured algorithm, it is not clear what should be required in place of conditions (c) and (d). The intuition is that we assume any necessary task involving any individual operator in the circuit can be efficiently performed, but we make no assumption regarding the interactions between several operators.

This can be made more precise. Section 3.5.3 (on query complexity) and section 3.7.2 (on communication complexity) each provided a framework in which the computational tractability conditions (c) and (d) were not relevant. We could use either of these to form a conjecture that avoids the need to state similar conditions. Of these two, communication complexity is representative

of a certain algorithmic structure. Consider algorithms that involve dealing with the elements of a circuit one at a time. For instance, when estimating (3.89) one could imagine carrying out some calculations involving $|\psi\rangle$, making notes of the result, carrying out further calculations involving U, and so on. The time complexity of such an algorithm is lower bounded by the amount of notes taken and the number of times attention is shifted from one circuit element to another. This can be quantified by imagining that each of $|\psi\rangle$, U, and M are stored in separate rooms, and considering how many notes need to be carried back and forth between the rooms by somebody who seeks to estimate (3.89). Equivalently, stated in terms of communication complexity, imagine that Alice has $|\psi\rangle$, Bob has U, and Charlie has M. How much communication is needed in order to estimate (3.89)? We conjecture that the amount of communication needed is polynomial in the interference of the circuit:

Conjecture 3.18. *Suppose that Alice has a classical description of a vector $|\psi\rangle$ of dimension N, Bob has a description of an $N \times N$ POVM element M, and T other parties have descriptions of $N \times N$ unitary matrices $U^{(1)}, \ldots, U^{(T)}$. Then, with probability less than δ of exceeding the error bound, the value of*

$$\left\langle \psi \middle| U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)} \middle| \psi \right\rangle \tag{3.91}$$

can be estimated to within additive error ϵ using $\mathrm{poly}(\log(\delta^{-1})\epsilon^{-1} \max\{1, \mathcal{I}\} \log(N))$ bits of classical communication where $\mathcal{I}$ is the interference of (3.91) as given by definition 3.3.

The reader may worry that this communication scenario has little bearing on the problem of simulating quantum circuits, however it is expected that any proof in the positive of this conjecture will be adaptable into an algorithm that can be used in the computation context. Indeed, the Markov chain technique of section 3.4 was first developed as as solution to a problem resembling conjecture 3.18.

We have been unable to prove this conjecture even for the simple case where there are no unitary operations and the goal is to estimate the expectation value $\langle\psi|M|\psi\rangle$. We present this simplified case formally, as it deserves some discussion.

Conjecture 3.19. *Conjecture 3.18 holds in the case $T = 0$. In other words, suppose that Alice has a classical description of a vector $|\psi\rangle$ of dimension N and Bob has a classical description of an $N \times N$ POVM element M. Then, with probability less than δ of exceeding the error bound, the value $\langle\psi|M|\psi\rangle$ can be estimated to within additive error ϵ using $\mathrm{poly}\left(\log(\delta^{-1})\epsilon^{-1} \max\{1, \mathcal{I}\} \log(N)\right)$ bits of classical communication where $\mathcal{I} = \left\langle \bar{\psi} \middle| \bar{M} \middle| \bar{\psi} \right\rangle$ is the interference of $\langle\psi|M|\psi\rangle$ as given by definition 3.3.*

Conjecture 3.19, being weaker than conjecture 3.18, should be easier to prove true. However, it would probably be very difficult to prove false since a proof that estimating $\langle\psi|M|\psi\rangle$ requires a large amount of classical communication in the general case (not assuming low interference) remained open for 11 years [RK11].

Conjecture 3.19 would be false if only one round of communication was allowed, from Alice to Bob. In [Mon11] the Perm-Invariance problem was defined and shown to be solved efficiently by a one-round quantum protocol, however no efficient one-round classical protocol exists. The quantum protocol has Bob measuring a POVM element M on a state $|\psi\rangle$ sent by Alice and this protocol is low interference, $\mathcal{I} = \left\langle \bar{\psi} \middle| \bar{M} \middle| \bar{\psi} \right\rangle \leq 1$. However, there can be no efficient one-round classical protocol for estimating $\langle\psi|M|\psi\rangle$, since such a protocol would efficiently solve Perm-Invariance. This does not provide a counterexample to conjecture 3.19 since we allow multiple rounds of communication, and there is indeed an efficient classical two round protocol, which can be constructed using the technique of section 3.7.2.

A potential problem with conjecture 3.18 is that the unitary portion of the circuit could create very large interference which could be masked by the final measurement. For example, consider the initial state $|\psi\rangle = |0\rangle^{\otimes n}$, acted upon by an arbitrary circuit involving all but the first qubit, followed by measurement of the observable $M = |1\rangle\langle 1| \otimes I^{\otimes n-1}$. For this circuit $\mathcal{I} = 0$ so conjecture 3.18

says the expectation value can be computed in $\text{poly}(\log(\delta^{-1})\epsilon^{-1}n)$ time, as indeed it can in this case. However, it seems there may be similar situations in which $\mathcal{I}$ is small because of the final measurement, but the circuit is nevertheless difficult to simulate. For this reason we provide an alternate definition that quantifies the interference just before the final measurement, computed by substituting $M = I$ in definition 3.3. This will be used to form a weaker conjecture.

Definition 3.20. *The interference of a quantum circuit without a measurement is*

$$\mathcal{J}(U^{(T)}, \ldots, U^{(1)}, \rho) = Tr\left\{\bar{U}^{(T)} \cdots \bar{U}^{(1)} \bar{\rho} \bar{U}^{(1)\dagger} \cdots \bar{U}^{(T)\dagger}\right\}. \tag{3.92}$$

In other words, $\mathcal{J}$ is the amount by which normalization is spoiled when destructive interference is turned into constructive interference by means of the absolute value applied to each path. This is nondecreasing in time,

$$\mathcal{J}(U^{(T)}, \ldots, U^{(1)}, \rho) \geq \mathcal{J}(U^{(T-1)}, \ldots, U^{(1)}, \rho) \tag{3.93}$$

and $\mathcal{J} = 1$ if all of the unitaries are permutation matrices as in a classical computation. We conjecture that a circuit can be efficiently simulated when $\mathcal{J}$ is small. Since $\mathcal{J}$ doesn't see the final measurement M, we need an extra constraint. We require M to be a projector diagonal in the computational basis.

Conjecture 3.21. *Suppose that Alice has a classical description of a vector $|\psi\rangle$ of dimension N, Bob has a description of an $N \times N$ projector M that is diagonal in the computational basis, and T other parties have descriptions of $N \times N$ unitary matrices $U^{(1)}, \ldots, U^{(T)}$. Then, with probability less than δ of exceeding the error bound, the value of*

$$\left\langle \psi \middle| U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)} \middle| \psi \right\rangle \tag{3.94}$$

can be estimated to within additive error ϵ using $\text{poly}(\log(\delta^{-1})\epsilon^{-1}\mathcal{J}\log(N))$ bits of classical communication where $\mathcal{J} = \mathcal{J}(U^{(T)}, \ldots, U^{(1)}, |\psi\rangle\langle\psi|)$ is the interference of (3.94) just before the final measurement, as given by definition 3.20.

3.9 Summary and open problems

We have provided an algorithm for efficiently simulating quantum circuits in which each operator has low interference producing capacity. Therefore, interference producing capacity is identified as a resource necessary for quantum speedup. The runtime of the simulation is quadratic in the interference producing capacities of each operator, so it is typically exponentially slow in the length of the circuit. However, for constant length circuits making use of operators with low interference producing capacity (many such operators are listed in section 3.6), the simulation runs in time polynomial in the number of qubits.

In general, our technique is able to estimate expressions of the form $\langle\psi|A \cdots Z|\phi\rangle$, of which quantum circuits $\langle\psi|U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)}|\psi\rangle$ are a special case, in time proportional to $\|\psi\|_p^2 \|\bar{A}\|_q^2 \cdots \|\bar{Z}\|_q^2 \|\phi\|_q^2$ for any $1/p + 1/q = 1$ where a bar over a vector or operator denotes entrywise absolute value in the computational basis, and where $\|\cdot\|_p$ denotes the ℓ^p-norm for vectors and the induced norm for operators. The choice $p = q = 2$ is most relevant for quantum mechanics, and $\|\bar{A}\|_2$ gives the interference producing capacity of A. The technique was also generalized to expressions of the form $\text{Tr}\{A \cdots Z\sigma\}$.

We formalized the conditions necessary for efficient simulation by introducing two definitions: EHT for the initial state σ and EPS for the operators $A, \ldots, Z$. These definitions consist of requirements having to do with the number of samples needed as well as requirements having to do with efficient computability. The latter requirements can for the most part be ignored if one is concerned with query complexity or communication complexity rather than time complexity. A wide range of initial

states and operators are EHT or EPS; many examples were listed in section 3.6. In addition to discussing circuits which can be efficiently simulated, we gave several examples of circuits which we cannot efficiently simulate, and explained why.

The choice $p = q = 2$ makes the most sense for simulating expressions of the form $\langle\psi|U^\dagger V^\dagger M V U|\psi\rangle$. However, using the Wigner representation this expression can also be written as $\langle \boldsymbol{M}|\boldsymbol{VU}|\boldsymbol{\rho}\rangle$, and here the choice $p = \infty$ and $q = 1$ works well, allowing efficient simulation of circuits that consist mainly of Clifford operations. We showed how our simulation technique can be applied to communication problems, with the conclusion that there can be no superpolynomial advantage of quantum communication over classical communication unless the quantum protocol uses operations with high interference producing capacity. Curiously, this result does not apply to one-round communication, since our simulation requires doubling the number of rounds. And indeed, there is an example of a one-round quantum protocol with low interference producing capacity which is exponentially more efficient than any one-round classical protocol.

Finally, we would like to suggest three open questions:

1) Can it be shown that interference, rather than interference producing capacity, is necessary for quantum speedup? In section 3.8 we formalized a series of conjectures on this topic, using the framework of communication complexity.

2) While we have shown interference producing capacity to be a necessary resource for quantum speedup, it is also fruitful to investigate sufficient resources for quantum speedup. For example [BaH10], building on the work of [Aar10], showed that any operator U having the property that $\max_{ij} |U_{ij}|$ is sufficiently small can be used to exhibit exponential quantum speedup. Can the gap between necessary (e.g. our result) and the sufficient (e.g. [BaH10]) conditions for quantum speedup be narrowed?

3) Can our technique be combined with existing Monte Carlo or other techniques to provide an improved simulation algorithm for systems of physical interest? Our algorithm in its present form is not likely to be more efficient than existing techniques for such problems.

3.10 Acknowledgments

The author thanks Robert Griffiths and Scott Cohen for many helpful comments and suggestions. This research received financial support from the National Science Foundation through Grant PHY-1068331.

3.A Generalized singular vectors

The goal of this appendix is to determine the minimum value of b such that a given operator A is $\mathrm{EPS}_p(b, f)$ and bounds on b such that an operator σ is $\mathrm{EHT}_p(b, f)$. We will show that conditions (a)-(b) of definition 3.6 require $b \geq \|\bar{A}\|_q$ and will construct probability distributions that satisfy this with equality. Whether these also satisfy conditions (c)-(d) of definition 3.6 needs to be determined on a case by case basis. Note that when $p = q = 2$ we have $\|\bar{A}\|_2 = \mathcal{I}_{\max}(A)$, the interference producing capacity of A. The end result of this appendix is the following theorem.[17]

Theorem 3.22. *Let A and σ be matrices, $p, q \in [1, \infty]$, and $1/p + 1/q = 1$. Then*

(a) It is not possible to satisfy conditions (a)-(b) of definition 3.6 unless $b \geq \|\bar{A}\|_q$. The same goes for (a) and (b) of definition 3.7 since they are stricter (i.e. $b \geq \|\bar{\sigma}\|_q$).

[17] In the case $p = q = 2$, claims (a) and (b) of theorem 3.22 are similar to results of [Mat90], although the techniques are different.

(b) It is possible to satisfy conditions (a)-(b) of definition 3.6 with $b = \|\bar{A}\|_q$. The k index is not needed (i.e. $k \in K = \{0\}$ and $\alpha_{mnk} = A_{mn}$).

(c) If one is concerned with query complexity rather than time complexity, and if A is not defined in terms of an oracle, then conditions (c)-(d) of definition 3.6 can be ignored, as explained in section 3.5.3. Therefore, A is $\mathrm{EPS}_p(\|\bar{A}\|_q, 0)$.

(d) Let w be the smallest value such that σ/w is a convex combination of normalized dyads. That is to say, let

$$w = \min\left\{ \sum_i |s_i| \middle| s_i \in \mathbb{C}, \sigma = \sum_i s_i \boldsymbol{v}^{(i)} \boldsymbol{u}^{(i)\top}, \|\boldsymbol{u}^{(i)}\|_p = \|\boldsymbol{v}^{(i)}\|_q = 1 \right\}. \tag{3.95}$$

It is possible to satisfy conditions (a)-(b) of definition 3.7 with $b = w$ (although this is not necessarily the smallest possible value of b). The k index is not needed (i.e. $k \in K = \{0\}$ and $\alpha_{mnk} = \sigma_{mn}$). Note that when $p = q = 2$, w is the trace norm of σ.

(e) If one is concerned with query complexity rather than time complexity, and if σ is not defined in terms of an oracle, then conditions (c)-(d) of definition 3.7 can be ignored. Therefore, σ is $\mathrm{EHT}_p(w, 0)$ (although this is not necessarily the smallest possible value of b).

We present immediately the proof of parts (a), (d), and (e). Parts (b) and (c) will require more preliminary discussion.

Proof of theorem 3.22(a). Let A be an $M \times N$ matrix. Suppose conditions (a)-(b) of definition 3.6 are satisfied by some b, K, α_{mnk}, $P(n,k|m)$, and $Q(m,k|n)$. Then, for all $m \in \{1,\dots,M\}$, $n \in \{1,\dots,N\}$, and $k \in K$, we have $A_{mn} = \sum_{k' \in K} \alpha_{mnk'}$ and

$$\frac{|\alpha_{mnk}|}{P(n,k|m)^{1/p} Q(m,k|n)^{1/q}} \leq b. \tag{3.96}$$

Rearranging this expression yields

$$|\alpha_{mnk}| \leq b \cdot P(n,k|m)^{1/p} Q(m,k|n)^{1/q}. \tag{3.97}$$

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be nonnegative vectors satisfying $\|\boldsymbol{u}\|_p = \|\boldsymbol{v}\|_q = 1$ and $\boldsymbol{u}^\top \bar{A} \boldsymbol{v} = \|\bar{A}\|_q$ (that such vectors exist is well known, but is also a consequence of theorem 3.25). Multiply both sides of (3.97) by $u_m v_n$ and sum over m, n, k to get

$$\begin{aligned}
\sum_{mnk} u_m \, |\alpha_{mnk}| \, v_n &\leq b \sum_{mnk} u_m P(n,k|m)^{1/p} Q(m,k|n)^{1/q} v_n && (3.98)\\
&= b \sum_{mnk} \left[P(n,k|m) u_m^p\right]^{1/p} \left[Q(m,k|n) v_n^q\right]^{1/q} && (3.99)\\
&\leq b \sum_{mnk} \left[\frac{1}{p} P(n,k|m) u_m^p + \frac{1}{q} Q(m,k|n) v_n^q\right] && (3.100)\\
&= b \sum_m \frac{1}{p} u_m^p + \sum_n \frac{1}{q} v_n^q && (3.101)\\
&= b(1/p + 1/q) && (3.102)\\
&= b && (3.103)
\end{aligned}$$

where (3.100) follows from the inequality of arithmetic and geometric means. We now place a lower bound on the left hand side. By the triangle inequality, $\sum_k |\alpha_{mnk}| \geq |\sum_k \alpha_{mnk}| = |A_{mn}|$ for all

m, n. Since $\boldsymbol{u}$ and $\boldsymbol{v}$ are nonnegative,

$$b \geq \sum_{mnk} u_m \left|\alpha_{mnk}\right| v_n \tag{3.104}$$

$$\geq \sum_{mn} u_m \left|A_{mn}\right| v_n \tag{3.105}$$

$$= \|\bar{A}\|_q. \tag{3.106}$$

□

Proof of theorem 3.22(d)-(e). Let σ be an $M \times N$ matrix. Let s_i, $\boldsymbol{u}^{(i)}$, and $\boldsymbol{v}^{(i)}$ take values achieving the minimum in (3.95). By absorbing phase into $\boldsymbol{u}^{(i)}$ we can assume that the s_i are positive. We then have $w = \sum_i s_i$, $\|\boldsymbol{u}^{(i)}\|_p = \|\boldsymbol{v}^{(i)}\|_q = 1$, and $\sigma = \sum_i s_i \boldsymbol{v}^{(i)} \boldsymbol{u}^{(i)\top}$. Define

$$P(n) = \sum_i \frac{s_i}{w} \left|u_n^{(i)}\right|^p, \tag{3.107}$$

$$Q(m) = \sum_i \frac{s_i}{w} \left|v_m^{(i)}\right|^q. \tag{3.108}$$

Since $\boldsymbol{u}^{(i)}$ and $\boldsymbol{v}^{(i)}$ are normalized for all i, and since $\sum_i s_i/w = 1$, these $P(n)$ and $Q(m)$ are convex combinations of probability distributions and hence are probability distributions themselves.

For any $m \in \{1, \ldots, M\}, n \in \{1, \ldots, N\}$, Hölder's inequality gives

$$\sum_i \frac{s_i^{1/p}}{w^{1/p}} \left|u_n^{(i)}\right| \cdot \frac{s_i^{1/q}}{w^{1/q}} \left|v_m^{(i)}\right| \leq \left[\sum_i \left(\frac{s_i^{1/p}}{w^{1/p}} \left|u_n^{(i)}\right|\right)^p\right]^{1/p} \left[\sum_i \left(\frac{s_i^{1/q}}{w^{1/q}} \left|v_m^{(i)}\right|\right)^q\right]^{1/q} \tag{3.109}$$

$$\implies \sum_i \frac{s_i}{w} \left|u_n^{(i)} v_m^{(i)}\right| \leq \left[\sum_i \frac{s_i}{w} \left|u_n^{(i)}\right|^p\right]^{1/p} \left[\sum_i \frac{s_i}{w} \left|v_m^{(i)}\right|^q\right]^{1/q} \tag{3.110}$$

$$\implies \left|\sum_i \frac{s_i}{w} u_n^{(i)} v_m^{(i)}\right| \leq P(n)^{1/p} Q(m)^{1/q} \tag{3.111}$$

$$\implies \frac{|\sigma_{mn}|}{w} \leq P(n)^{1/p} Q(m)^{1/q} \tag{3.112}$$

$$\implies \frac{|\sigma_{mn}|}{P(n)^{1/p} Q(m)^{1/q}} \leq w \tag{3.113}$$

Therefore conditions (a)-(b) of definition 3.7 are satisfied with $\alpha_{mn0} = \sigma_{mn}$ and $b = w$.

If one is concerned with query complexity rather than time complexity, and if σ is not defined in terms of an oracle, then conditions (c)-(d) of definition 3.7 are satisfied trivially with $f = 0$ since no oracle queries are needed in order to carry out the required operations. So σ is $\mathrm{EHT}_p(w, 0)$. □

We now begin construction of the probability distributions satisfying conditions (a)-(b) of definition 3.6 with $b = \|\bar{A}\|_q$. The bulk of the discussion concerns the $p \in (1, \infty)$ case; the reader interested only in $p = 1$ or $p = \infty$ may skip directly to the second half of the proof of theorem 3.22(b)-(c) at the end of this section.

It suffices to let k take only a single value, say $k = 0$, and to set $\alpha_{mn0} = A_{mn}$. Making this simplification, and plugging in the desired bound $b = \|\bar{A}\|_q$, conditions (a)-(b) of definition 3.6 become

$$\max_{mn} \left\{ \frac{|A_{mn}|}{P(n|m)^{1/p} Q(m|n)^{1/q}} \right\} \leq \|\bar{A}\|_q. \tag{3.114}$$

It will be convenient to derive the probability distributions from a pair of vectors. With A being an $M \times N$ matrix, let $\boldsymbol{u}$ be a positive vector of dimension M and let $\boldsymbol{v}$ be a positive vector of dimension N. Taking the probability distributions

$$P(n|m) = |A_{mn}|\, v_n / [\bar{A}\boldsymbol{v}]_m, \tag{3.115}$$

$$Q(m|n) = |A_{mn}|\, u_m / [\bar{A}^\top \boldsymbol{u}]_n \tag{3.116}$$

brings (3.114) to the form

$$\max_{mn} \left\{ \left(\frac{[\bar{A}\boldsymbol{v}]_m}{v_n} \right)^{1/p} \left(\frac{[\bar{A}^\top \boldsymbol{u}]_n}{u_m} \right)^{1/q} \right\} \leq \|\bar{A}\|_q. \tag{3.117}$$

Consider for a moment the case $p = q = 2$. If $\bar{A}$ is not block diagonal (even under permutations of rows and columns) then the left and right singular vectors of $\bar{A}$ will be positive. Taking these for $\boldsymbol{u}$ and $\boldsymbol{v}$ it is easy to see that (3.117) holds. If $p \neq 2$ we can use a sort of generalization of singular vectors: we will show the existence of positive vectors satisfying

$$(\bar{A}^\top \boldsymbol{u})_n \leq v_n^{q/p} \|\bar{A}\|_q, \tag{3.118}$$

$$(\bar{A}\boldsymbol{v})_m \leq u_m^{p/q} \|\bar{A}\|_q. \tag{3.119}$$

These vectors are easily seen to satisfy (3.117). If $\bar{A}$ is not block diagonal then $\boldsymbol{u}$ and $\boldsymbol{v}$ can be computed using the power method [Boy74, BV11] since $\bar{A}$ is nonnegative. In this case the inequalities (3.118)-(3.119) become equalities. On the other hand, if $\bar{A}$ is block diagonal then $\boldsymbol{u}$ and $\boldsymbol{v}$ can be built from the generalized left and right singular vectors of each block. The rest of this section is devoted to proving the existence of such vectors.

First we will need some basic facts about ℓ^p-norms. If $\boldsymbol{v}$ is a real vector normalized under the ℓ^2-norm then $\boldsymbol{u} = \boldsymbol{v}$ is the unique ℓ^2-normalized vector with the property that $\boldsymbol{u}^\top \boldsymbol{v} = 1$. This generalizes to arbitrary ℓ^p-norms, with some adaptation.

Definition 3.23. *Let $p, q \in [1, \infty]$ and $1/p + 1/q = 1$. Let $\boldsymbol{v} \in \ell^q$. Any $\boldsymbol{u} \in \ell^p$ satisfying the conditions $\boldsymbol{u}^\top \boldsymbol{v} = \|\boldsymbol{v}\|_q$ and $\|\boldsymbol{u}\|_p = 1$ is called a support functional of $\boldsymbol{v}$.*

Lemma 3.24. *Let $p, q \in (1, \infty)$ and $1/p + 1/q = 1$. For any nonzero $\boldsymbol{v} \in \ell^q$, the vector $\boldsymbol{u} \in \ell^p$ defined by*

$$u_i = \|\boldsymbol{v}\|_q^{-q/p}\, |v_i|^{q/p} \operatorname{sgn}(v_i) \tag{3.120}$$

is the unique support functional of $\boldsymbol{v}$. Similarly, for any nonzero $\boldsymbol{u} \in \ell^p$, the vector $\boldsymbol{v} \in \ell^q$ defined by

$$v_i = \|\boldsymbol{u}\|_p^{-p/q}\, |u_i|^{p/q} \operatorname{sgn}(u_i) \tag{3.121}$$

is the unique support functional of $\boldsymbol{u}$.

Proof. Uniqueness of the support functional when $1 < p < \infty$ follows from strict convexity of the norm (chapter 11 of [Car04]). That the specific vectors (3.120) and (3.121) are support functionals is easily verified through direct computation [Arm10]. □

We now describe generalized singular vectors. Ordinary ($p = 2$) left and right singular vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfy $\|A\boldsymbol{v}\|_2 = \|A^\top \boldsymbol{u}\|_2 = \|A\|_2$, furthermore $\boldsymbol{u}$ is the support functional of $A\boldsymbol{v}$ (since $p = 2$ this just means that $\boldsymbol{u} \propto A\boldsymbol{v}$), and $\boldsymbol{v}$ is the support functional of $A^\top \boldsymbol{u}$. These properties generalize to arbitrary ℓ^p-norms, as we now show.

Theorem 3.25. *Let $p, q \in [1, \infty]$ and $1/p + 1/q = 1$. Let A be a matrix. Then there are vectors $\boldsymbol{u} \in \ell^p$ and $\boldsymbol{v} \in \ell^q$ such that*

(a) $\|\boldsymbol{u}\|_p = \|\boldsymbol{v}\|_q = 1$

(b) $\boldsymbol{u}^\top A\boldsymbol{v} = \|A^\top \boldsymbol{u}\|_p = \|A\boldsymbol{v}\|_q = \|A\|_q = \|A^\top\|_p$

(c) $\boldsymbol{u}$ *is a support functional of* $A\boldsymbol{v}$

(d) $\boldsymbol{v}$ *is a support functional of* $A^\top \boldsymbol{u}$.

(e) If A *is nonnegative then* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are nonnegative.*

Proof. Let $\boldsymbol{v}$ be a vector satisfying $\|\boldsymbol{v}\|_q = 1$ and $\|A\boldsymbol{v}\|_q = \|A\|_q$. Such a vector is guaranteed to exist (see definition 5.6.1 of [HJ90]). Let $\boldsymbol{u}$ be a support functional of $A\boldsymbol{v}$. By the definition of a support functional, $\|\boldsymbol{u}\|_p = 1$ so claims (a) and (c) have been proved. With these two vectors defined, we have

$$\begin{aligned}
\|A\|_q &= \|A\boldsymbol{v}\|_q && && (3.122)\\
&= \boldsymbol{u}^\top A\boldsymbol{v} && (\boldsymbol{u} \text{ is the support functional of } A\boldsymbol{v}) && (3.123)\\
&= \boldsymbol{v}^\top (A^\top \boldsymbol{u}) && && (3.124)\\
&\le \|\boldsymbol{v}\|_q \|A^\top \boldsymbol{u}\|_p && (\text{Hölder's inequality}) && (3.125)\\
&= \|A^\top \boldsymbol{u}\|_p && && (3.126)\\
&\le \|A^\top\|_p \|\boldsymbol{u}\|_p && && (3.127)\\
&= \|A^\top\|_p. && && (3.128)
\end{aligned}$$

By symmetry we also have $\|A^\top\|_p \le \|A\|_q$, therefore the inequalities become equalities. Claim (b) is proved. Since $\|\boldsymbol{v}\|_q = 1$ and $\boldsymbol{v}^\top (A^\top \boldsymbol{u}) = \|A^\top \boldsymbol{u}\|_p$, claim (d) is proved as well.

To prove claim (e), assume that A is nonnegative. Then $\|\bar{\boldsymbol{u}}\|_p = \|\bar{\boldsymbol{v}}\|_q = 1$ and $\|A\bar{\boldsymbol{v}}\|_q \ge \bar{\boldsymbol{u}}^\top A\bar{\boldsymbol{v}} \ge \boldsymbol{u}^\top A\boldsymbol{v} = \|A\|_q$. It follows that $\|A\bar{\boldsymbol{v}}\|_q = \|A\|_q$, thus $\bar{\boldsymbol{u}}$ is a support functional of $A\bar{\boldsymbol{v}}$. Therefore $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$ could have been taken instead of $\boldsymbol{u}$ and $\boldsymbol{v}$ in the first steps of this proof, justifying the claim that $\boldsymbol{u}$ and $\boldsymbol{v}$ can be chosen to be nonnegative. □

The Perron–Frobenius theorem states that an irreducible nonnegative matrix has a first eigenvector that has positive components. A similar statement holds for the first singular vector: if $\bar{A}$ is a nonnegative matrix that is not block diagonal then the left and right singular vectors associated with the largest singular value of $\bar{A}$ have positive entries. This is true also for our generalized singular vectors, as we now show.

Definition 3.26. *A matrix* A *is* block diagonal *if there are permutation matrices* σ *and* τ *such that* A *can be decomposed as* $\bar{A} = \sigma^\top (A^{(1)} \oplus \cdots \oplus A^{(L)} \oplus \mathbf{0}^{M\times N})\tau$ *where the* $A^{(l)}$ *are nonzero and have nonvanishing dimension, and at least one of the inequalities* $L > 1$, $M > 0$, *or* $N > 0$ *holds.*[18] *A matrix is* not block diagonal *if no such decomposition is possible. In particular, a matrix that is not block diagonal has no totally zero rows or columns.*

Lemma 3.27. *Let* $q \in (1, \infty)$. *Let* $\bar{A}$ *be a nonnegative matrix that is not block diagonal. Let* $\boldsymbol{v}$ *be a nonzero, nonnegative vector that maximizes* $\|\bar{A}\boldsymbol{v}\|_q / \|\boldsymbol{v}\|_q$. *Then* $\boldsymbol{v}$ *is in fact a positive vector (has no zero entries).*

Proof. Let $Z = \{i : v_i = 0\}$. This will be a proof by contradiction; suppose that $\boldsymbol{v}$ has at least one zero entry, so that Z is nonempty. Since $\boldsymbol{v} \neq 0$, the complement Z^C is nonempty, therefore Z and Z^C partition the entries of $\boldsymbol{v}$ into two nonempty sets. Also, Z and Z^C can be considered as a partition of the columns of $\bar{A}$. Since $\bar{A}$ is not block diagonal, there must be indices $i \in Z$, $j \notin Z$, and k such that $\bar{A}_{ki} > 0$ and $\bar{A}_{kj} > 0$. We will show that $\boldsymbol{v}$ cannot maximize $\|\bar{A}\boldsymbol{v}\|_q / \|\boldsymbol{v}\|_q$ by showing that $\boldsymbol{v}$ is

[18] If $M > 0, N = 0$ then $\oplus \mathbf{0}^{M\times N}$ adds M rows of zeros. Similarly, if $M = 0, N > 0$ then $\oplus \mathbf{0}^{M\times N}$ adds N columns of zeros.

not a critical point of $\|\bar{A}\boldsymbol{v}\|_q/\|\boldsymbol{v}\|_q$, or equivalently of $\|\bar{A}\boldsymbol{v}\|_q^q/\|\boldsymbol{v}\|_q^q$. Without loss of generality take $\|\boldsymbol{v}\|_q = 1$. Let $\hat{\imath}$ be the unit vector corresponding to i. We have

$$\frac{\partial}{\partial\alpha}\frac{\|\bar{A}(\boldsymbol{v}+\alpha\hat{\imath})\|_q^q}{\|\boldsymbol{v}+\alpha\hat{\imath}\|_q^q}\bigg|_{\alpha=0} = \frac{\left(\frac{\partial}{\partial\alpha}\|\bar{A}(\boldsymbol{v}+\alpha\hat{\imath})\|_q^q\right)\|\boldsymbol{v}\|_q^q - \|\bar{A}\boldsymbol{v}\|_q^q\left(\frac{\partial}{\partial\alpha}\|\boldsymbol{v}+\alpha\hat{\imath}\|_q^q\right)}{\|\boldsymbol{v}\|_q^{2q}}\bigg|_{\alpha=0} \tag{3.129}$$

$$= \frac{\partial}{\partial\alpha}\|\bar{A}(\boldsymbol{v}+\alpha\hat{\imath})\|_q^q\bigg|_{\alpha=0} \tag{3.130}$$

$$= \frac{\partial}{\partial\alpha}\sum_l([\bar{A}\boldsymbol{v}]_l + \alpha\bar{A}_{li})^q\bigg|_{\alpha=0} \tag{3.131}$$

$$= \sum_l q\bar{A}_{li}[\bar{A}\boldsymbol{v}]_l^{q-1} \tag{3.132}$$

$$\geq q\bar{A}_{ki}[\bar{A}\boldsymbol{v}]_k^{q-1} \tag{3.133}$$

$$\geq q\bar{A}_{ki}(\bar{A}_{kj}v_j)^{q-1} \tag{3.134}$$

$$> 0. \tag{3.135}$$

Equality (3.130) follows from $\|\boldsymbol{v}\|_q = 1$ as well as ($v_i = 0 \implies \partial\|\boldsymbol{v}+\alpha\hat{\imath}\|_q^q/\partial\alpha = 0$). Inequality (3.133) follows from each term of the previous summation being nonnegative. Inequality (3.134) follows from each term of the sum $[\bar{A}\boldsymbol{v}]_k = \sum_n \bar{A}_{kn}v_n$ being nonnegative. □

Theorem 3.28. *Let $p, q \in (1,\infty)$ and $1/p + 1/q = 1$. Let $\bar{A}$ be a nonnegative matrix that is not block diagonal. Then there are positive vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfying*

$$(\bar{A}^\top\boldsymbol{u})_n = v_n^{q/p}\|\bar{A}\|_q, \tag{3.136}$$

$$(\bar{A}\boldsymbol{v})_m = u_m^{p/q}\|\bar{A}\|_q. \tag{3.137}$$

Note: if $p = q = 2$ then $\boldsymbol{u}$ and $\boldsymbol{v}$ will be the left and right singular vectors associated with the largest singular value of $\bar{A}$.

Proof. Theorem 3.25 guarantees the existence of nonnegative vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ that satisfy $\|\boldsymbol{u}\|_p = \|\boldsymbol{v}\|_q = 1$ and $\boldsymbol{u}^\top\bar{A}\boldsymbol{v} = \|\bar{A}\|_q = \|\bar{A}^\top\|_p$ with $\boldsymbol{u}$ being the support functional of $A\boldsymbol{v}$ and $\boldsymbol{v}$ being the support functional of $A^\top\boldsymbol{u}$. Lemma 3.24 give the exact form of these support functionals:

$$u_m = \|\bar{A}\boldsymbol{v}\|_q^{-q/p}(\bar{A}\boldsymbol{v})_m^{q/p}\mathrm{sgn}(\bar{A}\boldsymbol{v}) \tag{3.138}$$

$$v_n = \|\bar{A}^\top\boldsymbol{u}\|_p^{-p/q}(\bar{A}^\top\boldsymbol{u})_n^{p/q}\mathrm{sgn}(\bar{A}^\top\boldsymbol{u}). \tag{3.139}$$

Since $\bar{A}$, $\boldsymbol{u}$, and $\boldsymbol{v}$ are nonnegative, the sgn functions disappear. Theorem 3.25 gives $\|\bar{A}\boldsymbol{v}\|_q = \|\bar{A}^\top\boldsymbol{u}\|_p = \|\bar{A}\|_q$. With these simplifications, we get (3.136)-(3.137). That $\boldsymbol{u}$ and $\boldsymbol{v}$ have nonzero entries follows from Lemma 3.27. □

We now generalize theorem 3.28 to matrices that are not block diagonal. This is done by applying theorem 3.28 to each individual block of the matrix. Each block of $\bar{A}$ may have a different operator norm, but each of these is upper bounded by $\|\bar{A}\|_q$. For this reason, we end up with an inequality rather than an equality when generalizing (3.136)-(3.137).

Theorem 3.29. *Let $p, q \in (1,\infty)$ and $1/p+1/q = 1$. Let $\bar{A}$ be a nonnegative matrix that can possibly be block diagonal and that may have some totally zero rows or columns. Then there are positive vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfying*

$$(\bar{A}^\top\boldsymbol{u})_n \leq v_n^{q/p}\|\bar{A}\|_q, \tag{3.140}$$

$$(\bar{A}\boldsymbol{v})_m \leq u_m^{p/q}\|\bar{A}\|_q. \tag{3.141}$$

Proof. Let σ and τ be permutations matrices that bring out the block structure of $\bar{A}$, and let $A^{(1)}, \ldots, A^{(L)}$ be the blocks. Specifically, suppose $\sigma^\top(A^{(1)} \oplus \cdots \oplus A^{(L)} \oplus \mathbf{0}^{M\times N})\tau = \bar{A}$ where the $A^{(1)} \cdots A^{(L)}$ matrices are not block diagonal and $\mathbf{0}^{M\times N}$ is an M-by-N matrix of zeros (if there is no zero block then just take $M = N = 0$). It is easy to see that $\|A^{(l)}\|_q \leq \|\bar{A}\|_q$ for all $l \in \{1, \ldots, L\}$.

By theorem 3.28, there are positive vectors $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(L)}$ and $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}$ such that

$$(A^{(l)\top}\boldsymbol{u}^{(l)})_n = v_n^{(l)q/p}\|A^{(l)}\|_q \tag{3.142}$$
$$\leq v_n^{(l)q/p}\|\bar{A}\|_q, \tag{3.143}$$
$$(A^{(l)}\boldsymbol{v}^{(l)})_m = u_m^{(l)p/q}\|A^{(l)}\|_q \tag{3.144}$$
$$\leq u_m^{(l)p/q}\|\bar{A}\|_q \tag{3.145}$$

for all $l \in \{1, \ldots, L\}$. Define $\boldsymbol{u} = \sigma^\top(\boldsymbol{u}^{(1)} \oplus \cdots \oplus \boldsymbol{u}^{(L)} \oplus \mathbf{1}^M)$ and $\boldsymbol{v} = \tau^\top(\boldsymbol{v}^{(1)} \oplus \cdots \oplus \boldsymbol{v}^{(L)} \oplus \mathbf{1}^N)$ where $\mathbf{1}^M$ and $\mathbf{1}^N$ are the all-ones vectors of lengths M and N, respectively. Then (3.142)-(3.145) imply (3.140)-(3.141). Since the $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(L)}$ and $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}$ are positive, $\boldsymbol{u}$ and $\boldsymbol{v}$ are positive. □

We are now ready to complete the proof of theorem 3.22.

Proof of theorem 3.22(b)-(c) . Let A be a matrix. Set $K = \{0\}$ and $\alpha_{mn0} = A_{mn}$. Clearly condition (a) of definition 3.6 is satisfied.

Consider the case $p \in (1, \infty)$. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be positive vectors satisfying (3.140)-(3.141). The existence of such vectors is guaranteed by theorem 3.29. Define the probability distributions

$$P(n|m) = |A_{mn}|\, v_n/[\bar{A}\boldsymbol{v}]_m, \tag{3.146}$$
$$Q(m|n) = |A_{mn}|\, u_m/[\bar{A}^\top\boldsymbol{u}]_n. \tag{3.147}$$

These satisfy condition (b) of definition 3.6 with $b = \|\bar{A}\|_q$ since

$$\max_{mnk}\left\{\frac{|\alpha_{mnk}|}{P(n|m)^{1/p}Q(m|n)^{1/q}}\right\} = \max_{mn}\left\{\frac{|A_{mn}|}{P(n|m)^{1/p}Q(m|n)^{1/q}}\right\} \tag{3.148}$$
$$= \max_{mn}\left\{\left(\frac{[\bar{A}\boldsymbol{v}]_m}{v_n}\right)^{1/p}\left(\frac{[\bar{A}^\top\boldsymbol{u}]_n}{u_m}\right)^{1/q}\right\} \tag{3.149}$$
$$\leq \max_{mn}\left\{\left(\frac{u_m^{p/q}\|\bar{A}\|_q}{v_n}\right)^{1/p}\left(\frac{v_n^{q/p}\|\bar{A}\|_q}{u_m}\right)^{1/q}\right\} \tag{3.150}$$
$$= \|\bar{A}\|_q. \tag{3.151}$$

Now consider the case $p = 1$, $q = \infty$ (the case $p = \infty$, $q = 1$ follows by a symmetrical argument). Define $P(n|m) = |A_{mn}| \,/\, \sum_{n'} |A_{mn'}|$ and define $Q(m|n)$ arbitrarily. Condition (b) of definition 3.6 is satisfied with $b = \|\bar{A}\|_\infty$ since

$$\max_{mnk}\left\{\frac{|\alpha_{mnk}|}{P(n|m)^{1/p}Q(m|n)^{1/q}}\right\} = \max_{mn}\left\{\frac{|A_{mn}|}{P(n|m)^{1}Q(m|n)^{0}}\right\} \tag{3.152}$$
$$= \max_{mn}\left\{\frac{|A_{mn}|}{|A_{mn}| \,/\, \sum_{n'} |A_{mn'}|}\right\} \tag{3.153}$$
$$\leq \|\bar{A}\|_\infty. \tag{3.154}$$

If one is concerned with query complexity rather than time complexity, and if A is not defined in terms of an oracle, then conditions (c)-(d) of definition 3.6 are satisfied trivially with $f = 0$ since no oracle queries are needed in order to carry out the required operations. So A is $\mathrm{EPS}_p(\|\bar{A}\|_q, 0)$. □

3.B Proofs for section 3.5

In this section we prove theorem 3.10 and lemma 3.11. The proofs are conceptually rather simple, however they are notationally tedious. Since we will at times be manipulating infinite series, we begin by showing that these series converge absolutely. This will be useful, since absolutely convergent series allow permutation of terms and reordering of double summations.

Lemma 3.30. *Let b and α_{mnk} satisfy condition (b) of definition 3.6. Then series $\sum_{k\in K}\alpha_{mnk}$ is absolutely convergent for all m, n, and $\sum_{k\in K}|\alpha_{mnk}| \le b$.*

Proof. Rearranging (3.55) of condition (b) gives, for all m, n, k,

$$|\alpha_{mnk}| \le b \cdot P(n,k|m)^{1/p} Q(m,k|n)^{1/q} \tag{3.155}$$
$$\le b \cdot [P(n,k|m)/p + Q(m,k|n)/q]\,. \tag{3.156}$$

Therefore,

$$\sum_{k\in K}|\alpha_{mnk}| \le b \sum_{k\in K}[P(n,k|m)/p + Q(m,k|n)/q] \tag{3.157}$$
$$= b \cdot [P(n|m)/p + Q(m|n)/q] \tag{3.158}$$
$$\le b \tag{3.159}$$
$$< \infty. \tag{3.160}$$

□

We now prove that linear combinations of EPS operators are EPS. Theorem 3.10(a), regarding sums of EPS operators, follows as a corollary. This will also be used to prove theorem 3.10(c), regarding exponentials of EPS operators.

Theorem 3.31 (Linear combination of EPS)**.** *Let L be a finite or countable set. For $l \in L$ let s_l be a complex number and let $A^{(l)}$ be an $M \times N$ matrix that is $\mathrm{EPS}_p(b_l, f_l)$ for some f_l and b_l. Let $W(l)$ be a probability distribution[19] on l such that $W(l)$ can be sampled from, and $s_l/W(l)$ computed, in average time $\mathrm{O}(f_0)$. Let $b := \max_l\{|s_l|\, b_l/W(l)\} < \infty$ and $f := f_0 + \sum_l W(l) f_l$. Then $\sum_l s_l A^{(l)}$ is $\mathrm{EPS}_p(b, f)$.*

Proof. For each $l \in L$, $A^{(l)}$ is $\mathrm{EPS}_p(b_l, f_l)$ so there are K_l, $\alpha^{(l)}_{mnk}$, $P_l(n,k|m)$, and $Q_l(m,k|n)$ satisfying definition 3.6. Let $K = L \times \cup_{l\in L} K_l$. For $(l,k) \in K$ define

$$\alpha_{mn(l,k)} = \begin{cases} s_l \alpha^{(l)}_{mnk} & \text{if } k \in K_l \\ 0 & \text{otherwise.} \end{cases} \tag{3.161}$$

We first show that $\sum_{(l,k)\in K}\alpha_{mn(l,k)}$ is absolutely convergent, so that it can be expressed as a double sum. By lemma 3.30, $\sum_{k\in K_l}\left|\alpha^{(l)}_{mnk}\right| \le b_l$ for all $l \in L$, therefore

$$\sum_{(l,k)\in K}\left|\alpha_{mn(l,k)}\right| = \sum_{l\in L}|s_l| \sum_{k\in K_l}\left|\alpha^{(l)}_{mnk}\right| \tag{3.162}$$
$$\le \sum_{l\in L}|s_l b_l| \tag{3.163}$$
$$\le b. \tag{3.164}$$

[19] The lowest b is obtained when $W(l)$ is proportional to $|s_l|\, b_l$.

Since $b < \infty$ by assumption, the series $\sum_{(l,k)\in K} \alpha_{mn(l,k)}$ is absolutely convergent. We can then decompose it as a double series,

$$\sum_{(l,k)\in K} \alpha_{mn(l,k)} = \sum_{l\in L} s_l \sum_{k\in K_l} \alpha^{(l)}_{mnk} \tag{3.165}$$

$$= \sum_{l\in L} s_l A^{(l)}, \tag{3.166}$$

showing that condition (a) of definition 3.6 is satisfied.

Define the probability distributions

$$P(n,(l,k)|m) = \begin{cases} W(l)P_l(n,k|m) & \text{if } k \in K_l \\ 0 & \text{otherwise} \end{cases} \tag{3.167}$$

$$Q(m,(l,k)|n) = \begin{cases} W(l)Q_l(m,k|n) & \text{if } k \in K_l \\ 0 & \text{otherwise.} \end{cases} \tag{3.168}$$

We now show that condition (b) holds. Let $m \in \{1,\dots,M\}$, $n \in \{1,\dots,N\}$, and $(l,k) \in K$. We need only consider $k \in K_l$ since otherwise $\alpha_{mn(l,k)}$ vanishes.

$$\frac{\left|\alpha_{mn(l,k)}\right|}{\left[P(n,(l,k)|m)\right]^{1/p}\left[Q(m,(l,k)|n)\right]^{1/q}} = \frac{\left|s_l\alpha^{(l)}_{mnk}\right|}{\left[W(l)P_l(n,k|m)\right]^{1/p}\left[W(l)Q_l(m,k|n)\right]^{1/q}} \tag{3.169}$$

$$= \frac{|s_l|}{W(l)} \cdot \frac{\left|\alpha^{(l)}_{mnk}\right|}{\left[P_l(n,k|m)\right]^{1/p}\left[Q_l(m,k|n)\right]^{1/q}} \tag{3.170}$$

$$\le |s_l|\, b_l/W(l) \tag{3.171}$$

$$\le b. \tag{3.172}$$

Condition (c) requires that the distribution $P(n,(l,k)|m)$ can be sampled from, and the values $\alpha_{mn(l,k)}/P(n,(l,k)|m)$ and $\alpha_{mn(l,k)}/Q(m,(l,k)|n)$ can be computed, in average time $\mathrm{O}(f) = \mathrm{O}(f_0 + \sum_l W(l) f_l)$. This can be accomplished as follows:

(i) Draw l according to the distribution $W(l)$ and compute $s_l/W(l)$. This can be done in average time $\mathrm{O}(f_0)$.

(ii) Draw n,k according to the distribution $P_l(n,k|m)$ and compute $\alpha^{(l)}_{mnk}/P_l(n,k|m)$ and $\alpha^{(l)}_{mnk}/Q_l(m,k|n)$. This can be done in average time $\mathrm{O}(f_l)$.

(iii) The quantities $\alpha_{mn(l,k)}/P(n,(l,k)|m)$ and $\alpha_{mn(l,k)}/Q(m,(l,k)|n)$ can be directly computed from (3.161), (3.167), and (3.168) in time $\mathrm{O}(1)$ given the quantities that have been computed in the previous two steps.

The average time needed for a given l is $\mathrm{O}(f_0 + f_l)$, therefore the average time needed given that l is drawn according to $W(l)$ is $\mathrm{O}(f) = \mathrm{O}(f_0 + \sum_l W(l) f_l)$. Condition (c) is satisfied. Condition (d) follows from a symmetric argument. □

Proof of theorem 3.10(a). This follows directly from theorem 3.31. Specifically, apply theorem 3.31 with $L = \{A,B\}$, $s_A = s_B = 1$, $W(A) = b_A/(b_A+b_B)$, and $W(B) = b_B/(b_A+b_B)$. Then $b = \max_l\{|s_l|\, b_l/W(l)\} = b_A + b_B$ and $f = \mathrm{O}(1) + \sum_l W(l) f_l = \mathrm{O}(\max\{b_A, b_B\})$. □

Proof of theorem 3.10(b). Since A is $\mathrm{EPS}_p(b_A, f_A)$, there are K_A, $\alpha^{(A)}_{lmk}$, $P_A(m,k|l)$, and $Q_A(l,k|m)$ satisfying definition 3.6 with $l \in \{1,\dots,L\}$, $m \in \{1,\dots,M\}$, and $k \in K_A$. Likewise, since B

is $\mathrm{EPS}_p(b_B, f_B)$, there are K_B, $\alpha^{(B)}_{mnk}$, $P_B(n,k|m)$, and $Q_B(m,k|n)$ satisfying definition 3.6 with $m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, and $k \in K_B$.

Let $K = K_A \times K_B \times \{1,\ldots,M\}$ and

$$\alpha_{ln(k',k'',m)} = \alpha^{(A)}_{lmk'}\alpha^{(B)}_{mnk''}. \tag{3.173}$$

We first show that $\sum_{(k',k'',m)\in K} \alpha_{ln(k',k'',m)}$ is absolutely convergent, so that it can be expressed as a double series. By lemma 3.30, $\sum_{k'\in K_A} \left|\alpha^{(A)}_{lmk'}\right| \le b_A$ and $\sum_{k''\in K_B} \left|\alpha^{(B)}_{mnk''}\right| \le b_B$, therefore

$$\sum_{(k',k'',m)\in K} \left|\alpha_{ln(k',k'',m)}\right| = \sum_{m\in\{1,\ldots,M\}} \sum_{k'\in K_A} \left|\alpha^{(A)}_{lmk'}\right| \sum_{k''\in K_B} \left|\alpha^{(B)}_{mnk''}\right| \tag{3.174}$$

$$\le M b_A b_B \tag{3.175}$$

$$\le \infty. \tag{3.176}$$

Being absolutely convergent, $\sum_{(k',k'',m)\in K} \alpha_{ln(k',k'',m)}$ can be expressed as a double series, giving

$$\sum_{(k',k'',m)\in K} \alpha_{ln(k',k'',m)} = \sum_{m\in\{1,\ldots,M\}} \sum_{k'\in K_A} \alpha^{(A)}_{lmk'} \sum_{k''\in K_B} \alpha^{(B)}_{mnk''} \tag{3.177}$$

$$= \sum_m A_{lm}B_{mn} \tag{3.178}$$

$$= (AB)_{ln} \tag{3.179}$$

so condition (a) of definition 3.6 is satisfied.

Define the probability distributions

$$P(n,(k',k'',m)|l) = P_A(m,k'|l)P_B(n,k''|m), \tag{3.180}$$

$$Q(l,(k',k'',m)|n) = Q_A(l,k'|m)Q_B(m,k''|n). \tag{3.181}$$

These satisfy condition (b) of definition 3.6 since for all l,m,n,k',k'',

$$b_A b_B \ge \frac{\left|\alpha^{(A)}_{lmk'}\right|}{P_A(m,k'|l)^{1/p}Q_A(l,k'|m)^{1/q}} \frac{\left|\alpha^{(B)}_{mnk''}\right|}{P_B(n,k''|m)^{1/p}Q_B(m,k''|n)^{1/q}} \tag{3.182}$$

$$= \frac{\left|\alpha_{ln(k',k'',m)}\right|}{P(n,(k',k'',m)|l)^{1/p}Q(l,(k',k'',m)|n)^{1/q}}. \tag{3.183}$$

Condition (c) requires that it be possible in average time $\mathrm{O}(f_A + f_B)$ to sample from the probability distribution $P(n,(k',k'',m)|l)$ and to compute $\frac{\alpha_{ln(k',k'',m)}}{P(n,(k',k'',m)|l)}$ and $\frac{\alpha_{ln(k',k'',m)}}{Q(l,(k',k'',m)|n)}$. This can be accomplished as follows:

(i) Draw m,k' from $P_A(m,k'|l)$ and compute $\frac{\alpha^{(A)}_{lmk'}}{P_A(m,k'|l)}$ and $\frac{\alpha^{(A)}_{lmk'}}{Q_A(l,k'|m)}$. This can be done in average time $\mathrm{O}(f_A)$.

(ii) Draw n,k'' from $P_B(n,k''|m)$ and compute $\frac{\alpha^{(B)}_{mnk''}}{P_B(n,k''|m)}$ and $\frac{\alpha^{(B)}_{mnk''}}{Q_B(m,k''|n)}$. This can be done in average time $\mathrm{O}(f_B)$.

(iii) Compute

$$\frac{\alpha_{ln(k',k'',m)}}{P(n,(k',k'',m)|l)} = \frac{\alpha^{(A)}_{lmk'}}{P_A(m,k'|l)} \cdot \frac{\alpha^{(B)}_{mnk''}}{P_B(n,k''|m)} \tag{3.184}$$

$$\frac{\alpha_{ln(k',k'',m)}}{Q(l,(k',k'',m)|n)} = \frac{\alpha^{(A)}_{lmk'}}{Q_A(l,k'|m)} \cdot \frac{\alpha^{(B)}_{mnk''}}{Q_B(m,k''|n)}. \tag{3.185}$$

This can be done in time O(1) since the factors on the right hand sides of these expressions have already been computed in the previous two steps.

So condition (c) is satisfied. Condition (d) follows from a symmetric argument. □

Proof of theorem 3.10(c). Let A be a square matrix that is $\text{EPS}_p(b, f)$. We will show that e^A is $\text{EPS}_p(e^b, bf)$.

This follows from applying theorem 3.31 and theorem 3.10(b) to $e^A = \sum_{j=0}^{\infty} A^j/j!$. Specifically, let $L = \{0, 1, \dots\}$, $A^{(l)} = A^l$, $s_l = 1/l!$, and $W(l) = b^l/(l!e^b)$. By repeated application of theorem 3.10(b), $A^{(l)}$ is $\text{EPS}_p(b^l, lf)$. Assume for now that $W(l)$ can be sampled in average time $\text{O}(b)$. Then by theorem 3.31, $e^A = \sum_{j=0}^{\infty} A^j/j!$ is $\text{EPS}_p(b', f')$ with $b' = \max_l\{|s_l|\, b_l/W(l)\} = e^b$ and

$$f' = b + \sum_{l=0}^{\infty} W(l) f_l \tag{3.186}$$

$$= b + \sum_{l=0}^{\infty} \frac{l f b^l}{l! e^b} \tag{3.187}$$

$$= b + \frac{bf}{e^b} \sum_{l=1}^{\infty} \frac{b^{l-1}}{(l-1)!} \tag{3.188}$$

$$= b + bf \tag{3.189}$$

$$= \text{O}(bf) \tag{3.190}$$

It remains only to show that $W(l)$ can be sampled in time $\text{O}(b)$. The procedure is as follows. Flip a weighted coin that lands heads with probability $W(0)$, and if it lands heads take $l = 0$. This can be done in time O(1). If the coin landed tails then flip another coin that lands heads with probability $W(1)/(1 - W(0))$, and if it lands heads take $l = 1$. Continue, each iteration flipping a coin that lands heads with probability $W(l)/(1 - \sum_{j=0}^{l-1} W(j))$. Each iteration requires computing $W(l)/(1 - \sum_{j=0}^{l-1} W(j))$, which in turn requires computing $W(l)$ and updating the partial sum with the previous $W(l-1)$. This can be done in O(1) time. The expected number of iterations is $\sum_l l W(l) = b$. Therefore, this sampling algorithm takes average time b. □

Proof of lemma 3.11. Since σ is $\text{EHT}_p(b_\sigma, f_\sigma)$, there are $\alpha^{(\sigma)}_{nmk}$, $P_\sigma(m, k)$, and $Q_\sigma(n, k)$ with $k \in K_\sigma$ satisfying definition 3.7 (note that m and n have been swapped since σ is an $N \times M$ operator). Similarly, since A is $\text{EPS}_p(b_A, f_A)$ there are $\alpha^{(A)}_{mnk'}$, $P_A(n, k'|m)$, and $Q_A(m, k'|n)$ with $k' \in K_A$ satisfying definition 3.6.

We have

$$\text{Tr}(A\sigma) = \sum_{mn} A_{mn}\sigma_{nm} \tag{3.191}$$

$$= \sum_{mnkk'} \alpha^{(A)}_{mnk'} \alpha^{(\sigma)}_{nmk}. \tag{3.192}$$

Define the probability distribution

$$R(m, n, k, k') = \frac{1}{p} P_\sigma(m, k) P_A(n, k'|m) + \frac{1}{q} Q_\sigma(n, k) Q_A(m, k'|n). \tag{3.193}$$

By the inequality of arithmetic and geometric means,

$$R(m, n, k, k') \geq [P_\sigma(m, k) P_A(n, k'|m)]^{1/p} [Q_\sigma(n, k) Q_A(m, k'|n)]^{1/q}. \tag{3.194}$$

Setting $V(m,n,k,k') = \alpha^{(A)}_{mnk'}\alpha^{(\sigma)}_{nmk}$ we get the bound

$$b_{\max} := \max_{mnkk'} \left\{ \frac{|V(m,n,k,k')|}{R(m,n,k,k')} \right\} \tag{3.195}$$

$$\leq \max_{mnkk'} \left\{ \frac{\left|\alpha^{(A)}_{mnk'}\alpha^{(\sigma)}_{nmk}\right|}{[P_\sigma(m,k)P_A(n,k'|m)]^{1/p}[Q_\sigma(n,k)Q_A(m,k'|n)]^{1/q}} \right\} \tag{3.196}$$

$$\leq \max_{mnk'} \left\{ \frac{\left|\alpha^{(A)}_{mnk'}\right|}{P_A(n,k'|m)^{1/p}Q_A(m,k'|n)]^{1/q}} \right\} \cdot \max_{mnk} \left\{ \frac{\left|\alpha^{(\sigma)}_{nmk}\right|}{P_\sigma(m,k)^{1/p}Q_\sigma(n,k)^{1/q}} \right\} \tag{3.197}$$

$$\leq b_A b_\sigma. \tag{3.198}$$

By corollary 3.2, the sum (3.192) can be estimated at the cost of drawing $\mathrm{O}(\log(\delta^{-1})\epsilon^{-2}b_\sigma^2 b_A^2)$ samples from $R(m,n,k,k')$ and evaluating the corresponding $V(m,n,k,k')/R(m,n,k,k')$. Each of these samples can be computed in average time $\mathrm{O}(f_\sigma + f_A)$ as follows.

(i) Flip a weighted coin that lands heads with probability $1/p$.

(ii) If it lands heads, sample m,k according to $P_\sigma(m,k)$ and then sample n,k' according to $P_A(n,k'|m)$.

(iii) If it lands tails, sample n,k according to $Q_\sigma(n,k)$ and then sample m,k' according to $Q_A(m,k'|n)$.

(iv) The previous steps produce a sample according to $R(m,n,k,k')$ and can be accomplished in time $\mathrm{O}(f_\sigma + f_A)$ by conditions (c) and (d) of definition 3.6 and (c) and (d) of definition 3.7, with the side effect of producing values $\alpha^{(\sigma)}_{nmk}/P_\sigma(m,k)$, $\alpha^{(A)}_{mnk'}/P_A(n,k'|m)$, $\alpha^{(\sigma)}_{nmk}/Q_\sigma(n,k)$, and $\alpha^{(A)}_{mnk'}/Q_A(m,k'|n)$.

(v) These values can be used to compute $V(m,n,k,k')/R(m,n,k,k')$ since

$$\frac{V(m,n,k,k')}{R(m,n,k,k')} = \frac{\alpha^{(A)}_{mnk'}\alpha^{(\sigma)}_{nmk}}{R(m,n,k,k')} \tag{3.199}$$

$$= \left[\frac{1}{p}\frac{P_A(n,k'|m)}{\alpha^{(A)}_{mnk'}} \cdot \frac{P_\sigma(m,k)}{\alpha^{(\sigma)}_{nmk}} + \frac{1}{q}\frac{Q_A(m,k'|n)}{\alpha^{(A)}_{mnk'}} \cdot \frac{Q_\sigma(n,k)}{\alpha^{(\sigma)}_{nmk}}\right]^{-1} \tag{3.200}$$

Therefore, the sum (3.192) can be estimated in average time $\mathrm{O}[\log(\delta^{-1})\epsilon^{-2}b_\sigma^2 b_A^2(f_\sigma + f_A)]$. □

3.C Proofs for section 3.6

In section 3.6 several matrices and classes of matrices were claimed to be $\mathrm{EPS}_2(b,f)$ or $\mathrm{EPS}_p(b,f)$ for small values of b and f. In this appendix we provide proofs for these claims.

We first prove that the efficiently computable sparse (ECS) matrices from [VdN11] (definition reproduced below) are $\mathrm{EPS}_p(\mathrm{polylog}(N), \mathrm{polylog}(N))$. This covers a rather large class of matrices including permutation matrices, Pauli matrices, controlled phase matrices, and arbitrary unitaries on a constant number of qudits. The original definition from [VdN11] was in terms of qubits, but we adapt it to systems of arbitrary dimension.

Definition 3.32 (ECS)**.** *A matrix A is efficiently computable sparse (ECS) if*

(a) Each row and column of A has at most $\mathrm{polylog}(N)$ *nonzero entries.*

(b) For any given row index m, it is possible in $\mathrm{polylog}(N)$ *time to list the indices of the nonzero entries in that row,* $\{n : A_{mn} \neq 0\}$*, and to compute their values* A_{mn}*.*

(c) For any given column index n, it is possible in $\mathrm{polylog}(N)$ *time to list the indices of the nonzero entries in that column,* $\{m : A_{mn} \neq 0\}$*, and to compute their values* A_{mn}*.*

Theorem 3.33 (ECS is EPS)**.** *Let A be an ECS matrix satisfying* $\max_{mn}\{|A_{mn}|\} = \mathrm{polylog}(N)$*. Unitaries and Hermitian matrices whose eigenvalues are in the* $[-1, 1]$ *range satisfy this bound. Then A is* $\mathrm{EPS}_p(\mathrm{polylog}(N), \mathrm{polylog}(N))$ *for any* $p \in [1, \infty]$*.*

Proof. Theorem 3.12 is applicable here with $f = \mathrm{polylog}(N)$. Let $P(n|m)$ and $Q(m|n)$ be the probability distributions defined in (3.67). Given any m and n, the value A_{mn} can be computed in $\mathrm{polylog}(N)$ time. Since each row and column contains $\mathrm{polylog}(N)$ nonzero entries, which can be enumerated and computed in $\mathrm{polylog}(N)$ time, the sums $\sum_{n'} |A_{mn'}|$ and $\sum_{m'} |A_{m'n}|$ can be computed in $\mathrm{polylog}(N)$ time. Thus condition (c) of theorem 3.12 is satisfied.

For any given m, the distribution $P(n|m)$ has support of size $\mathrm{polylog}(N)$, the indices of which can be enumerated in $\mathrm{polylog}(N)$ time, and each individual probability can be computed in time $\mathrm{polylog}(N)$. Therefore, this distribution can be sampled from in time $\mathrm{polylog}(N)$. Similarly for $Q(m|n)$, so conditions (a) and (b) of theorem 3.12 are satisfied and A is $\mathrm{EPS}_p(\|A\|_\infty^{1/p}\|A\|_1^{1/q}, \mathrm{polylog}(N))$. Each row and column of A has at most $\mathrm{polylog}(N)$ nonzero entries, each bounded by $\max_{mn}\{|A_{mn}|\} = \mathrm{polylog}(N)$. It follows that $\|A\|_\infty = \mathrm{polylog}(N)$ and $\|A\|_1 = \mathrm{polylog}(N)$, giving $\|A\|_\infty^{1/p}\|A\|_1^{1/q} = \mathrm{polylog}(N)$. □

A block diagonal matrix is $\mathrm{EPS}_p(b, f)$ if each of its blocks is $\mathrm{EPS}_p(b, f)$. This is rather powerful in that it can be used to show the EPS property for operations on subsystems, for controlled-unitaries, and for some rather exotic projectors. This will be the subject of the following theorem and corollaries.

Theorem 3.34 (Block diagonal)**.** *For $r \in \{1, \ldots, R\}$, let $A^{(r)}$ be an* $\mathrm{EPS}_p(b_r, f)$ *matrix of dimension $M_r \times N_r$. Let A be the block diagonal matrix $A = \oplus_r A^{(r)}$ of dimension $\sum_r M_r \times \sum_r N_r$. Suppose that it is possible in time* $\mathrm{O}(f)$ *to convert between row/column indices of A and the corresponding block indices (i.e. $m' \to (r, m)$ and $n' \to (s, n)$ and their inverse maps, with $A_{m'n'} = \delta_{rs} A^{(r)}_{mn}$). Then A is* $\mathrm{EPS}_p(\max_r\{b_r\}, f)$*.*

Proof. Since $A^{(r)}$ is $\mathrm{EPS}_p(b_r, f)$ for each r, there are K_r, $\alpha^{(r)}_{mnk}$, $P_r(n, k|m)$, and $Q_r(m, k|n)$ satisfying definition 3.6, with $m \in \{1, \ldots, M_r\}$, $n \in \{1, \ldots, N_r\}$, and $k \in K_r$. Since we can convert between row/column indices of A and the corresponding block indices in time $\mathrm{O}(f)$, go ahead and label the indices of A using block indices: $A_{(r,m),(s,n)} = \delta_{rs} A^{(r)}_{mn}$. Define $K = \cup_r K_r$ and

$$\alpha_{(r,m),(s,n),k} = \begin{cases} \alpha^{(r)}_{mnk} & \text{if } r = s \text{ and } k \in K_r \\ 0 & \text{otherwise.} \end{cases} \tag{3.201}$$

This satisfies condition (a) of definition 3.6 since

$$\sum_{k \in K} \alpha_{(r,m),(s,n),k} = \delta_{rs} \sum_{k \in K_r} \alpha^{(r)}_{mnk} \tag{3.202}$$

$$= \delta_{rs} A^{(r)}_{mn} \tag{3.203}$$

$$= A_{(r,m),(s,n)}. \tag{3.204}$$

Define the probability distributions

$$P((s, n), k|(r, m)) = \delta_{rs} P_r(n, k|m), \tag{3.205}$$

$$Q((r, m), k|(s, n)) = \delta_{rs} Q_s(m, k|n). \tag{3.206}$$

That $\alpha_{(r,m),(s,n),k}$, $P((s,n),k|(r,m))$, and $Q((r,m),k|(s,n))$ satisfy conditions (c) and (d) of definition 3.6 directly follows from the fact that $\alpha^{(r)}_{mnk}$, $P_r(n,k|m)$, and $Q_s(m,k|n)$ satisfy conditions (c) and (d) for all r. Condition (b) is satisfied as well, since

$$\max_{(r,m),(s,n),k} \left\{ \frac{\left|\alpha_{(r,m),(s,n),k}\right|}{P((s,n),k|(r,m))^{1/p} Q((r,m),k|(s,n))^{1/q}} \right\} \tag{3.207}$$

$$= \max_r \max_{mnk} \left\{ \frac{\left|\alpha^{(r)}_{mnk}\right|}{P_r(n,k|m)^{1/p} Q_r(m,k|n)^{1/q}} \right\} \tag{3.208}$$

$$\leq \max_r \{b_r\}. \tag{3.209}$$

□

Corollary 3.35. *For $r \in \{1,\ldots,R\}$, let $A^{(r)}$ be matrices on a space of dimension N. Suppose that each $A^{(r)}$ is $\mathrm{EPS}_p(b,f)$ with $f = \Omega(\log^2(N))$. Then $A = \sum_{r=1}^{R} |r\rangle\langle r| \otimes A^{(r)}$, where the $|r\rangle$ are computational basis states, is $\mathrm{EPS}_p(b,f)$.*

Proof. This is essentially a restatement of theorem 3.34 for the case where all the $A^{(r)}$ are the same size. We require $f = \Omega(\log^2(N))$ because converting row or column indices of A to indices of the blocks (as required for application of theorem 3.34) requires the operation of computing the quotient and remainder of division by N. The $f = \Omega(\log^2(N))$ requirement can be dropped if one is dealing with query complexity rather than computational complexity. □

Corollary 3.36. *Let U denote a unitary matrix on n qubits whose rows are CT states (e.g. the Fourier transform). Let $g : \{0,\ldots,2^{n-1}\} \to \{0,\ldots,2^{n-1}\}$ be a $\mathrm{poly}(n)$ time computable function. Then the projector $\sum_{x=0}^{2^{n-1}} |x\rangle\langle x| \otimes U^\dagger |g(x)\rangle\langle g(x)| U$ is $\mathrm{EPS}_2(1,\mathrm{poly}(n))$. This projector corresponds to measuring half of the system in the computational basis to get measurement result x, measuring the other half of the system in the basis determined by U to get y, and returning true if $y = g(x)$. The measurement depicted in fig. 3.1 is of this form.*

Proof. Apply corollary 3.35 with $A^{(x)} = U^\dagger |g(x)\rangle\langle g(x)| U$. $U^\dagger |g(x)\rangle$ is a CT state, so by theorem 3.13 $A^{(x)}$ is $\mathrm{EHT}_2(1,\mathrm{poly}(n))$ and therefore also $\mathrm{EPS}_2(1,\mathrm{poly}(n))$. □

Corollary 3.37. *Let I_{M_1} and I_{M_2} denote the identity operator on spaces of dimension M_1 and M_2. Let A be an $\mathrm{EPS}_p(b,f)$ matrix of dimension $N_1 \times N_2$ with $f = \Omega(\log^2(M_1 M_2 N_1 N_2))$. Then $I_{M_1} \otimes A \otimes I_{M_2}$ is $\mathrm{EPS}_p(b,f)$. This somewhat trivial result is important in that it allows the matrix to act on subsystems of the full state.*

Proof. Apply theorem 3.34 with all of the $A^{(r)}$ blocks being equal. We require $f = \Omega(\log^2(M_1 M_2 N_1 N_2))$ in order to allow converting row or column indices of $I_{M_1} \otimes A \otimes I_{M_2}$ to indices of A in time $\mathrm{O}(f)$. □

We now turn to the Grover reflection operation. We will show this operator to be $\mathrm{EPS}_2(3,\log(N))$. Since a unitary operator incurs a time expense of b^4 as per (3.69), each round of Grover's algorithm multiplies the simulation time by $3^4 = 81$. This time is constant in the number of qubits, but is exponential in the number of rounds. Our technique is therefore perfectly capable of simulating a small number of Grover reflections placed anywhere in a circuit, but would perform very poorly, $\exp(\Theta(\sqrt{N}))$ time, if applied to the $\Theta(\sqrt{N})$ rounds required by Grover's algorithm.

Theorem 3.38. *Let $|+\rangle = N^{-1/2} \sum_{i=0}^{N-1} |i\rangle$. The Grover reflection $I - 2|+\rangle\langle+|$ is $\mathrm{EPS}_2(3,\log(N))$.*

Proof. Let δ_{mn} be the Kronecker delta. The identity operator can be seen to be $\mathrm{EPS}_p(1, \log(N))$, for any p but in particular $p = 2$, by simple inspection of definition 3.6 with $K = \{0\}$ and $\alpha_{mnk} = P(n, k|m) = Q(m, k|n) = \delta_{mn}$. Note that we must take $f = \log(N)$ rather than $f = 1$ since it takes $\Omega(\log(N))$ time to even write the indices m and n, which are $\log(N)$ bits long.

By Theorem 3.13, the projector $|+\rangle\langle+|$ is $\mathrm{EHT}_2(1, \log(N))$, and therefore also $\mathrm{EPS}_2(1, \log(N))$. By theorem 3.9 the operator $(-2)|+\rangle\langle+|$ is $\mathrm{EPS}_2(2, \log(N))$ and by theorem 3.10(a) the operator $I - 2|+\rangle\langle+|$ is $\mathrm{EPS}_2(3, \log(N))$. One cannot do much better than $b = 3$ since $\|I - 2(|+\rangle\langle+|)\|_2 \to 3$ as $N \to \infty$. □

Next we show that the Haar wavelet transform on n qubits, denoted G_n, is $\mathrm{EPS}_2(\sqrt{n+1}, n)$. This is the lowest possible value of b, since $\|\bar{G}_n\|_2 = \sqrt{n+1}$.

Definition 3.39. *The Haar wavelet transform on n qubits is defined to be*

$$G_n = (|0\rangle\langle+|)^{\otimes n} + \sum_{m=0}^{n-1} (|0\rangle\langle+|)^{\otimes m} \otimes |1\rangle\langle-| \otimes I^{\otimes n-m-1}. \tag{3.210}$$

Note that there are other conventions that differ from this by a permutation in the computational basis. Such permutations do not affect whether the Haar transform is $\mathrm{EPS}_2(\sqrt{n+1}, n)$.

As an example, the Haar transform on three qubits is implemented by the circuit depicted in fig. 3.4 and in the computational basis takes the form

$$G_3 = \begin{bmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} \\ \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} & 0 & \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} & 0 \\ 0 & \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} & 0 & \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} \\ \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} \end{bmatrix}. \tag{3.211}$$

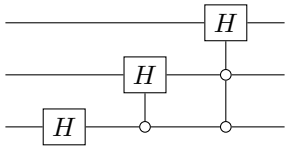


Figure 3.4: This circuit implements the Haar transform of definition 3.39, on three qubits [Hoy97]. The gates in this circuit are controlled-Hadamard gates, and the open circles denote that the Hadamard gates are active when all of the controls are in the $|0\rangle$ state.

Theorem 3.40. *The Haar transform on n qubits is* $\mathrm{EPS}_2(\sqrt{n+1}, n)$.

Proof. Since we are dealing with spaces of dimension 2^n, made of qubits, it will be convenient to index the space using bit strings $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$. We will denote the corresponding basis vectors using the notation $|\boldsymbol{x}\rangle = |x_0\rangle \otimes \cdots \otimes |x_{n-1}\rangle$. To avoid notational confusion regarding subscripts, define $A = G_n$. Then $A_{\boldsymbol{xy}}$ refers to the matrix element $\langle \boldsymbol{x}|G_n|\boldsymbol{y}\rangle$.

Take $K = \{0\}$ (i.e. don't make use of the index k), and set $\alpha_{\boldsymbol{xy}k} = A_{\boldsymbol{xy}}$. This satisfies condition (a) of definition 3.6 trivially. Take the probability distributions $P(\boldsymbol{y}|\boldsymbol{x})$ and $Q(\boldsymbol{x}|\boldsymbol{y})$ to be uniform over

the nonzero elements of the given row or column of $A_{\boldsymbol{xy}}$. Despite the apparent simplicity of this choice, analysis will be tedious due to the somewhat complicated definition of A. These probability distributions can be expressed as follows.

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{2^n}[\boldsymbol{x}=0] + \sum_{m=0}^{n-1} \frac{1}{2^{m+1}} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) [x_m = 1] \left(\prod_{i=m+1}^{n-1} [y_i = x_i] \right) \tag{3.212}$$

$$Q(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{n+1} \left\{ [\boldsymbol{x}=0] + \sum_{m=0}^{n-1} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) [x_m = 1] \left(\prod_{i=m+1}^{n-1} [x_i = y_i] \right) \right\} \tag{3.213}$$

These can be sampled from in time $\mathrm{O}(n)$. Consider first $P(\boldsymbol{y}|\boldsymbol{x})$. Given an $\boldsymbol{x}$, only a single one of the $n+1$ terms of (3.212) doesn't vanish, and this term can be identified in time $\mathrm{O}(n)$, by searching for the smallest (if any) m for which $x_m = 1$. The nonvanishing term defines the value of y_i for some of the i, and gives a uniform distribution for each of the remaining y_i. For $Q(\boldsymbol{x}|\boldsymbol{y})$, each of the $n+1$ terms of (3.213) is nonvanishing for a single value of $\boldsymbol{x}$, and each occurs with equal probability. Therefore, sampling from $Q(\boldsymbol{x}|\boldsymbol{y})$ is accomplished by drawing from a uniform distribution over $n+1$ possibilities.

To satisfy conditions (c) and (d) of definition 3.6 we must also show that $A_{\boldsymbol{xy}}/P(\boldsymbol{y}|\boldsymbol{x})$ and $A_{\boldsymbol{xy}}/Q(\boldsymbol{x}|\boldsymbol{y})$ can be computed in time $\mathrm{O}(n)$. We begin by writing an expression for $A_{\boldsymbol{xy}}$. In the equations below, square brackets denote the Iverson bracket, which takes a value of 1 if the enclosed expression is true and 0 otherwise.

$$A_{\boldsymbol{xy}} = \langle \boldsymbol{x} | \left((|0\rangle \langle +|)^{\otimes n} + \sum_{m=0}^{n-1} (|0\rangle \langle +|)^{\otimes m} \otimes |1\rangle \langle -| \otimes I^{\otimes n-m-1} \right) |\boldsymbol{y}\rangle \tag{3.214}$$

$$= \frac{1}{\sqrt{2^n}}[\boldsymbol{x}=0] + \sum_{m=0}^{n-1} \frac{(-1)^{y_m}}{\sqrt{2^{m+1}}} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) [x_m = 1] \left(\prod_{i=m+1}^{n-1} [x_i = y_i] \right), \tag{3.215}$$

Since only a single term for each of (3.212), (3.213), and (3.215) is nonvanishing for each given $\boldsymbol{x}, \boldsymbol{y}$ pair, we can divide these equations term-by-term to get

$$\frac{A_{\boldsymbol{xy}}}{P(\boldsymbol{y}|\boldsymbol{x})} = \sqrt{2^n}[\boldsymbol{x}=0] + \sum_{m=0}^{n-1} (-1)^{y_m} \sqrt{2^{m+1}} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) [x_m = 1] \left(\prod_{i=m+1}^{n-1} [x_i = y_i] \right), \tag{3.216}$$

$$\frac{A_{\boldsymbol{xy}}}{Q(\boldsymbol{x}|\boldsymbol{y})} = (n+1) \left\{ \frac{1}{\sqrt{2^n}}[\boldsymbol{x}=0] + \sum_{m=0}^{n-1} \frac{(-1)^{y_m}}{\sqrt{2^{m+1}}} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) [x_m = 1] \left(\prod_{i=m+1}^{n-1} [x_i = y_i] \right) \right\}. \tag{3.217}$$

At most a single term of these expressions is nonvanishing for each given $\boldsymbol{x}, \boldsymbol{y}$ pair, and this term can be identified in time $\mathrm{O}(n)$ by searching for the smallest (if any) m for which $x_m = 1$. The value of nonvanishing terms is of the form $\pm\sqrt{2^s}$ or $\pm(n+1)/\sqrt{2^s}$ for some s, and this can be computed in $\mathrm{O}(1)$ time.

That condition (b) of definition 3.6 is satisfied is checked directly,

$$\max_{\boldsymbol{xy}} \left\{ \frac{|A_{\boldsymbol{xy}}|}{P(\boldsymbol{y}|\boldsymbol{x})^{1/2} Q(\boldsymbol{x}|\boldsymbol{y})^{1/2}} \right\} = \max_{\boldsymbol{xy}} \left\{ \left(\frac{|A_{\boldsymbol{xy}}|}{P(\boldsymbol{y}|\boldsymbol{x})} \frac{|A_{\boldsymbol{xy}}|}{Q(\boldsymbol{x}|\boldsymbol{y})} \right)^{1/2} \right\} \tag{3.218}$$

$$= \max_{\boldsymbol{xy}} \left\{ (n+1)^{1/2} \right\} \tag{3.219}$$

$$= \sqrt{n+1}, \tag{3.220}$$

where (3.219) follows from the fact that only a single term from each of (3.216) and (3.217) is nonvanishing, so they can be multiplied term-by-term. □

Chapter 4

Bounds on Entanglement Assisted Source-channel Coding via the Lovász ϑ Number and its Variants[1]

[1] Preprint available: Toby Cubitt, Laura Mančinska, David Roberson, Simone Severini, Dan Stahlke, and Andreas Winter, *Bounds on Entanglement Assisted Source-channel Coding via the Lovász ϑ Number and its Variants*, arXiv:1310.7120 [quant-ph] (2013).

4.1 Abstract

We study zero-error entanglement assisted source-channel coding (communication in the presence of side information). Adapting a technique of Beigi, we show that such coding requires existence of a set of vectors satisfying orthogonality conditions related to suitably defined graphs G and H. Such vectors exist if and only if $\vartheta(\overline{G}) \leq \vartheta(\overline{H})$ where ϑ represents the Lovász number. We also obtain similar inequalities for the related Schrijver ϑ' and Szegedy ϑ^+ numbers.

These inequalities reproduce several known bounds and also lead to new results. We provide a lower bound on the entanglement assisted cost rate. We show that the entanglement assisted independence number is bounded by the Schrijver number: $\alpha^*(G) \leq \vartheta'(G)$. Therefore, we are able to disprove the conjecture that the one-shot entanglement-assisted zero-error capacity is equal to the integer part of the Lovász number. Beigi introduced a quantity β as an upper bound on α^* and posed the question of whether $\beta(G) = \lfloor \vartheta(G) \rfloor$. We answer this in the affirmative and show that a related quantity is equal to $\lceil \vartheta(G) \rceil$. We show that a quantity $\chi_{\text{vect}}(G)$ recently introduced in the context of Tsirelson's problem is equal to $\lceil \vartheta^+(\overline{G}) \rceil$.

In an appendix we investigate multiplicativity properties of Schrijver's and Szegedy's numbers, as well as projective rank.

4.2 Introduction

The source-channel coding problem is as follows: Alice and Bob can communicate only through a noisy channel. Alice wishes to send a message to Bob, and Bob already has some side information regarding Alice's message. (Note that Alice's message may be several bits long.) Alice encodes her message and sends a transmission through the channel. Given the (noisy) channel output along with his side information, Bob must be able to deduce Alice's message with zero probability of error (we always require zero error throughout this entire paper). An (m, n)-*coding scheme* consists of encoding and decoding operations which allow sending m messages via n uses of the noisy channel (again, each of the m messages may be several bits long). The *cost rate* η is the infimum of n/m over all (m, n)-coding schemes.

There are two special cases which are particularly noteworthy. If the messages are bits and there is no side information then the inverse of the cost rate, $1/\eta$, is the *Shannon capacity* [Sha56], the number of zero-error bits that can be transmitted per channel use in the limit of many uses of the channel. On the other hand, communication over a perfect channel with side information was considered by Witsenhausen [Wit76]; the corresponding cost rate is known as the *Witsenhausen rate*. The general problem, with both side information and a noisy channel, was considered by Nayak, Tuncel, and Rose [NTR06].

The Shannon capacity of a channel is very difficult to compute, and is not even known to be decidable. However, a useful upper bound on Shannon capacity is provided by the ϑ number introduced by Lovász [Lov79]. The Lovász ϑ number also provides a lower bound on the Witsenhausen rate [NTR06] and, in general, the cost rate.

Recently it has been of interest to study a version of this problem in which the parties may make use of an entangled quantum state, which can in certain cases increase the zero-error capacity of a classical channel [CLMW10, LMM+12]. The Lovász ϑ number upper bounds entanglement assisted Shannon capacity, just as it does classical Shannon capacity [Bei10, DSW13]. Beigi's proof [Bei10] proceeds through a relaxation of the channel coding problem, with the relaxed constraints consisting of various orthogonality conditions imposed upon a set of vectors. We study a relaxation of the entanglement assisted source-channel coding problem inspired by this technique of Beigi. This relaxation leads to a set of constraints that are exactly characterized by monotonicity of ϑ. This has a number of consequences. Beigi defined a function β as an upper bound on entanglement assisted independence number and posed the question of whether β is equal to $\lfloor \vartheta \rfloor$. We answer this in the affirmative and show that a similarly defined quantity is equal to $\lceil \vartheta \rceil$. We show that ϑ provides a

bound for the source-channel coding problem. As a special case this reproduces both Beigi's result as well as that of Briët et al. [BBL+13] in which it is shown that ϑ is a lower bound on the entanglement assisted Witsenhausen rate.

A slightly different relaxation of source-channel coding leads to three necessary conditions for the existence of a $(1,1)$-coding scheme in terms of ϑ and two variants: Schrijver's ϑ' and Szegedy's ϑ^+. This reproduces or strengthens results from [Bei10, BBL+13, RM12] under a unified framework, with simpler proofs. In particular, we produce a tighter bound on the entanglement assisted independence number: $\alpha^* \leq \vartheta'$.

The technical results, theorems 4.6 and 4.10, should be accessible to the reader who is familiar with graph theory but not information theory or quantum mechanics, which merely provide a motivation for the problem.

4.3 Source-channel coding

We will make use of the following graph theoretical concepts. A *graph* G consists of a set of *vertices* $V(G)$ along with a symmetric binary relation $x \sim_G y$ among vertices (we abbreviate $x \sim y$ when the graph can be inferred from context). A pair of vertices (x, y) satisfying $x \sim y$ are said to be *adjacent*. Equivalently, it is said that there is an *edge* between x and y. Vertices are not adjacent to themselves, so $x \not\sim x$ for all $x \in V(G)$. The *complement* of a graph G, denoted $\overline{G}$, has the same set of vertices but has edges between distinct pairs of vertices which are not adjacent in G (i.e. for $x \neq y$ we have $x \sim_{\overline{G}} y \iff x \not\sim_G y$). A set of vertices no two of which form an edge is known as an *independent set*; the size of the largest independent set is the *independence number* $\alpha(G)$. A set of vertices such that every pair is adjacent is known as a *clique*; the size of the largest clique is the *clique number* $\omega(G)$. Clearly $\omega(G) = \alpha(\overline{G})$. An assignment of colors to vertices such that adjacent vertices are given distinct colors is called a *proper coloring*; the minimum number of colors needed is the *chromatic number* $\chi(G)$. A function mapping the vertices of one graph to those of another, $f : V(G) \to V(H)$, is a *homomorphism* if $x \sim_G y \implies f(x) \sim_H f(y)$. Since vertices are not adjacent to themselves it is necessary that $f(x) \neq f(y)$ when $x \sim y$. If such a function exists, we say that G *is homomorphic to* H and write $G \to H$. The *complete graph* on n vertices, denoted K_n, has an edge between every pair of vertices. It is not hard to see that $\omega(G)$ is equal to the largest n such that $K_n \to G$, and $\chi(G)$ is equal to the smallest n such that $G \to K_n$. Many other graph properties can be expressed in terms of homomorphisms; for details see [HT97, HN04]. The *strong product* of two graphs, $G \boxtimes H$, has vertex set $V(G) \times V(H)$ and has edges

$$\begin{aligned}(x_1, y_1) \sim (x_2, y_2) \iff &(x_1 = x_2 \text{ and } y_1 \sim y_2) \text{ or} \\ &(x_1 \sim x_2 \text{ and } y_1 = y_2) \text{ or} \\ &(x_1 \sim x_2 \text{ and } y_1 \sim y_2).\end{aligned}$$

The n-fold strong product is written $G^{\boxtimes n} := G \boxtimes G \boxtimes \cdots \boxtimes G$. The *disjunctive product* $G * H$ has edges

$$(x_1, y_1) \sim (x_2, y_2) \iff x_1 \sim x_2 \text{ or } y_1 \sim y_2.$$

It is easy to see that $\overline{G * H} = \overline{G} \boxtimes \overline{H}$. The n-fold disjunctive product is written $G^{*n} := G * G * \ldots * G$.

Suppose that Alice communicates to Bob through a noisy classical channel $\mathcal{N} : S \to V$. She wishes to send a message to Bob with zero chance of error. Let $\mathcal{N}(v|s)$ denote the probability that $\mathcal{N}$ will produce symbol v when given symbol s as input, and define the graph H with vertices S and edges

$$s \sim_H t \iff \mathcal{N}(v|s)\mathcal{N}(v|t) = 0 \text{ for all } v \in V. \tag{4.1}$$

Bob can distinguish codewords s and t if and only if they have no chance of being mapped to the same output by $\mathcal{N}$. Therefore, Alice's set of codewords must form a clique of H; the size of the largest such set is the clique number $\omega(H)$. We will call this the *distinguishability graph* of $\mathcal{N}$. The complementary graph $\overline{H}$ is known as the *confusability graph* of $\mathcal{N}$. Note that standard convention is to denote the confusability graph by H rather than $\overline{H}$. We break convention in order to make notation in this paper much simpler. However, to minimize confusion when discussing prior results, we will follow the tradition of using the independence number when speaking of the number of codewords that Alice can send (equal to $\alpha(\overline{H}) = \omega(H)$ in our notation).

The number of bits (the base-2 log of the number of distinct codewords) that Alice can send with a single use of $\mathcal{N}$ is known as the *one-shot zero-error capacity* of $\mathcal{N}$, and is equal to $\log \alpha(\overline{H})$. The average number of bits that can be sent per channel use (again with zero error) in the limit of many uses of a channel is known as the *Shannon capacity*. With n parallel uses of $\mathcal{N}$ the distinguishability graph is H^{*n}. The Shannon capacity of $\mathcal{N}$ is therefore

$$\begin{aligned}\Theta(\overline{H}) :&= \lim_{n\to\infty} \frac{1}{n} \log \omega(H^{*n}) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \alpha(\overline{H}^{\boxtimes n}).\end{aligned}$$

This quantity is in general very difficult to compute, with the capacity of the five cycle graph $\overline{H} = C_5$ having been open for over 20 years and the capacity of C_7 being unknown to this day. The capacity of C_5 was solved by Lovász [Lov79] who introduced a function $\vartheta(\overline{H})$, the definition of which will be postponed until section 4.4. Lovász proved a sandwich theorem which, using the notation $\bar{\vartheta}(H) := \vartheta(\overline{H})$, takes the form

$$\alpha(\overline{H}) = \omega(H) \leq \bar{\vartheta}(H) \leq \chi(H).$$

He also showed that $\bar{\vartheta}(H^{*n}) = \bar{\vartheta}(H)^n$, therefore $\Theta(\overline{H}) \leq \log \bar{\vartheta}(H)$. This bound also applies to entanglement assisted communication [Bei10], which we will investigate in detail, and has been generalized to quantum channels [DSW13].

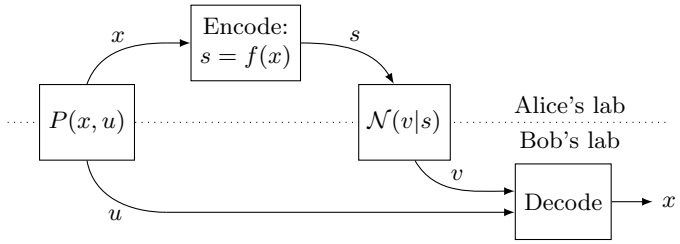


Figure 4.1: A zero-error source-channel $(1, 1)$-coding scheme.

We now introduce the *source-channel coding* problem. As before, Alice wishes to send Bob a message $x \in X$, and she can only communicate through a noisy channel $\mathcal{N} : S \to V$. Now, however, Bob has some side information about Alice's message. Specifically, Alice and Bob each receive one part of a pair (x, u) drawn according to a probability distribution $P(x, u)$. This is known as a *dual source*. Alice encodes her input x using a function $f : X \to S$ and sends the result through $\mathcal{N}$. Bob must deduce x with zero chance of error using the output of $\mathcal{N}$ along with his side information u. Such a protocol is called a *zero-error source-channel* $(1, 1)$*-coding scheme*, and is depicted in fig. 4.1. An (m, n)-coding scheme transmits m independent instances of the source using n copies of the channel.

Again the analysis involves graphs. Let H again be the distinguishability graph (4.1) and define the *characteristic graph* G with vertices X and edges

$$x \sim_G y \iff \exists u \in U \text{ such that } P(x, u)P(y, u) \neq 0.$$

In [NTR06] it was shown that decoding is possible if and only if Alice's encoding f is a homomorphism from G to H.[2] Therefore a zero-error $(1,1)$-coding scheme exists if and only if $G \to H$. A zero-error (m,n)-coding scheme is possible if and only if

$$G^{\boxtimes m} \to H^{*n}. \tag{4.2}$$

The smallest possible ratio n/m (in the limit $m \to \infty$) is called the *cost rate*, $\eta(G, \overline{H})$. More precisely, the cost rate is defined as

$$\eta(G, \overline{H}) = \lim_{m\to\infty} \frac{1}{m} \min\left\{n : G^{\boxtimes m} \to H^{*n}\right\}. \tag{4.3}$$

The $\bar{\vartheta}$ quantity is monotone under graph homomorphisms in the sense that $G \to H \implies \bar{\vartheta}(G) \leq \bar{\vartheta}(H)$ [dCST13]. Consequently, a zero-error $(1,1)$-coding scheme requires $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$. Since $\bar{\vartheta}(G^{\boxtimes m}) = \bar{\vartheta}(G)^m$ [Knu94] and $\bar{\vartheta}(H^{*n}) = \bar{\vartheta}(H)^n$ [Lov79], it follows that an (m,n)-coding scheme is possible only if $\log \bar{\vartheta}(G)/\log \bar{\vartheta}(H) \leq n/m$. Thus we have the bound

$$\eta(G, \overline{H}) \geq \frac{\log \bar{\vartheta}(G)}{\log \bar{\vartheta}(H)}.$$

(Cf. [NTR06] for the special case of the Witsenhausen rate.)

We will return to this in section 4.4 when we prove an analogous bound for entanglement assisted zero-error source-channel coding.

When Bob has no side information (equivalently, when U is a singleton), G is the complete graph. In this case zero-error transmission of x is possible if and only if $K_n \to H$ where $n = |X|$, which in turn holds if and only if $n \leq \omega(H) = \alpha(\overline{H})$. This is the expected result, since as mentioned before $\alpha(\overline{H})$ is the number of unambiguously decodable codewords that Alice can send through $\mathcal{N}$. On the other hand, consider the case where there is side information and $\mathcal{N}$ is a noiseless channel of size $n = |S|$. Now H becomes the complete graph K_n, so x can be perfectly transmitted if and only if $G \to K_n$. This holds if and only if $n \geq \chi(G)$. These two examples provide an operational interpretation to the independence number and chromatic number of a graph. The analogous communication problems in the presence of an entangled state (which we examine shortly) define the entanglement assisted independence and chromatic numbers.

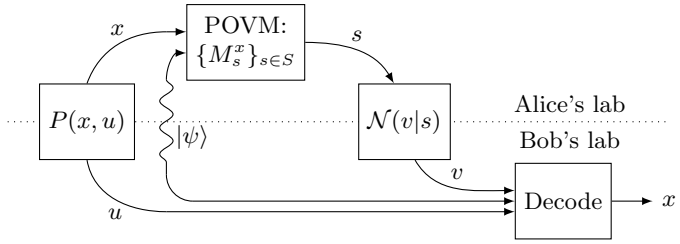


Figure 4.2: An entanglement assisted zero-error source-channel $(1,1)$-coding scheme.

If Alice and Bob share an entangled state they can use the strategy depicted in fig. 4.2, which is described in greater detail in [BBL+13]. Alice, upon receiving $x \in X$, performs a POVM $\{M_s^x\}_{s\in S}$ on her half of the entanglement resource $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and receives measurement outcome $s \in S$. Without loss of generality this can be assumed to be a projective measurement since any POVM can be converted to a projective measurement by enlarging the entangled state. So for each $x \in X$, the collection $\{M_s^x\}_{s\in S}$ consists of projectors on $\mathcal{H}_A$ which sum to the identity. Alice sends

[2] Basically, G represents the information that needs to be sent and H represents the information that survives the channel. A homomorphism $G \to H$ ensures that the needed information makes it through the channel intact.

the measurement outcome s through the channel $\mathcal{N}$ to Bob, who receives some $v \in V$ such that $\mathcal{N}(v|s) > 0$. Bob then measures his half of the entangled state using a projective measurement depending on v and his side information u. An *entanglement assisted zero-error* $(1,1)$*-coding scheme* is one in which Bob is able to determine x with zero chance of error; an *entanglement assisted zero-error* (m,n)*-coding scheme* involves sending m independent samples of the source using n copies of the channel.

After Alice's measurement, Bob's half of the entanglement resource is in the state

$$\rho_s^x = \mathrm{Tr}_A\{(M_s^x \otimes I)|\psi\rangle\langle\psi|\}.$$

An error free decoding operation exists for Bob if and only if these states are orthogonal for every $x \in X$ consistent with the information in Bob's possession (i.e. u and v). We then have the following necessary and sufficient condition [BBL+13]. Let G be the characteristic graph of the source and H be the distinguishability graph of the channel. There must be a bipartite pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ for some Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, and for each $x \in X$ there must be a projective decomposition of the identity $\{M_s^x\}_{s\in S}$ on $\mathcal{H}_A$ such that

$$\rho_s^x \perp \rho_t^y \text{ for all } x \sim_G y \text{ and } s \not\sim_H t,$$

where orthogonality is in terms of the Hilbert–Schmidt inner product.

Recall that without entanglement a zero-error $(1,1)$-coding scheme was possible if and only if $G \to H$. By analogy we say there is an *entanglement assisted homomorphism* $G \xrightarrow{*} H$ when there exists an entanglement assisted zero-error $(1,1)$-coding scheme:

Definition 4.1. *Let G and H be graphs. There is an entanglement assisted homomorphism from G to H, written $G \xrightarrow{*} H$, if there is a bipartite state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ (for some Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$) and, for each $x \in V(G)$, a projective decomposition of the identity $\{M_s^x\}_{s\in V(H)}$ on $\mathcal{H}_A$ such that*

$$\rho_s^x \perp \rho_t^y \text{ for all } x \sim_G y \text{ and } s \not\sim_H t, \tag{4.4}$$

where

$$\rho_s^x := Tr_A\{(M_s^x \otimes I)|\psi\rangle\langle\psi|\}. \tag{4.5}$$

Analogous to (4.2), there is an entanglement assisted (m,n)-coding scheme if and only if $G^{\boxtimes m} \xrightarrow{*} H^{*n}$. The entangled cost rate [BBL+13] is analogous to (4.3),

$$\eta^*(G,\overline{H}) = \lim_{m\to\infty} \frac{1}{m} \min\left\{n : G^{\boxtimes m} \xrightarrow{*} H^{*n}\right\}. \tag{4.6}$$

In the absence of side information (i.e. with U being a singleton set), G becomes the complete graph. We saw above that without entanglement and without side information, n distinct codewords can be sent error-free through a noisy channel if and only if $K_n \to H$; the largest such n is $\omega(H) = \alpha(\overline{H})$. With the help of entanglement the largest number of codewords is the largest n such that $K_n \xrightarrow{*} H$; this defines the *entanglement assisted independence number*, $\alpha^*(\overline{H})$. Since an entanglement resource never hurts, $\alpha^*(\overline{H}) \geq \alpha(\overline{H})$ always. In some cases $\alpha^*(\overline{H})$ can be strictly larger than $\alpha(\overline{H})$ [CLMW10].

We saw above that $\alpha(\overline{H}) \leq \bar{\vartheta}(H)$. Indeed, this was the original application of $\bar{\vartheta}$. Beigi showed that also $\alpha^*(\overline{H}) \leq \bar{\vartheta}(H)$ [Bei10] (this has been generalized to quantum channels as well [DSW13]; however, we consider here only classical channels). Beigi proved his bound by showing that if n distinct codewords can be sent through a noisy channel with zero-error using entanglement ($K_n \xrightarrow{*} H$ in our notation) then there are vectors $|w\rangle \neq 0$ and $|w_s^x\rangle$ with $x \in \{1,\ldots,n\}$ and $s \in V(H)$ such

that[3]

$$\sum_s |w_s^x\rangle = |w\rangle \tag{4.7}$$

$$\langle w_s^x | w_t^y \rangle = 0 \text{ for all } x \neq y, s \not\sim_H t \tag{4.8}$$

$$\langle w_s^x | w_t^x \rangle = 0 \text{ for all } s \neq t. \tag{4.9}$$

Denote by $\beta(\overline{H})$ the largest n such that vectors of this form exist. Then $\beta(\overline{H}) \geq \alpha^*(\overline{H})$. Beigi showed that the existence of such vectors implies $n \leq \bar{\vartheta}(H)$, therefore $\alpha^*(\overline{H}) \leq \beta(\overline{H}) \leq \bar{\vartheta}(H)$. Since ϑ is multiplicative under the strong graph product, $\bar{\vartheta}(H)$ is in fact an upper bound on the entanglement assisted Shannon capacity. Beigi left open the question of whether $\beta(\overline{H})$ was equal to $\lfloor \bar{\vartheta}(H) \rfloor$. We will answer this question in the affirmative (corollary 4.7).

In fact, we show something more general. We generalize Beigi's vectors so that they apply to the source-channel coding problem (i.e. with G not necessarily being K_n) and give a bound in terms of the Lovász ϑ number. The conditions we will introduce can be thought of as a relaxation of the condition (4.4), which defines $G \overset{*}{\to} H$. A related but different relaxation will give bounds in terms of two variations of the Lovász number: the Schrijver number [Sch79, MRRJ78] and the Szegedy number [Sze94]. We denote the first relaxation $G \overset{B}{\to} H$ since it generalizes Beigi's condition, and denote the second $G \overset{+}{\to} H$ since it contains a positivity condition. A third relaxation, $G \overset{V}{\to} H$, is defined here but the significance is discussed later.

Definition 4.2. *Let G and H be graphs. Write $G \overset{B}{\to} H$ if there are vectors $|w\rangle \neq 0$ and $|w_s^x\rangle \in \mathbb{C}^d$ for each $x \in V(G)$, $s \in V(H)$, for some $d \in \mathbb{N}$, such that*

1. $\sum_s |w_s^x\rangle = |w\rangle$
2. $\langle w_s^x | w_t^y \rangle = 0$ *for all* $x \sim_G y$, $s \not\sim_H t$
3. $\langle w_s^x | w_t^x \rangle = 0$ *for all* $s \neq t$.

Write $G \overset{+}{\to} H$ if there are vectors satisfying

1. $\sum_s |w_s^x\rangle = |w\rangle$
2. $\langle w_s^x | w_t^y \rangle = 0$ *for all* $x \sim_G y$, $s \not\sim_H t$
3. $\langle w_s^x | w_t^y \rangle \geq 0$.

Write $G \overset{V}{\to} H$ if there are vectors satisfying the conditions for both $G \overset{B}{\to} H$ and $G \overset{+}{\to} H$, i.e.,

1. $\sum_s |w_s^x\rangle = |w\rangle$
2. $\langle w_s^x | w_t^y \rangle = 0$ *for all* $x \sim_G y$, $s \not\sim_H t$
3. $\langle w_s^x | w_t^x \rangle = 0$ *for all* $s \neq t$
4. $\langle w_s^x | w_t^y \rangle \geq 0$.

Without loss of generality one could consider only real vectors, since complex vectors can be turned real via the recipe $|\hat{w}_s^x\rangle = \mathrm{Re}(|w_s^x\rangle) \oplus \mathrm{Im}(|w_s^x\rangle)$ while preserving the inner product properties required by the above definitions.

It is enlightening to consider the Gram matrices of the $|w_s^x\rangle$ vectors. In fact, it is this formulation that will be used to prove our main theorems.

[3] Recall that we take $\overline{H}$ to be the confusability graph rather than H. So Beigi's definition is worded differently.

Theorem 4.3. *$G \overset{B}{\to} H$ if and only if there is a positive semidefinite matrix $C : \mathcal{L}(\mathbb{C}^{|V(G)|}) \otimes \mathcal{L}(\mathbb{C}^{|V(H)|})$ satisfying*

$$\sum_{s,t} C_{xyst} = 1 \tag{4.10}$$

$$C_{xyst} = 0 \text{ for } x \sim_G y \text{ and } s \not\sim_H t \tag{4.11}$$

$$C_{xxst} = 0 \text{ for } s \neq t. \tag{4.12}$$

$G \overset{+}{\to} H$ if and only if there is a positive semidefinite matrix satisfying (4.10), (4.11), and

$$C_{xyst} \geq 0. \tag{4.13}$$

$G \overset{V}{\to} H$ if and only if there is a positive semidefinite matrix satisfying (4.10)-(4.13).

Proof. We prove only the $G \overset{B}{\to} H$ claim; the proofs for $G \overset{+}{\to} H$ and $G \overset{V}{\to} H$ are analogous. Suppose $G \overset{B}{\to} H$ and let $|w\rangle$ and $|w_s^x\rangle$ be the vectors described in definition 4.2. Without loss of generality rescale so that $\langle w|w\rangle = 1$. Define the matrix $C : \mathcal{L}(\mathbb{C}^{|V(G)|}) \otimes \mathcal{L}(\mathbb{C}^{|V(H)|})$ with entries $C_{xyst} = \langle w_s^x | w_t^y \rangle$ for $x, y \in V(G)$ and $s, t \in V(H)$. Since C is a Gram matrix, it is positive semidefinite. Properties (4.10)-(4.12) follow directly from the three properties listed in definition 4.2 for $G \overset{B}{\to} H$ (the first of these uses $\langle w|w\rangle = 1$).

For the converse, note that any positive semidefinite matrix $C : \mathcal{L}(\mathbb{C}^{|V(G)|}) \otimes \mathcal{L}(\mathbb{C}^{|V(H)|})$ is a Gram matrix of some vectors $|w_s^x\rangle$. The three properties of definition 4.2 for $G \overset{B}{\to} H$ follow from (4.10)-(4.12). Only the first of these is nontrivial. We have that for all $x, y \in V(G)$,

$$1 = \sum_{st} C_{xyst} = \sum_{st} \langle w_s^x | w_t^y \rangle = \left(\sum_s \langle w_s^x | \right) \left(\sum_t |w_t^y\rangle \right).$$

For $x = y$, the above implies that $\sum_s |w_s^x\rangle$ is a unit vector for all x. These unit vectors must have unit inner product amongst themselves by the $x \neq y$ cases above, and therefore they must all be the same vector. Call this $|w\rangle$. Clearly $|w\rangle \neq 0$. □

It is interesting to note that a matrix with properties (4.10), (4.11), and (4.13) (those associated with $G \overset{+}{\to} H$) can be interpreted as a conditional probability distribution, $P(s,t|x,y) = C_{xyst}$. With this interpretation, $G \overset{+}{\to} H$ if and only if there exists a conditional probability distribution $P(s,t|x,y)$ such that $C_{sx;ty} = P(s,t|x,y)$ is a positive semidefinite matrix and $P(s \sim_H t | x \sim_G y) = 1$. A similar interpretation holds for $G \overset{V}{\to} H$.

We now show that $G \overset{B}{\to} H$ and $G \overset{+}{\to} H$ are indeed relaxations of $G \overset{*}{\to} H$ (the significance of $G \overset{V}{\to} H$ will be explained later). Since definition 4.2 reduces to Beigi's criteria when considering $K_n \overset{B}{\to} H$, the argument that follows provides an alternative and simpler proof of Beigi's result that $\alpha^*(\overline{H}) \leq \beta(\overline{H})$.

Theorem 4.4. *If $G \overset{*}{\to} H$ then $G \overset{B}{\to} H$ and $G \overset{+}{\to} H$.*

Proof. ($G \overset{*}{\to} H \implies G \overset{B}{\to} H$): Suppose that $G \overset{*}{\to} H$. Let $|\psi\rangle$ and M_s^x for $x \in V(G)$ and $s \in V(H)$ satisfy condition (4.4) (with ρ_s^x given by (4.5)). Define $|w\rangle = |\psi\rangle$ and

$$|w_s^x\rangle = (M_s^x \otimes I) |\psi\rangle .$$

Since $\{M_s^x\}_{s\in S}$ is a projective decomposition of the identity,

$$\sum_s M_s^x = I \implies \sum_s |w_s^x\rangle = |w\rangle\,,$$
$$M_s^x M_t^x = 0 \implies \langle w_s^x | w_t^x\rangle = 0 \text{ for } s \neq t.$$

For all $x \sim_G y$ and $s \not\sim_H t$, condition (4.4) gives that the reduced density operators (tracing over $\mathcal{H}_A$) of the post-measurement states $(M_s^x \otimes I)\,|\psi\rangle$ and $(M_t^y \otimes I)\,|\psi\rangle$ are orthogonal. But this is only possible if the pure states (without tracing out $\mathcal{H}_A$) are orthogonal. So,

$$\big\langle\psi\big|(M_s^x \otimes I)^\dagger (M_t^y \otimes I)\big|\psi\big\rangle = 0 \implies \langle w_s^x | w_t^y\rangle = 0.$$

($G \overset{*}{\to} H \implies G \overset{+}{\to} H$): Suppose that $G \overset{*}{\to} H$. Let $|\psi\rangle$ and M_s^x for $x \in V(G)$ and $s \in V(H)$ satisfy condition (4.4) (with ρ_s^x given by (4.5)). Define $|w_s^x\rangle$ to be the vectorization of the post-measurement reduced density operator,

$$|w_s^x\rangle = \mathrm{vec}(\rho_s^x).$$

Since $\{M_s^x\}_{s\in S}$ sum to identity,

$$\begin{aligned}\sum_s |w_s^x\rangle &= \mathrm{vec}\left(\sum_s \mathrm{Tr}_A\{(M_s^x \otimes I)|\psi\rangle\langle\psi|\}\right)\\ &= \mathrm{vec}(\mathrm{Tr}_A\{|\psi\rangle\langle\psi|\}) =: |w\rangle\,.\end{aligned}$$

For all $x \sim_G y$ and $s \not\sim_H t$, condition (4.4) gives $\langle w_s^x | w_t^y\rangle = 0$. Density operators are positive, giving positive inner products $\langle w_s^x | w_t^y\rangle \geq 0$. □

4.4 Monotonicity theorems

Our main results concern monotonicity properties of the Lovász number ϑ, Schrijver number ϑ', and Szegedy number ϑ^+ for graphs that are related by the generalized homomorphisms of definition 4.2. These will lead to various bounds relevant to entanglement assisted zero-error source-channel coding. These three quantities are defined as follows.

Definition 4.5. *In this definition we use real matrices. For convenience we state the definitions in terms of the complement of a graph, since this form is used throughout the theorems.*

The Lovász number of the complement, $\bar\vartheta(G) := \vartheta(\overline{G})$, is given by either of the following two semidefinite programs, which are equivalent [Lov79, Knu94, Lov03]:

$$\begin{aligned}\bar\vartheta(G) = \max\{&\|I + T\| : I + T \succeq 0,\\ &T_{ij} = 0 \text{ for } i \not\sim j\},\end{aligned} \tag{4.14}$$

$$\begin{aligned}\bar\vartheta(G) = \min\{&\lambda : \exists Z \succeq 0, Z_{ii} = \lambda - 1,\\ &Z_{ij} = -1 \text{ for } i \sim j\},\end{aligned} \tag{4.15}$$

where $\|\cdot\|$ denotes the operator norm (the largest singular value) and $\succeq 0$ means that a matrix is positive semidefinite. The Schrijver number of the complement, $\bar\vartheta'(G) := \vartheta'(\overline{G})$, (sometimes written ϑ^-) is [Sch79, MRRJ78]

$$\begin{aligned}\bar\vartheta'(G) = \min\{&\lambda : \exists Z \succeq 0, Z_{ii} = \lambda - 1,\\ &Z_{ij} \leq -1 \text{ for } i \sim j\}.\end{aligned} \tag{4.16}$$

The Szegedy number of the complement, $\bar{\vartheta}^+(G) := \vartheta^+(\overline{G})$, is [Sze94]

$$\begin{aligned}\bar{\vartheta}^+(G) = \min\{\lambda : \exists Z \succeq 0, Z_{ii} = \lambda - 1,\\ Z_{ij} = -1 \text{ for } i \sim j,\\ Z_{ij} \geq -1 \text{ for all } i, j\}.\end{aligned} \tag{4.17}$$

Clearly $\bar{\vartheta}'(G) \leq \bar{\vartheta}(G) \leq \bar{\vartheta}^+(G)$.

Our first result is that $G \overset{B}{\to} H$ exactly characterizes ordering of $\bar{\vartheta}$. This will lead to a bound on entanglement assisted cost rate.

Theorem 4.6. $G \overset{B}{\to} H \iff \bar{\vartheta}(G) \leq \bar{\vartheta}(H)$.

Proof. ($\Longleftarrow$): Suppose $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$. We will explicitly construct a matrix $C \succeq 0$ satisfying properties (4.10)-(4.12) of theorem 4.3. Let $\lambda = \bar{\vartheta}(H)$. By definition, there is a matrix T such that $\|I + T\| = \lambda$, $I + T \succeq 0$, and $T_{st} = 0$ for $s \not\sim t$. With $|\psi\rangle$ denoting the vector corresponding to the largest eigenvalue of $I + T$, and with $\circ$ denoting the Schur–Hadamard (i.e. entrywise) product, define the matrices

$$\begin{aligned}D &= |\psi\rangle\langle\psi| \circ I,\\ B &= |\psi\rangle\langle\psi| \circ (I + T).\end{aligned}$$

With J being the all-ones matrix and $\langle\cdot,\cdot\rangle$ denoting the Hilbert–Schmidt inner product, it is readily verified that

$$\begin{aligned}\langle D, J\rangle &= \langle\psi|\psi\rangle = 1,\\ \langle B, J\rangle &= \langle\psi|I + T|\psi\rangle = \lambda.\end{aligned}$$

Since the Schur–Hadamard product of two matrices is a principal submatrix of their tensor product, this operation preserves positive semidefiniteness. As a consequence, $B \succeq 0$ and

$$\|I + T\| = \lambda \implies \lambda I - (I + T) \succeq 0 \implies \lambda D - B \succeq 0.$$

Since $\lambda \geq \bar{\vartheta}(G)$, there is a matrix Z such that $Z \succeq 0$, $Z_{xx} = \lambda - 1$ for all x, and $Z_{xy} = -1$ for all $x \sim_G y$. Note that definition 4.5 gives existence of a matrix with $\bar{\vartheta}(G) - 1$ on the diagonal, but since $\lambda \geq \bar{\vartheta}(G)$ we can add a multiple of the identity to get $\lambda - 1$ on the diagonal.

We now construct C. Define

$$C = \lambda^{-1}\left[J \otimes B + (\lambda - 1)^{-1} Z \otimes (\lambda D - B)\right].$$

Since J, B, Z, and $\lambda D - B$ are all positive semidefinite, and $\lambda - 1 \geq 0$, we have that C is positive semidefinite. The other desired conditions on C are easy to verify. For all x, y we have

$$\begin{aligned}\sum_{st} C_{xyst} &= \lambda^{-1}\left[\langle B, J\rangle + (\lambda - 1)^{-1} Z_{xy}[\lambda\langle D, J\rangle - \langle B, J\rangle]\right]\\ &= 1.\end{aligned}$$

Note that the J in the above equation is indexed by $V(H)$, whereas the J in the definition of C is indexed by $V(G)$. For $x \sim_G y$ and $s \not\sim_H t$,

$$\begin{aligned}C_{xyst} &= \lambda^{-1}\left[B_{st} + (\lambda - 1)^{-1} Z_{xy}(\lambda D_{st} - B_{st})\right]\\ &= \lambda^{-1}\left[B_{st} + (\lambda - 1)^{-1}(-1)(\lambda B_{st} - B_{st})\right] = 0.\end{aligned}$$

For all x and for $s \neq t$,

$$\begin{aligned} C_{xxst} &= \lambda^{-1}\left[B_{st} + (\lambda-1)^{-1} Z_{xx}(\lambda D_{st} - B_{st})\right] \\ &= \lambda^{-1}\left[B_{st} + (0 - B_{st})\right] = 0. \end{aligned}$$

($\Longrightarrow$): Suppose $G \overset{B}{\to} H$. By theorem 4.3, there is a matrix $C \succeq 0$ satisfying properties (4.10)-(4.12). Let Z achieve the optimal value (call it λ) for the minimization (4.15) for $\bar{\vartheta}(H)$. We will provide a feasible solution for (4.15) for $\bar{\vartheta}(G)$ to show that $\bar{\vartheta}(G) \leq \lambda = \bar{\vartheta}(H)$. To this end, let $|\mathbf{1}\rangle$ be the all ones vector and define

$$Y = \big(I \otimes \langle \mathbf{1}|\,\big)\big[\big(J \otimes Z\big) \circ C\big]\big(I \otimes |\mathbf{1}\rangle\,\big).$$

Since $C \succeq 0$ and $Z \succeq 0$, and positive semidefiniteness is preserved by conjugation, we have that $Y \succeq 0$. Also note that

$$Y_{xy} = \sum_{st} Z_{st} C_{xyst}.$$

Using the fact that $Z_{ss} = \lambda - 1$ and $C_{xxst} = 0$ for $s \neq t$, we have

$$Y_{xx} = \sum_{st} Z_{st} C_{xxst} = (\lambda - 1) \sum_{st} C_{xxst} = \lambda - 1.$$

Using the fact that $Z_{st} = -1$ for $s \sim_H t$ and $C_{xyst} = 0$ for $x \sim_G y$, $s \not\sim_H t$, we have that for $x \sim_G y$,

$$\begin{aligned} Y_{xy} &= \sum_{st} Z_{st} C_{xyst} = \sum_{s \sim_H t} Z_{st} C_{xyst} = (-1) \sum_{s \sim_H t} C_{xyst} \\ &= (-1) \sum_{st} C_{xyst} = -1. \end{aligned}$$

Now define a matrix Y' consisting of the real part of Y (i.e. with coefficients $Y'_{xy} = \mathrm{Re}[Y_{xy}]$). This matrix is real, positive semidefinite,[4] and satisfies $Y_{xx} = \lambda - 1$ for all x and $Y_{xy} = -1$ for $x \sim y$. Therefore Y' is feasible for (4.15) with value $\lambda = \bar{\vartheta}(H)$. Since $\bar{\vartheta}(G)$ is the minimum possible value of (4.15), we have $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$. □

We are now prepared to answer in the affirmative an open question posed by Beigi [Bei10].

Corollary 4.7. *Let $\beta(\overline{H})$ be the largest n such that there exist vectors $|w\rangle \neq 0$ and $|w_s^x\rangle$ with $x \in \{1, \ldots, n\}$ and $s \in V(H)$ which satisfy conditions (4.7)-(4.9). Then $\beta(\overline{H}) = \lfloor \bar{\vartheta}(H) \rfloor$.*

Proof. Considering $K_n \overset{B}{\to} H$, the conditions of definition 4.2 are equivalent to (4.7)-(4.9). Since $\bar{\vartheta}(K_n) = n$, theorem 4.6 gives $K_n \overset{B}{\to} H \iff n \leq \bar{\vartheta}(H)$. Since $\beta(\overline{H})$ is the largest n such that $K_n \overset{B}{\to} H$, we have that $\beta(\overline{H}) = \lfloor \bar{\vartheta}(H) \rfloor$. □

A related corollary can be formed by considering $G \overset{B}{\to} K_n$ rather than $K_n \overset{B}{\to} H$. This defines a set of vectors $|w_s^x\rangle$ satisfying conditions in some sense complementary to Beigi's (4.7)-(4.9). Now we approach ϑ from above:

Corollary 4.8. *Let $\beta_\chi(G)$ be the smallest n such that there exist vectors $|w\rangle \neq 0$ and $|w_s^x\rangle$ with $x \in V(G)$ and $s \in \{1, \ldots, n\}$ for which*

1. $\sum_s |w_s^x\rangle = |w\rangle$

[4] The entrywise complex conjugate of a positive semidefinite matrix is positive semidefinite, so $Y' = (Y + \mathrm{conj}(Y))/2 \succeq 0$.

2. $\langle w_s^x | w_s^y \rangle = 0$ *for all* $x \sim_G y$

3. $\langle w_s^x | w_t^x \rangle = 0$ *for all* $s \neq t$.

Then $\beta_\chi(G) = \lceil \bar{\vartheta}(G) \rceil$.

Proof. Considering $G \xrightarrow{B} K_n$, the conditions of definition 4.2 are equivalent to the conditions stated above. Since $\bar{\vartheta}(K_n) = n$, theorem 4.6 gives $G \xrightarrow{B} K_n \iff \bar{\vartheta}(G) \leq n$. Since $\beta_\chi(G)$ is the smallest n such that $G \xrightarrow{B} K_n$, we have $\beta_\chi(G) = \lceil \bar{\vartheta}(G) \rceil$. □

Corollary 4.9. *The entanglement assisted cost rate is bounded as follows:*

$$\eta^*(G, \overline{H}) \geq \frac{\log \bar{\vartheta}(G)}{\log \bar{\vartheta}(H)}.$$

Proof. Since $\bar{\vartheta}(G^{\boxtimes m}) = \bar{\vartheta}(G)^m$ [Knu94] and $\bar{\vartheta}(H^{*n}) = \bar{\vartheta}(H)^n$ [Lov79], it follows that

$$\begin{aligned}
G^{\boxtimes m} \xrightarrow{*} H^{*n} &\implies G^{\boxtimes m} \xrightarrow{B} H^{*n} && \text{(by theorem 4.4)} \\
&\implies \bar{\vartheta}(G^{\boxtimes m}) \leq \bar{\vartheta}(H^{*n}) && \text{(by theorem 4.6)} \\
&\implies \bar{\vartheta}(G)^m \leq \bar{\vartheta}(H)^n \\
&\implies \frac{\log \bar{\vartheta}(G)}{\log \bar{\vartheta}(H)} \leq \frac{n}{m}.
\end{aligned}$$

Therefore,

$$\eta^*(G, \overline{H}) = \lim_{m \to \infty} \min_n \left\{ \frac{n}{m} : G^{\boxtimes m} \xrightarrow{*} H^{*n} \right\} \geq \frac{\log \bar{\vartheta}(G)}{\log \bar{\vartheta}(H)}.$$

□

Something similar to theorem 4.6 holds for the relation $G \xrightarrow{+} H$. In this case there is an inequality not just for the Lovász ϑ number but also for Schrijver's ϑ' and Szegedy's ϑ^+. Unfortunately, this will no longer be an if-and-only-if statement (but see theorem 4.13 for a weakened converse, and appendix 4.B for a somewhat more complicated if-and-only-if involving $\bar{\vartheta}'$).

Theorem 4.10. *Suppose* $G \xrightarrow{+} H$. *Then* $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$, $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$, *and* $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$.

Proof. As per theorem 4.3, let C be a positive semidefinite matrix satisfying properties (4.10), (4.11), and (4.13). We give the proof for $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$; the others are proved in a similar way. The proof is very similar to that of theorem 4.6, with slight modification due to the fact that the last condition on C is different. Let Z achieve the optimal value for the minimization program (4.16) for $\bar{\vartheta}'(H)$. We will provide a feasible solution for (4.16) for $\bar{\vartheta}'(G)$ to show that $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$. Specifically, let $Y_{xy} = \sum_{st} Z_{st} C_{xyst}$. Since C and Z are positive semidefinite, so is Y.

A feasible solution for (4.16), with value $\bar{\vartheta}'(H)$, requires $Y_{xx} = \bar{\vartheta}'(H) - 1$. However, it suffices to show $Y_{xx} \leq \bar{\vartheta}'(H) - 1$ since equality can be achieved by adding a non-negative diagonal matrix to Y. We have

$$\begin{aligned}
Y_{xx} &= \sum_{st} Z_{st} C_{xxst} \\
&\leq \max_{st} |Z_{st}| \sum_{st} C_{xxst} && \text{(since } C_{xxst} \geq 0\text{)} \\
&\leq \max_{s} |Z_{ss}| \sum_{st} C_{xxst} && \text{(since } Z \succeq 0\text{)} \\
&= \bar{\vartheta}'(H) - 1.
\end{aligned}$$

Similarly, for $x \sim_G y$ we have

$$\begin{aligned} Y_{xy} &= \sum_{st} Z_{st} C_{xyst} = \sum_{s \sim_H t} Z_{st} C_{xyst} \\ &\leq (-1) \sum_{s \sim_H t} C_{xyst} = (-1) \sum_{st} C_{xyst} = -1. \end{aligned} \tag{4.18}$$

Therefore Y is feasible for (4.16) with value $\lambda = \bar{\vartheta}'(H)$. Since $\bar{\vartheta}'(G)$ is the minimum possible value of (4.16), we have $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$.

To show $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$ or $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$, replace inequality with equality in (4.18). For $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$ we have $Z_{st} \geq -1$ for all s, t and need to show $Y_{xy} \geq -1$ for all x, y. This is readily verified:

$$Y_{xy} = \sum_{st} Z_{st} C_{xyst} \geq (-1) \sum_{st} C_{xyst} = -1.$$

□

It is well known that $\alpha(G) \leq \vartheta'(G) \leq \vartheta(G) \leq \vartheta^+(G) \leq \chi(\overline{G})$. We show that similar inequalities hold for the entanglement assisted independence and chromatic numbers. Since ϑ' and ϑ^+ are not multiplicative under the required graph products (appendix 4.A), these do not lead to bounds on asymptotic quantities such as entanglement assisted Shannon capacity or entanglement assisted cost rate.

Corollary 4.11. $\alpha^*(H) \leq \vartheta'(H)$.

Proof. By definition $\alpha^*(H)$ is the largest n such that $K_n \overset{*}{\to} \overline{H}$. But $K_n \overset{*}{\to} \overline{H} \implies K_n \overset{+}{\to} \overline{H} \implies \bar{\vartheta}'(K_n) \leq \vartheta'(H)$. Since $\bar{\vartheta}'(K_n) = n$, the conclusion follows. □

The following corollary was already shown in [BBL+13] via a different method.

Corollary 4.12. *Define $\chi^*(G)$ to be the smallest n such that $G \overset{*}{\to} K_n$. Then $\chi^*(G) \geq \bar{\vartheta}^+(G)$.*

Proof. By definition, $\chi^*(G)$ is the smallest n such that $G \overset{*}{\to} K_n$. But $G \overset{*}{\to} K_n \implies G \overset{+}{\to} K_n \implies \bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(K_n) = n$. □

It would be nice to have a converse to theorem 4.10, like there was with theorem 4.6. Is it the case that $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$, $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$, and $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$ together imply $G \overset{+}{\to} H$? We do not know. However, it is the case that $\bar{\vartheta}^+(G) \leq \bar{\vartheta}'(H) \implies G \overset{+}{\to} H$. In fact, something stronger can be said. We have the following theorem, the consequences of which will be further explored in section 4.5.

Theorem 4.13. $\bar{\vartheta}^+(G) \leq \bar{\vartheta}'(H) \implies G \overset{V}{\to} H$.

Proof. The proof mirrors that of the ($\Longleftarrow$) portion of theorem 4.6, so we only describe the differences. Let $\lambda = \bar{\vartheta}'(H)$. Theorem 4.29 in appendix 4.B gives that

$$\begin{aligned} \bar{\vartheta}'(H) = \max\{ &\|I + T\| : I + T \succeq 0, \\ &T_{st} = 0 \text{ for } s \not\sim t, \\ &T_{st} \geq 0 \text{ for all } s, t\}. \end{aligned}$$

So there is a matrix T such that $\|I + T\| = \lambda$, $I + T \succeq 0$, $T_{st} = 0$ for $s \not\sim t$, and $T_{st} \geq 0$ for all s, t. Since $\lambda \geq \bar{\vartheta}^+(G)$, there is a matrix Z such that $Z \succeq 0$, $Z_{xx} = \lambda - 1$ for all x, $Z_{xy} = -1$ for all $x \sim_G y$, and $Z_{xy} \geq -1$ for all x, y. Note that T and Z satisfy all conditions required by theorem 4.6 plus the additional conditions $T_{st} \geq 0$ for all s, t and $Z_{xy} \geq -1$ for all x, y.

Define B and D as in theorem 4.6. The eigenvector $|\psi\rangle$ corresponding to the maximum eigenvalue of $I+T$ can be chosen to be entrywise non-negative (this follows from the Perron–Frobenius theorem and the fact that $I+T$ is entrywise non-negative). It follows that B can be chosen entrywise non-negative. As before, define

$$C = \lambda^{-1}\left[J \otimes B + (\lambda-1)^{-1} Z \otimes (\lambda D - B)\right].$$

Since T and Z satisfy all conditions needed by theorem 4.6, C satisfies (4.10)-(4.12). To get $G \overset{V}{\to} H$ it remains only to show satisfaction of (4.13): $C_{xyst} \geq 0$ for all x, y, s, t. When $s = t$,

$$\begin{aligned} C_{xyss} &= \lambda^{-1}[D_{ss} + (\lambda-1)^{-1} Z_{xy}(\lambda D_{ss} - D_{ss})] \\ &= \lambda^{-1}(1 + Z_{xy}) D_{ss} \geq 0. \end{aligned}$$

The last inequality follows from $Z_{xy} \geq -1$ and $D_{ss} \geq 0$. When $s \neq t$,

$$\begin{aligned} C_{xyst} &= \lambda^{-1}[B_{st} + (\lambda-1)^{-1} Z_{xy}(0 - B_{st})] \\ &= \lambda^{-1}(\lambda-1)^{-1}[(\lambda-1) - Z_{xy}] B_{st} \geq 0. \end{aligned}$$

The last inequality follows from $Z_{xy} \leq \max\{Z_{xx}, Z_{yy}\} = \lambda - 1$ (since $Z \succeq 0$) and $B_{st} \geq 0$. □

Finally, we show that the two conditions $G \overset{+}{\to} H$ and $G \overset{B}{\to} H$ are not equivalent: the second one is weaker.

Theorem 4.14. *If $G \overset{+}{\to} H$ then $G \overset{B}{\to} H$, but there are graphs for which the converse does not hold.*

Proof. The forward implication is an immediate consequence of theorems 4.6 and 4.10:

$$G \overset{+}{\to} H \implies \bar{\vartheta}(G) \leq \bar{\vartheta}(H) \implies G \overset{B}{\to} H.$$

To see that the converse does not hold, take H to be any graph such that $\lfloor \bar{\vartheta}'(H) \rfloor < \lfloor \bar{\vartheta}(H) \rfloor$. For example, a graph with $\bar{\vartheta}'(H) = 4$ but $\bar{\vartheta}(H) = 16/3 > 5$ is given at the end of [Sch79]. Then $5 = \bar{\vartheta}(K_5) \leq \bar{\vartheta}(H) \implies K_5 \overset{B}{\to} H$ but $5 = \bar{\vartheta}'(K_5) > \bar{\vartheta}'(H) \implies K_5 \overset{+}{\not\to} H$. □

4.5 Quantum homomorphisms

Suppose Alice and Bob share an entangled state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ on Hilbert spaces of arbitrary dimension. A referee asks Alice a question $x \in X$ and Bob a question $y \in Y$. Based on x, Alice performs a (without loss of generality, projective) measurement $\{E_s^x\}_s$ and reports outcome $s \in S$ to the referee. Similarly, Bob performs measurement $\{F_t^y\}_t$ and reports $t \in T$. The sets X, Y, S, T are finite. The probability distribution of Alice and Bob's answer, conditioned upon the referee's question, is

$$P(s,t|x,y) = \langle\psi| E_s^x \otimes F_t^y |\psi\rangle \tag{4.19}$$

where $\sum_s E_s^x = I$, $\sum_t F_t^y = I$, and $\langle\psi|\psi\rangle = 1$.

The assumption that Alice and Bob's measurements take such a tensor product form is associated with non-relativistic quantum mechanics. One may alternatively consider a model in which there is only a single Hilbert space, $|\psi\rangle \in \mathcal{H}_A$ and $E_s^x, F_t^y \in \mathcal{L}(\mathcal{H}_A)$, but in which each E_s^x commutes with each F_t^y. The conditional probability distribution in this model is

$$P(s,t|x,y) = \langle\psi| E_s^x F_t^y |\psi\rangle\,. \tag{4.20}$$

Tsirelson's problem is the question of whether these two models differ. That is to say, is there a conditional probability distribution realizable as (4.20) but not as (4.19)? Tsirelson showed that if the Hilbert spaces are finite dimensional then the two models are the same. A simplified proof appears in [SW08]. In addition to its importance to quantum mechanics, Tsirelson's problem is of mathematical interest since it is closely related to Connes' embedding problem [JNP+11]. Note that any correlation of the form (4.19) can be written in the form (4.20) since $\langle\psi|E_s^x \otimes F_t^y|\psi\rangle = \langle\psi|(E_s^x \otimes I)(I \otimes F_t^y)|\psi\rangle$ and $E_s^x \otimes I$ commutes with $I \otimes F_t^y$.

Graph G is said to have a *quantum homomorphism* to H (written $G \overset{q}{\to} H$) if there is a probability distribution of the form (4.19), with $X = Y = V(G)$, $S = T = V(H)$, and finite dimensional $|\psi\rangle$, satisfying [RM12]

$$\begin{aligned} P(s \neq t | x = y) &= 0 \\ P(s \not\sim_H t | x \sim_G y) &= 0. \end{aligned} \tag{4.21}$$

This is called a "quantum homomorphism" because if Alice and Bob are not allowed to share an entangled state [equivalently, if $\dim(\mathcal{H}_A) = \dim(\mathcal{H}_B) = 1$] then such a conditional probability distribution is achievable if and only if $G \to H$. Although $G \overset{q}{\to} H \implies G \overset{*}{\to} H$, it is an open question whether the converse holds.

In [RM12] it is shown that $G \overset{q}{\to} H$ if and only if there exist projection operators (i.e., Hermitian matrices with eigenvalues in $\{0,1\}$) E_s^x for $x \in V(G)$ and $s \in V(H)$ such that

$$\begin{aligned} &\sum_s E_s^x = I \\ &E_s^x E_t^y = 0 \text{ for all } x \sim_G y, s \not\sim_H t \\ &E_s^x E_t^x = 0 \text{ for all } s \neq t. \end{aligned}$$

Note that the first condition actually implies the third. Define $|w_s^x\rangle = \mathrm{vec}(E_s^x)$. Since $\langle w_s^x | w_t^y \rangle = \mathrm{Tr}(E_s^x E_t^y)$, this gives a set of vectors satisfying the conditions of definition 4.2 for $G \overset{V}{\to} H$. So $G \overset{q}{\to} H \implies G \overset{V}{\to} H$. Since $G \overset{V}{\to} H \implies G \overset{+}{\to} H$, theorem 4.10 gives the following corollary which was previously shown in [Rob13].

Corollary 4.15. *Suppose $G \overset{q}{\to} H$. Then $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$, $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$, and $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$.*

The *quantum chromatic number* $\chi_q(G)$ is the least n such that $G \overset{q}{\to} K_n$, in analogy to the chromatic number $\chi(G)$ which is the least n such that $G \to K_n$. As a means of studying Tsirelson's problem, a number of variations of χ_q were considered in [PT13]. For instance, they define χ_{qr} by taking the correlation model to be (4.20) with infinite dimensional $|\psi\rangle$ rather than (4.19) with finite dimensional $|\psi\rangle$ as was used to define $G \overset{q}{\to} H$ (and thus χ_q). Also, they consider a semidefinite relaxation χ_{vect}. In the language of our paper, $\chi_{\mathrm{vect}}(G)$ is the least n such that $G \overset{V}{\to} K_n$. In fact, our $G \overset{V}{\to} H$ definition was inspired by their work[5]. One could also define $\omega_{\mathrm{vect}}(H)$ as the largest n such that $K_n \overset{V}{\to} H$. Both $\chi_{\mathrm{vect}}(G)$ and $\omega_{\mathrm{vect}}(H)$ can be computed using the tools of section 4.4.

Corollary 4.16. *$\chi_{vect}(G) = \lceil \bar{\vartheta}^+(G) \rceil$ and $\omega_{vect}(H) = \lfloor \bar{\vartheta}'(H) \rfloor$.*

Proof. Theorem 4.13 gives (for integer n) $\bar{\vartheta}^+(G) \leq n \implies G \overset{V}{\to} K_n$ and $n \leq \bar{\vartheta}'(H) \implies K_n \overset{V}{\to} H$. Theorem 4.10 gives the converse, so $\bar{\vartheta}^+(G) \leq n \iff G \overset{V}{\to} K_n$ and $n \leq \bar{\vartheta}'(H) \iff K_n \overset{V}{\to} H$. □

The authors of [PT13] posed the question of whether $\chi_{\mathrm{vect}}(G) = \chi_q(G)$, that is to say whether χ_q is equivalent to its semidefinite relaxation. In fact, these two quantities are not equal.

[5] The first version of the present paper was posted before [PT13]. We later amended this paper to address the question posed in [PT13].

Theorem 4.17. *There is a graph G such that $\chi_{vect}(G) < \chi_q(G)$. Therefore $G \overset{V}{\to} H$ does not imply $G \overset{q}{\to} H$.*

Proof. In light of corollary 4.16, the goal is to find G such that $\lceil \bar{\vartheta}^+(G) \rceil < \chi_q(G)$. The projective rank of a graph, $\xi_f(G)$, is the infimum of d/r such that the vertices of a graph can be assigned rank-r projectors in $\mathbb{C}^d$ such that adjacent vertices have orthogonal projectors. Since $\xi_f(G) \leq \chi_q(G)$ [Rob13], it suffices to find a gap between $\lceil \bar{\vartheta}^+(G) \rceil$ and $\xi_f(G)$.

The five cycle has $\bar{\vartheta}^+(C_5) = \sqrt{5} < \xi_f(C_5) = 5/2$ [Rob13]. But this is not enough since $\lceil \sqrt{5} \rceil = 3 > 5/2$. Fortunately, we can amplify the difference by taking the disjunctive product with a complete graph. $\bar{\vartheta}^+$ is sub-multiplicative under disjunctive product, as feasible solutions $Z + J$ to (4.17) can be combined by tensor product. Theorem 4.27 states that ξ_f is multiplicative under the disjunctive (and lexicographical) product, so

$$\left\lceil \bar{\vartheta}^+(C_5 * K_3) \right\rceil \leq \left\lceil 3\sqrt{5} \right\rceil = 7 < 3 \cdot \frac{5}{2} = \xi_f(C_5 * K_3).$$

□

Subsequently, this result has been strengthened to $\chi_{\text{vect}}(G) < \chi_{\text{qr}}(G)$ [PSS+14].

4.6 Conclusion

Beigi provided a vector relaxation of the entanglement assisted zero-error communication problem, leading to an upper bound on the entanglement assisted independence number: $\alpha^* \leq \lfloor \vartheta \rfloor$ [Bei10]. We generalized Beigi's construction to apply it to entanglement assisted zero-error source-channel coding, defining a relaxed graph homomorphism $G \overset{B}{\to} H$. This ends up exactly characterizing monotonicity of ϑ, and shows that ϑ can be used to provide a lower bound on the cost rate for entanglement assisted source-channel coding. As a corollary we answer in the affirmative an open question posed by Beigi of whether a quantity β that he defined is equal to $\lfloor \vartheta \rfloor$. Applying a Beigi-style argument to chromatic number rather than independence number yields a quantity analogous to β which is equal to $\lceil \vartheta \rceil$. We defined a similar (and stronger) relaxation, $G \overset{+}{\to} H$, which yields bounds involving Schrijver's number ϑ' and Szegedy's number ϑ^+. This leads to a stronger bound on entanglement assisted independence number: $\alpha^* \leq \lfloor \vartheta' \rfloor$. In addition to these new bounds, we reproduce previously known bounds from [Bei10, BBL+13, RM12, Rob13]. We also answer an open question from [PT13] regarding the relation of the quantum chromatic number to its semidefinite relaxation.

A number of open questions remain. Since there is a graph for which $\vartheta' < \vartheta - 1$ [Sch79], our bound $\alpha^* \leq \lfloor \vartheta' \rfloor$ shows a gap between one-shot entanglement assisted zero-error capacity and $\lfloor \vartheta \rfloor$. However, since ϑ' is not multiplicative, it is still not known whether there can be a gap between the *asymptotic* capacity (i.e. the entanglement assisted Shannon capacity) and ϑ. To show such a gap requires a stronger bound on entanglement assisted Shannon capacity. Haemers provided a bound on Shannon capacity which is sometimes stronger than Lovász' bound [Hae78, Hae79, Alo98, Pee96]; however, this bound does not apply to the entanglement assisted case [LMM+12].

The standard notion of graph homomorphism, along with two of its quantum generalizations, and our three relaxations, form a hierarchy as outlined in fig. 4.3. In some cases we do not know whether converses hold. $G \overset{*}{\to} H$ is equivalent to $G \overset{q}{\to} H$ if and only if projective measurements and a maximally entangled state always suffice for entanglement assisted zero-error source-channel coding. Equivalence between $G \overset{*}{\to} H$ and $G \overset{+}{\to} H$ seems unlikely but would have two important consequences. First, we would have a much simpler characterization (vector rather than operator) of entanglement assisted homomorphisms and, in particular, entanglement assisted zero-error communication. Second, since $G \overset{V}{\to} H \implies G \overset{+}{\to} H$, the gap that we found between $G \overset{q}{\to} H$ and $G \overset{V}{\to} H$ would give a gap between $G \overset{q}{\to} H$ and $G \overset{*}{\to} H$.

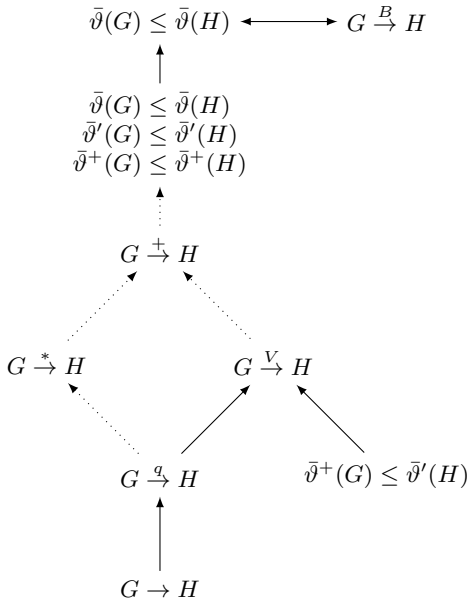


Figure 4.3: Implications between various conditions discussed in this paper. Double ended arrows mean if-and-only-if, solid arrows mean the converse is known to not hold, and dotted arrows mean we do not know whether the converse holds.

After completing this work, we became aware of a previous investigation of a similar problem. Semidefinite relaxations of the homomorphism game (outside of the quantum context) were investigated in [FL92], and this was further developed in [BM95]. What they call a *hoax* corresponds to our $G \overset{V}{\to} H$, and what they call a *semi-hoax* corresponds to our $G \overset{B}{\to} H$. Though they studied the same problem, they reached a different conclusion: they showed (using our terminology)

$$G \overset{B}{\to} H \iff \bar{\vartheta}(G \circ H) = |V(G)| \tag{4.22}$$

$$G \overset{V}{\to} H \iff \bar{\vartheta}'(G \circ H) = |V(G)| \tag{4.23}$$

where $\circ$ denotes the *hom-product* with vertices $V(G) \times V(H)$ and edges

$$(x, s) \sim (y, t) \iff (x \neq y) \text{ and } (x \sim y \implies s \sim t).$$

Combining our theorem 4.6 with (4.22) gives $\bar{\vartheta}(G \circ H) = |V(G)| \iff \bar{\vartheta}(G) \le \bar{\vartheta}(H)$ and combining theorems 4.10 and 4.13 with (4.23) gives

$$\begin{aligned} \bar{\vartheta}^+(G) \le \bar{\vartheta}'(H) &\implies \bar{\vartheta}'(G \circ H) = |V(G)| \\ &\implies \bar{\vartheta}(G) \le \bar{\vartheta}(H), \bar{\vartheta}'(G) \le \bar{\vartheta}'(H), \\ &\quad\ \bar{\vartheta}^+(G) \le \bar{\vartheta}^+(H). \end{aligned}$$

Considering the special case $H = K_n$, we have $G \circ K_n = G \,\square\, K_n$ (Cartesian product) giving $\bar{\vartheta}(G \,\square\, K_n) = |V(G)| \iff \bar{\vartheta}(G) \le n$ and $\bar{\vartheta}'(G \,\square\, K_n) = |V(G)| \iff \bar{\vartheta}^+(G) \le n$, reproducing Theorem 2.7 of [GL08].

4.7 Acknowledgments

We thank Vern Paulsen and Ivan Todorov for inspiring us to define $G \xrightarrow{V} H$ and to find the gap between this and $G \xrightarrow{q} H$.

TC is supported by the Royal Society. LM is supported by the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant "Random numbers from quantum processes" (MOE2012-T3-1-009). DR is supported by an NTU start-up grant awarded to D.V. Pasechnik. SS is supported by the Royal Society and the British Heart Foundation. DS is supported by the National Science Foundation through Grant PHY-1068331. AW is supported by the European Commission (STREPs "QCS" and "RAQUEL"), the European Research Council (Advanced Grant "IRQUAT") and the Philip Leverhulme Trust; furthermore by the Spanish MINECO, project FIS2008-01236, with the support of FEDER funds.

4.A Multiplicativity

In Lovász' original paper [Lov79] on the ϑ function, he proved that

$$\vartheta(G \boxtimes H) = \vartheta(G)\vartheta(H),$$

i.e. ϑ is multiplicative with respect to the strong product. To do this he proved the following two inequalities:

$$\vartheta(G)\vartheta(H) \leq \vartheta(G \boxtimes H) \leq \vartheta(G)\vartheta(H).$$

This sufficed for Lovász because it was only required to show that ϑ is multiplicative with respect to the strong product in order to prove that it was an upper bound on Shannon capacity. However, Lovász also noted that his proof of the first inequality above also proves the following stronger statement:

$$\vartheta(G)\vartheta(H) \leq \vartheta(G * H).$$

Together these inequalities imply that ϑ is multiplicative with respect to both the strong and disjunctive products. Our aim in this appendix is to show that ϑ' is not multiplicative with respect to the strong product and ϑ^+ is not multiplicative with respect to the disjunctive product, but ϑ' is multiplicative with respect to the disjunctive product. Also we will show multiplicativity of projective rank ξ_f.

4.A.1 Counterexamples

Some of the inequalities involving ϑ above can be proved for ϑ' as well. Adapting Lovász' proof of the analogous statement for ϑ, it can be shown that

$$\vartheta'(G \boxtimes H) \geq \vartheta'(G * H) \geq \vartheta'(G)\vartheta'(H).$$

Similarly, it can be shown that

$$\vartheta^+(G * H) \leq \vartheta^+(G \boxtimes H) \leq \vartheta^+(G)\vartheta^+(H).$$

Therefore, in order to show that neither ϑ' or ϑ^+ are multiplicative with respect to both the strong and disjunctive products, we must find counterexamples to both of the following inequalities:

$$\vartheta'(G \boxtimes H) \leq \vartheta'(G)\vartheta'(H), \quad \vartheta^+(G * H) \geq \vartheta^+(G)\vartheta^+(H).$$

At the end of [Sch79], Schrijver gives an example of a graph, which we refer to as G_S, that satisfies $\vartheta'(G_S) < \vartheta(G_S)$. The vertices of G_S are the 0-1-strings of length six, and two strings are adjacent if

their Hamming distance is at most three. In other words, Schrijver's graph G_S is an instance of a *Hamming graph*. Note that this graph is vertex transitive. We will see how to use the graph G_S to construct counterexamples to both of the above inequalities. To do this we will need two lemmas, the first of which is from [Lov79].

Lemma 4.18. *For any graph G,*

$$\vartheta(G)\vartheta(\overline{G}) \geq |V(G)|,$$

with equality when G is vertex transitive.

An analogous statement involving ϑ' and ϑ^+ was proved by Szegedy in [Sze94]:

Lemma 4.19. *For any graph G,*

$$\vartheta'(G)\vartheta^+(\overline{G}) \geq |V(G)|,$$

with equality when G is vertex transitive.

One easy consequence of these lemmas is that if G is a vertex transitive graph such that $\vartheta'(G) < \vartheta(G)$, then

$$\vartheta^+(\overline{G}) = \frac{|V(G)|}{\vartheta'(G)} > \frac{|V(G)|}{\vartheta(G)} = \vartheta(\overline{G}).$$

In particular this implies that $\vartheta^+(\overline{G_S}) > \vartheta(\overline{G_S})$.

More pertinent to our discussion are the following lemmas.

Lemma 4.20. *If G is a vertex transitive graph such that $\vartheta'(G) < \vartheta(G)$, then*

$$\vartheta'(G \boxtimes \overline{G}) > \vartheta'(G)\vartheta'(\overline{G}).$$

Proof. First note the vertices of the form (v, v) in $G \boxtimes \overline{G}$ form an independent set of size $|V(G)|$. Therefore, $\vartheta'(G \boxtimes \overline{G}) \geq |V(G)|$, and we have the following:

$$\vartheta'(G)\vartheta'(\overline{G}) < \vartheta(G)\vartheta(\overline{G}) = |V(G)| \leq \vartheta'(G \boxtimes \overline{G}),$$

since $\vartheta'(\overline{G}) \leq \vartheta(\overline{G})$. □

Since G_S satisfies the hypotheses of lemma 4.20, we have the following desired corollary:

Corollary 4.21. *The parameter ϑ' is not multiplicative with respect to the strong product.*

We are also able to use lemma 4.20 to prove a similar lemma for ϑ^+.

Lemma 4.22. *If G is a vertex transitive graph such that $\vartheta'(G) < \vartheta(G)$, then*

$$\vartheta^+(G * \overline{G}) < \vartheta^+(G)\vartheta^+(\overline{G}).$$

Proof. Suppose that G is such a graph. By lemma 4.20, we have that

$$\vartheta'(G \boxtimes \overline{G}) > \vartheta'(G)\vartheta'(\overline{G}).$$

Since G is vertex transitive, so is $G * \overline{G}$ and thus we can apply lemma 4.19 to obtain

$$\begin{aligned}\vartheta^+(G * \overline{G}) &= \frac{|V(G)|^2}{\vartheta'\left(\overline{G * \overline{G}}\right)} = \frac{|V(G)|^2}{\vartheta'(\overline{G} \boxtimes G)} \\ &< \frac{|V(G)|}{\vartheta'(\overline{G})}\frac{|V(G)|}{\vartheta'(G)} = \vartheta^+(G)\vartheta^+(\overline{G}).\end{aligned}$$

□

Similarly to the above, this implies the following:

Corollary 4.23. *The parameter ϑ^+ is not multiplicative with respect to the disjunctive product.*

Even though ϑ' is not multiplicative with respect to the strong product, nor is ϑ^+ with respect to the disjunctive product, one could ask whether they are at least multiplicative with respect to the corresponding graph powers, as this would be enough to prove an analogue of corollary 4.9. It turns out that they are not, as we now show. Non-multiplicativity for ϑ' was shown already in [BFL10] but with a much smaller gap.

Corollary 4.24. *The parameter ϑ' is not multiplicative under strong graph powers $G^{\boxtimes n}$, and ϑ^+ is not multiplicative under disjunctive graph powers G^{*n}.*

Proof. Let G_S be a vertex transitive graph such that $\vartheta'(G_S) < \vartheta(G_S)$, whose existence was discussed above. Let $H = G_S \oplus \overline{G_S}$ where $\oplus$ denotes disjoint union. Since ϑ' is additive under disjoint union and is super-multiplicative under the strong product,

$$\begin{aligned}
\vartheta'(H^{\boxtimes 2}) &= \vartheta'[G_S^{\boxtimes 2} \oplus (G_S \boxtimes \overline{G_S}) \oplus (\overline{G_S} \boxtimes G_S) \oplus \overline{G_S}^{\boxtimes 2}] \\
&= \vartheta'(G_S^{\boxtimes 2}) + \vartheta'(G_S \boxtimes \overline{G_S}) \\
&\quad + \vartheta'(\overline{G_S} \boxtimes G_S) + \vartheta'(\overline{G_S}^{\boxtimes 2}) \\
&\geq \vartheta'(G_S)^2 + \vartheta'(G_S \boxtimes \overline{G_S}) \\
&\quad + \vartheta'(\overline{G_S} \boxtimes G_S) + \vartheta'(\overline{G_S})^2 \\
&> \vartheta'(G_S)^2 + \vartheta'(G_S)\vartheta'(\overline{G_S}) \\
&\quad + \vartheta'(\overline{G_S})\vartheta'(G_S) + \vartheta'(\overline{G_S})^2 \\
&= [\vartheta'(G_S) + \vartheta'(\overline{G_S}]^2 \\
&= \vartheta'(H)^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\vartheta^+(H^{*2}) &= \vartheta^+[(G_S \oplus \overline{G_S}) * (G_S \oplus \overline{G_S})] \\
&\leq \vartheta^+[G_S^{\boxtimes 2} \oplus (G_S * \overline{G_S}) \\
&\quad \oplus (\overline{G_S} * G_S) \oplus \overline{G_S}^{\boxtimes 2}] \qquad (4.24) \\
&\leq \vartheta^+(G_S)^2 + \vartheta^+(G_S * \overline{G_S}) \\
&\quad + \vartheta^+(\overline{G_S} * G_S) + \vartheta^+(\overline{G_S})^2 \\
&< \vartheta^+(G_S)^2 + \vartheta^+(G_S)\vartheta^+(\overline{G_S}) \\
&\quad + \vartheta^+(\overline{G_S})\vartheta^+(G_S) + \vartheta^+(\overline{G_S})^2 \\
&= [\vartheta^+(G_S) + \vartheta^+(\overline{G_S}]^2 \\
&= \vartheta^+(H)^2.
\end{aligned}$$

where (4.24) follows from the fact that $\vartheta^+(G_1) \geq \vartheta^+(G_2)$ when G_1 is a subgraph of G_2. □

4.A.2 ϑ' and the disjunctive product

Though ϑ' is not multiplicative with respect to the strong product, we are able to show that it is multiplicative with respect to the disjunctive product and the lexicographical product. The lexicographical product $G[H]$ has vertices $V(G) \times V(H)$ and edges $(x, y) \sim (x', y')$ if $x \sim_G x'$ or $(x = x'$ and $y \sim_H y')$.

Theorem 4.25. *Schrijver's number is multiplicative under the disjunctive and the lexicographical products:* $\vartheta'(G * H) = \vartheta'(G[H]) = \vartheta'(G)\vartheta'(H)$.

Proof. We use the following formulation for Schrijver's number:

$$\begin{aligned}\vartheta'(G) = \max\{\langle B, J\rangle : B \succeq 0, \mathrm{Tr}B = 1,\\ B_{ij} \geq 0 \text{ for all } i, j,\\ B_{ij} = 0 \text{ for } i \sim j\}.\end{aligned} \tag{4.25}$$

It is easy to show that $\vartheta'(G*H) \geq \vartheta'(G)\vartheta'(H)$: if B_G and B_H are optimal solutions of (4.25) for $\vartheta'(G)$ and $\vartheta'(H)$ then $B_G \otimes B_H$ is feasible for (4.25) for $\vartheta'(G * H)$ with value $\vartheta'(G)\vartheta'(H)$. Since $G[H]$ is a subgraph of $G * H$, we have $\vartheta'(G[H]) \geq \vartheta'(G * H)$. It remains only to show $\vartheta'(G[H]) \leq \vartheta'(G)\vartheta'(H)$.

Let B be an optimal solution for (4.25) for $\vartheta'(G[H])$. This can be considered as an operator $B \in \mathcal{L}(\mathbb{C}^{|V(G)|}) \otimes \mathcal{L}(\mathbb{C}^{|V(H)|})$, and we have $\langle B, J_G \otimes J_H\rangle = \vartheta'(G[H])$ where J_G is the all ones matrix indexed by $V(G)$ and similarly for J_H. Note that $B_{xyx'y'} = 0$ when $x \sim_G x'$ or $(x = x'$ and $y \sim_H y')$. For $x \in V(G)$ let $|x\rangle$ denote the corresponding basis vector in $\mathbb{C}^{|V(G)|}$ and define

$$B^x = (\langle x| \otimes I)B(|x\rangle \otimes I) \in \mathcal{L}(\mathbb{C}^{|V(H)|}).$$

Since $B^x \succeq 0$, $B^x_{yy'} \geq 0$ for all y, y', and $B^x_{yy'} = 0$ when $y \sim_H y'$, it holds that $B^x/\mathrm{Tr}B^x$ is feasible for (4.25) for $\vartheta'(H)$. The value of this solution is $\langle B^x, J_H\rangle/\mathrm{Tr}B^x$, thus

$$\frac{\langle B^x, J_H\rangle}{\mathrm{Tr}B^x} \leq \vartheta'(H).$$

Also, $\sum_x \mathrm{Tr}B^x = \mathrm{Tr}B = 1$, giving

$$\sum_x \langle B^x, J_H\rangle \leq \sum_x \vartheta'(H)\mathrm{Tr}B^x = \vartheta'(H).$$

Define $B' = (I \otimes \langle \mathbf{1}|)B(I \otimes |\mathbf{1}\rangle) \in \mathcal{L}(\mathbb{C}^{|V(G)|})$ where $|\mathbf{1}\rangle$ is the all ones vector. Note that $B' = \mathrm{Tr}_H\{(I \otimes J_H)B\}$. Since $B' \succeq 0$, $B'_{xx'} \geq 0$ for all x, x', and $B'_{xx'} = 0$ when $x \sim_G x'$, it holds that $B'/\mathrm{Tr}B'$ is feasible for (4.25) for $\vartheta'(G)$. The value of this solution is $\langle B', J_G\rangle/\mathrm{Tr}B'$, thus $\langle B', J_G\rangle/\mathrm{Tr}B' \leq \vartheta'(G)$. Finally,

$$\begin{aligned}\vartheta'(G[H]) &= \langle B, J_G \otimes J_H\rangle = \langle B', J_G\rangle\\ &\leq \vartheta'(G)\mathrm{Tr}B'\\ &= \vartheta'(G)\sum_x \langle B^x, J_H\rangle\\ &\leq \vartheta'(G)\vartheta'(H).\end{aligned}$$

□

4.A.3 What About ϑ^+?

Based on other results concerning ϑ' and ϑ^+, theorem 4.25 seems to suggest that one should be able to prove that ϑ^+ is multiplicative with respect to the strong product. We already noted above that one of the needed inequalities, namely $\vartheta^+(G \boxtimes H) \leq \vartheta^+(G)\vartheta^+(H)$, does hold, so we would only need to show that $\vartheta^+(G \boxtimes H) \geq \vartheta^+(G)\vartheta^+(H)$ holds as well. For now, a proof of this fact eludes us, but we are able to prove the multiplicativity of ϑ^+ in the case of vertex transitive graphs using lemma 4.19 and the multiplicativity of ϑ' with respect to the disjunctive product.

Theorem 4.26. *If G and H are vertex transitive, then*

$$\vartheta^+(G \boxtimes H) = \vartheta^+(G)\vartheta^+(H).$$

Proof. Since G and H are vertex transitive, so is $G \boxtimes H$. Therefore

$$\begin{aligned}\vartheta^+(G \boxtimes H) &= \frac{|V(G)| \cdot |V(H)|}{\vartheta'(\overline{G} * \overline{H})} = \frac{|V(G)| \cdot |V(H)|}{\vartheta'(\overline{G})\vartheta'(\overline{H})} \\ &= \vartheta^+(G)\vartheta^+(H).\end{aligned}$$

□

This seems to be pretty strong evidence that ϑ^+ is multiplicative with respect to the strong product in general.

4.A.4 Projective Rank

The *projective rank* of a graph, $\xi_f(G)$, is the infimum of d/r such that the vertices of a graph can be assigned rank-r projectors in $\mathbb{C}^d$ such that adjacent vertices have orthogonal projectors. Such an assignment is called a *d/r-representation*. The 'f' subscript in the notation for projective rank indicates that it can be thought of as a fractional version of orthogonal rank: the minimum dimension of an assignment of vectors such that adjacent vertices receive orthogonal vectors. We will show ξ_f to be multiplicative under both the disjunctive and the lexicographical products. As a reminder, the lexicographical product $G[H]$ has edges $(x, y) \sim (x', y')$ if $x \sim_G x'$ or $(x = x'$ and $y \sim_H y')$.

Theorem 4.27. *Projective rank is multiplicative under the disjunctive and the lexicographical products: $\xi_f(G * H) = \xi_f(G[H]) = \xi_f(G)\xi_f(H)$.*

Proof. A d_1/r_1-representation for G and a d_2/r_2-representation for H can be turned into a d_1d_2/r_1r_2-representation for $G * H$ by taking the tensor products of the projectors associated with each graph. So $\xi_f(G * H) \leq \xi_f(G)\xi_f(H)$.

On the other hand, let U_{xy} for $x \in V(G), y \in V(H)$ be the subspaces associated with a d/r-representation of $G[H]$. For each x, the subspaces $\{U_{xy} : y\}$ form an r'_x/r projective representation of H where r'_x is the dimension of $\text{span}\{U_{xy} : y\}$, so it must hold that $r'_x/r \geq \xi_f(H)$. Let $r' = \min\{r'_x\}$ and for each x let V_x be an r' dimensional subspace of $\text{span}\{U_{xy} : y\}$. These form a d/r' representation of G, so $d/r' \geq \xi_f(G)$. Then, $d/r = (d/r')(r'/r) \geq \xi_f(G)\xi_f(H)$ so $\xi_f(G[H]) \geq \xi_f(G)\xi_f(H)$. Since $G[H] \subseteq G * H$ we have $\xi_f(G[H]) \geq \xi_f(G)\xi_f(H) \geq \xi_f(G * H) \geq \xi_f(G[H])$. □

4.B An if-and-only-if for Schrijver's number

Monotonicity of Schrijver's number admits an if-and-only-if statement along the lines of theorem 4.6; however, the corresponding conditions on the $|w_s^x\rangle$ vectors are a bit more complicated and there is seemingly no direct connection to entanglement assisted source-channel coding. Specifically, we have the following result:

Theorem 4.28. *$\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$ if and only if there are vectors $|w\rangle \neq 0$ and $|w_s^x\rangle \in \mathbb{C}^d$ for each $x \in V(G)$, $s \in V(H)$, for some $d \in \mathbb{N}$, such that*

1. $\sum_s |w_s^x\rangle = |w\rangle$
2. $\langle w_s^x | w_t^y\rangle = 0$ *for* $s \not\sim_H t$, $s \neq t$
3. $\langle w_s^x | w_s^y\rangle \leq 0$ *for* $x \sim_G y$
4. $\langle w_s^x | w_t^x\rangle = 0$ *for* $s \neq t$
5. $\langle w_s^x | w_t^y\rangle \geq 0$ *for* $s \neq t$.

The proof is a straightforward modification of the proof for theorem 4.6. Before proceeding with this, it is necessary to express $\bar{\vartheta}'$ in a form analogous to (4.14). This characterization appears without proof in [Gal00]; we give the proof below. We do not know how to provide such a formulation for $\bar{\vartheta}^+$, so it may be possible that $\bar{\vartheta}^+$ does not admit an if-and-only-if statement along the lines of theorems 4.6 and 4.28.

Theorem 4.29.

$$\begin{aligned}\bar{\vartheta}'(G) = \max\{&\|I+T\| : I+T \succeq 0,\\ &T_{ij} = 0 \text{ for } i \not\sim j,\\ &T_{ij} \geq 0 \text{ for all } i,j\}.\end{aligned} \tag{4.26}$$

Proof. The dual to the semidefinite program (4.16) is [Sch79]

$$\begin{aligned}\bar{\vartheta}'(G) = \max\{&\langle B, J\rangle : B \succeq 0,\\ &\mathrm{Tr}B = 1,\\ &B_{ij} = 0 \text{ for } i \not\sim j, i \neq j,\\ &B_{ij} \geq 0 \text{ for all } i,j\}.\end{aligned} \tag{4.27}$$

Let T be the optimal solution for (4.26). We will show that this induces a feasible solution for (4.27) via the recipe

$$B = |\psi\rangle \langle\psi| \circ (I+T),$$

where $|\psi\rangle$ is the eigenvector corresponding to the largest eigenvalue of $I+T$. This is positive semidefinite (being the Schur–Hadamard product of two positive semidefinite matrices), and $\langle B, J\rangle = \langle\psi|I+T|\psi\rangle = \lambda$. T_{ii} vanishes, so the diagonal of B is equal to the diagonal of $|\psi\rangle\langle\psi|$; consequently $\mathrm{Tr}B = 1$. The matrix $I+T$ has nonnegative entries so its eigenvector $|\psi\rangle$ can be chosen nonnegative, leading to to $B_{ij} \geq 0$. So B is feasible for (4.27) and (4.27) $\geq$ (4.26).

Conversely, suppose that B is feasible for (4.27) with value λ. Let D be the diagonal component of B. Let $D^{-1/2}$ be the diagonal matrix having entries $D_{ii} = 1/\sqrt{B_{ii}}$ with the convention $1/0 = 0$ (note that $D^{-1/2}$ is the Moore–Penrose pseudoinverse of $D^{1/2}$). Define

$$T = D^{-1/2}(B-D)D^{-1/2}.$$

When $i \not\sim j$, this matrix satisfies $T_{ij} = 0$. Since D and $B-D$ have nonnegative entries, T does as well. We have

$$\begin{aligned}I+T &\succeq D^{-1/2}DD^{-1/2} + T\\ &= D^{-1/2}BD^{-1/2}\\ &\succeq 0.\end{aligned} \tag{4.28}$$

So T is feasible for (4.26). Let $|\psi\rangle$ be the vector with coefficients $\psi_i = \sqrt{B_{ii}}$. Since $\mathrm{Tr}B = 1$, this is a unit vector. Making use of (4.28),

$$\begin{aligned}\langle\psi|I+T|\psi\rangle &\geq \left\langle\psi\middle|D^{-1/2}BD^{-1/2}\middle|\psi\right\rangle\\ &= \sum_{\substack{ij \text{ s.t.}\\ B_{ii}B_{jj}\neq 0}} B_{ij}\\ &= \sum_{ij} B_{ij}\\ &= \langle J, B\rangle = \lambda.\end{aligned} \tag{4.29}$$

Equality (4.29) holds because B is positive semidefinite and so satisfies $B_{ij} = 0$ when $B_{ii}B_{jj} = 0$. Since T is feasible for (4.26),

$$(4.26) \geq \|I + T\| \geq \lambda = (4.27).$$

□

Proof of theorem 4.28. As in the proof of theorem 4.6, we work with the Gram matrix of the $|w_s^x\rangle$ vectors. The existence of vectors satisfying the conditions in the theorem statement is easily seen to be equivalent to the existence of a matrix $C : \mathcal{L}(\mathbb{C}^{|V(G)|}) \otimes \mathcal{L}(\mathbb{C}^{|V(H)|})$ satisfying

$$\begin{aligned}
&C \succeq 0 \\
&\sum_{st} C_{xyst} = 1 \\
&C_{xyst} = 0 \text{ for } s \not\sim t, s \neq t \\
&C_{xyss} \leq 0 \text{ for } x \sim y \\
&C_{xxst} = 0 \text{ for } s \neq t \\
&C_{xyst} \geq 0 \text{ for } s \neq t
\end{aligned}$$

Using this characterization, we proceed with the proof.

($\Longrightarrow$): Suppose $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$. We will explicitly construct a matrix C having the above properties. Let $\lambda = \bar{\vartheta}'(H)$. By theorem 4.29 there is a matrix T such that $\|I + T\| = \lambda$, $I + T \succeq 0$, $T_{st} = 0$ for $s \not\sim t$, and $T_{st} \geq 0$ for all s, t. Let $|\psi\rangle$ be the vector corresponding to the largest eigenvalue of $I + T$, which can be chosen nonnegative since T is entrywise nonnegative. With $\circ$ denoting the Schur–Hadamard product, define the matrices

$$\begin{aligned}
D &= |\psi\rangle\langle\psi| \circ I, \\
B &= |\psi\rangle\langle\psi| \circ (I + T).
\end{aligned}$$

These are entrywise nonnegative. With J being the all-ones matrix and $\langle\cdot,\cdot\rangle$ denoting the Hilbert–Schmidt inner product, it is readily verified that

$$\begin{aligned}
\langle D, J\rangle &= \langle\psi|\psi\rangle = 1, \\
\langle B, J\rangle &= \langle\psi|I + T|\psi\rangle = \lambda.
\end{aligned}$$

Schur–Hadamard products between positive semidefinite matrices yield positive semidefinite matrices. As a consequence, $B \succeq 0$ and

$$\|I + T\| = \lambda \implies \lambda I - (I + T) \succeq 0 \implies \lambda D - B \succeq 0.$$

Since $\lambda \geq \bar{\vartheta}'(G)$, there is a matrix Z such that $Z \succeq 0$, $Z_{xx} = \lambda - 1$ for all x, and $Z_{xy} \leq -1$ for all $x \sim y$. Note that (4.16) gives existence of a matrix with $\bar{\vartheta}'(G) - 1$ on the diagonal, but since $\lambda \geq \bar{\vartheta}'(G)$ we can add a multiple of the identity to get $\lambda - 1$ on the diagonal.

We now construct C. Define

$$C = \lambda^{-1}\left[J \otimes B + (\lambda - 1)^{-1} Z \otimes (\lambda D - B)\right].$$

Since J, B, Z, and $\lambda D - B$ are all positive semidefinite, and $\lambda - 1 \geq 0$, we have that C is positive semidefinite. The other desired conditions on C are easy to verify. For all x, y we have

$$\begin{aligned}
\sum_{st} C_{xyst} &= \lambda^{-1}\left[\langle B, J\rangle + (\lambda - 1)^{-1} Z_{xy}[\lambda\langle D, J\rangle - \langle B, J\rangle]\right] \\
&= 1.
\end{aligned}$$

For $s \not\sim t$, $s \neq t$, we have that $B_{st} = D_{st} = 0$ so $C_{xyst} = 0$. For $x \sim y$,

$$\begin{aligned} C_{xyss} &= \lambda^{-1}\left[B_{ss} + (\lambda-1)^{-1} Z_{xy}(\lambda D_{ss} - B_{ss})\right] \\ &= \lambda^{-1}\left[D_{ss} + (\lambda-1)^{-1} Z_{xy}(\lambda D_{ss} - D_{ss})\right] \\ &= \lambda^{-1} D_{ss}\left[1 + Z_{xy}\right] \\ &\leq 0. \end{aligned}$$

For all x and for $s \neq t$,

$$\begin{aligned} C_{xxst} &= \lambda^{-1}\left[B_{st} + (\lambda-1)^{-1} Z_{xx}(\lambda D_{st} - B_{st})\right] \\ &= \lambda^{-1} B_{st}\left[1 - (\lambda-1)^{-1} Z_{xx}\right] \\ &= 0. \end{aligned}$$

For all x, y and for $s \neq t$,

$$\begin{aligned} C_{xyst} &= \lambda^{-1}\left[B_{st} + (\lambda-1)^{-1} Z_{xy}(\lambda D_{st} - B_{st})\right] \\ &= \lambda^{-1} B_{st}\left[1 - (\lambda-1)^{-1} Z_{xy}\right] \\ &\geq 0, \end{aligned}$$

where the last inequality follows from the fact that $Z \succeq 0 \implies |Z_{xy}| \leq \max\{Z_{xx}, Z_{yy}\} = \lambda - 1$.

($\Longleftarrow$): Let Z achieve the optimal value (call it λ) for the minimization program (4.16) for $\bar{\vartheta}'(H)$. We will provide a feasible solution for (4.16) for $\bar{\vartheta}'(G)$ to show that $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$. Specifically, let

$$Y = \big(I \otimes \langle \mathbf{1}| \,\big)\big[(J \otimes Z) \circ C\big]\big(I \otimes |\mathbf{1}\rangle\,\big),$$

as in the proof of theorem 4.6. Since $C \succeq 0$ and $Z \succeq 0$, and positive semidefiniteness is preserved by conjugation, we have that $Y \succeq 0$. Considering the entries of Y we see that $Y_{xy} = \sum_{st} Z_{st} C_{xyst}$.

Using the fact that $Z_{ss} = \lambda - 1$ and $C_{xxst} = 0$ for $s \neq t$, we have

$$Y_{xx} = \sum_{st} Z_{st} C_{xxst} = (\lambda - 1) \sum_{st} C_{xxst} = \lambda - 1.$$

For $x \sim y$ we have

$$\begin{aligned} Y_{xy} &= \sum_{st} Z_{st} C_{xyst} \\ &= \sum_{s \sim t} \underbrace{Z_{st}}_{\leq -1} \underbrace{C_{xyst}}_{\geq 0} + \sum_{s \not\sim t, s \neq t} Z_{st} \underbrace{C_{xyst}}_{=0} + \sum_{s} \underbrace{Z_{ss}}_{\geq -1} \underbrace{C_{xyss}}_{\leq 0} \\ &\leq \sum_{s \sim t} (-1) C_{xyst} + \sum_{s \not\sim t, s \neq t} (-1) C_{xyst} + \sum_{s} (-1) C_{xyss} \\ &= \sum_{st} (-1) C_{xyst} = -1. \end{aligned}$$

Now define a matrix Y' consisting of the real part of Y (i.e. with coefficients $Y'_{xy} = \mathrm{Re}[Y_{xy}]$). This matrix is real, positive semidefinite, and satisfies $Y_{xx} = \lambda - 1$ for all x and $Y_{xy} \leq -1$ for $x \sim y$. Therefore Y' is feasible for (4.16) with value $\lambda = \bar{\vartheta}'(H)$. Since $\bar{\vartheta}'(G)$ is the minimum possible value of (4.16), we have $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$. □

By setting $G = K_n$ or $H = K_n$ it is possible to formulate corollaries analogous to corollaries 4.7 and 4.8. We describe only the first of these here.

Corollary 4.30. *Let $\beta^-(H)$ be the largest n such that there are vectors $|w\rangle \neq 0$ and $|w_s^x\rangle \in \mathbb{C}^d$ for each $x \in \{1, \ldots, n\}$, $s \in V(H)$, for some $d \in \mathbb{N}$, such that*

1. $\sum_s |w_s^x\rangle = |w\rangle$
2. $\langle w_s^x | w_t^y \rangle = 0$ *for* $s \sim_H t$
3. $\langle w_s^x | w_s^y \rangle \leq 0$ *for* $x \neq y$
4. $\langle w_s^x | w_t^x \rangle = 0$ *for* $s \neq t$
5. $\langle w_s^x | w_t^y \rangle \geq 0$ *for* $s \neq t$.

Then $\beta^-(H) = \lfloor \vartheta'(H) \rfloor$.

Chapter 5

Quantum source-channel coding and non-commutative graph theory[1]

[1] Preprint available: Dan Stahlke, *Quantum source-channel coding and non-commutative graph theory*, arXiv:1405.5254 [quant-ph] (2014).

5.1 Abstract

Alice and Bob receive a bipartite state (possibly entangled) from some finite collection or from some subspace. Alice sends a message to Bob through a noisy quantum channel such that Bob may determine the initial state, with zero chance of error. This framework encompasses, for example, teleportation, dense coding, entanglement assisted quantum channel capacity, and one-way communication complexity of function evaluation.

With classical sources and channels, this problem can be analyzed using graph homomorphisms. We show this quantum version can be analyzed using homomorphisms on non-commutative graphs (an operator space generalization of graphs). Previously the Lovász ϑ number has been generalized to non-commutative graphs; we show this to be a homomorphism monotone, thus providing bounds on quantum source-channel coding. We generalize the Schrijver and Szegedy numbers, and show these to be monotones as well. As an application we construct a quantum channel whose entanglement assisted zero-error one-shot capacity can only be unlocked by using a non-maximally entangled state.

These homomorphisms allow definition of a chromatic number for non-commutative graphs. Many open questions are presented regarding the possibility of a more fully developed theory.

5.2 Introduction

We investigate a quantum version of zero-error source-channel coding (communication over a noisy channel with side information). This includes such problems as zero-error quantum channel capacity (with or without entanglement assistance) [DSW13, Dua09, FS13, BS08], dense coding [BW92], teleportation [BBC+93], function evaluation using one-way (classical or quantum) communication [Wit76, dW01], and measurement of bipartite states using local operations and one-way communication (LOCC-1) [Nat13]. Unless otherwise mentioned all discussion is in the context of zero-error information theory—absolutely no error is allowed.

The problem we consider is as follows. Alice and Bob each receive half of a bipartite state $|\psi_i\rangle$ from some finite collection that has been agreed to in advance (the *source*). Alice sends a message through a noisy quantum channel, and Bob must determine i using Alice's noisy message and his half of the input $|\psi_i\rangle$. The goal is to determine whether such a protocol is possible for a given collection of input states and a given noisy channel. One may also ask how many channel uses are needed per input state if several different input states arrive in parallel. This is known as the *cost rate*. We also consider a variation in which the discrete index i is replaced by a quantum register.

For classical inputs and a classical channel, source-channel coding is possible if and only if there is a graph homomorphism between two suitably defined graphs. Since the Lovász ϑ number of a graph is a homomorphism monotone, it provides a lower bound on the cost rate [NTR06]. This bound also applies if Alice and Bob can make use of an entanglement resource [CMR+13, BBL+13]. We extend the notion of graph homomorphism to non-commutative graphs and show the generalized Lovász ϑ number of [DSW13] to be monotone under these homomorphisms, providing a lower bound on cost rate for quantum source-channel coding.

Schrijver's ϑ' and Szegedy's ϑ^+, which are variations on Lovász's ϑ, are also homomorphism monotones. We generalize these for non-commutative graphs, providing stronger bounds on one-shot quantum channel capacity in particular and on quantum source-channel coding in general. Although ϑ' and ϑ^+ provide only mildly stronger bounds as compared to ϑ for classical graphs, with non-commutative graphs the differences are often dramatic. For classical graphs ϑ' and ϑ^+ are monotone under entanglement assisted homomorphisms [CMR+13], but oddly this is not the case for non-commutative graphs. As a consequence, these quantities can be used to study the power of entanglement assistance. We construct a channel with large one-shot entanglement assisted capacity but no one-shot capacity when assisted by a maximally entangled state.

In section 5.3 we review graph theory and (slightly generalized) classical source-channel coding. In section 5.4 we review the theory of non-commutative graphs and define a homomorphism for

these graphs. In section 5.5 we build the theory of quantum source-channel coding and provide a few basic examples. In section 5.6 we prove that ϑ is monotone under entanglement assisted homomorphisms of non-commutative graphs. In section 5.7 we consider parallel repetition and define various products on non-commutative graphs. In section 5.8 we define Schrijver ϑ' and Szegedy ϑ^+ numbers for non-commutative graphs; we then revisit some examples from the literature and also show that one-shot entanglement assisted capacity for a quantum channel can require a non-maximally entangled state. We conclude with a list of many open questions in section 5.9.

5.3 Classical source-channel coding

We will make use of the following graph theory terminology. A graph G consists of a finite set of *vertices* $V(G)$ along with a symmetric binary relation $x \sim_G y$ (the edges of G). The absence of an edge is denoted $x \not\sim_G y$. The subscript will be omitted when the graph can be inferred from context. We allow loops on vertices. That is to say, we allow $x \sim x$ for some of the $x \in V(G)$. Typically we will be dealing with graphs that do not have loops (*simple graphs*), but allow the possibility due to the utility and insight that loops will afford. We will note the subtleties that this causes, as they arise. We denote by $\overline{G}$ the *complement* of G, having vertices $V(G)$ and edges $x \sim_{\overline{G}} y \iff x \not\sim_G y, x \neq y$. For graphs with loops it is also common to use as the complement the graph with edges $x \sim_{\overline{G}} y \iff x \not\sim_G y$. Fortunately, we will only consider the complement of loop graphs that have loops on all vertices, and in this case the two definitions coincide. A *clique* is a set of vertices $C \subseteq V(G)$ such that $x \sim y$ for all $x, y \in C, x \neq y$. An *independent set* is a clique of $\overline{G}$, equivalently a set $C \subseteq V(G)$ such that $x \not\sim y$ for all $x, y \in C, x \neq y$. The *clique number* $\omega(G)$ is the size of the largest clique, and the *independence number* $\alpha(G)$ is the size of the largest independent set. A *proper coloring* of G is a map $f : G \to \{1, \cdots, n\}$ (an assignment of *colors* to the vertices of G) such that $f(x) \neq f(y)$ whenever $x \sim y$ (note that this is only possible for graphs with no loops). The *chromatic number* $\chi(G)$ is the smallest possible number of colors needed. If no proper coloring exists (i.e. if G has loops) then $\chi(G) = \infty$. The *complete graph* K_n has vertices $\{1, \cdots, n\}$ and edges $x \sim y \iff x \neq y$ (note in particular that K_n does not have loops). G is a *subgraph* of H if $V(G) \subseteq V(H)$ and $x \sim_G y \implies x \sim_H y$.

Suppose Alice wishes to send a message to Bob through a noisy classical channel $\mathcal{N} : S \to V$ such that Bob can decode Alice's message with zero chance of error. How big of a message can be sent? Denote by $\mathcal{N}(v|s)$ the probability that sending $s \in S$ through $\mathcal{N}$ will result in Bob receiving $v \in V$, and define the graph H with vertex set S and with edges

$$s \sim_H t \iff \mathcal{N}(v|s)\mathcal{N}(v|t) = 0 \text{ for all } v \in V. \tag{5.1}$$

Two codewords s and t can be distinguished with certainty by Bob if they are never mapped to the same v. Therefore, the largest set of distinguishable codewords corresponds to the largest clique in H, and the number of such codewords is the clique number $\omega(H)$. We will call H the *distinguishability graph* of the channel $\mathcal{N}$. It is traditional to instead deal with the *confusability graph*, of which (5.1) is the complement. We choose to break with this tradition as this will lead to cleaner notation. Also the distinguishability graph has the advantage of not having loops, making it more natural from a graph-theoretic perspective. In order to facilitate comparison to prior results we will sometimes speak of $\alpha(\overline{H})$ rather than $\omega(H)$ (note that these are equal).

If Bob already has some side information regarding the message Alice wishes to send, the communication task becomes easier: the number of codewords is no longer limited to $\omega(H)$. This situation is known as *source-channel coding*. We will use a slightly generalized version of source-channel coding, as this will aid in the quantum generalization in section 5.5. Suppose Charlie chooses a value i and sends a value x to Alice and u to Bob with probability $P(x, u|i)$. Alice sends Bob a message through a noisy channel. Bob uses Alice's noisy message, along with his side information u, to deduce Charlie's input i (fig. 5.1). This reduces to standard source-channel coding if $P(x, u|i) \neq 0$

only when $x = i$. In other words, the standard scenario has no Charlie, x and u come in with probability $P(x, u)$, and Bob is supposed to produce x.

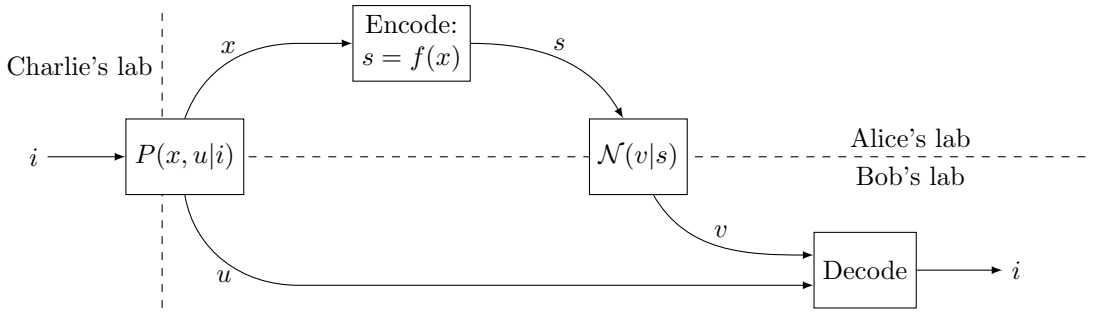


Figure 5.1: Zero-error source-channel coding.

There are a number of reasons one might wish to consider such a scenario. For instance, suppose that $x = i$ always. The side information u might have originated from a previous noisy transmission of x from Alice to Bob. The goal is to resend using channel $\mathcal{N}$ in order to fill in the missing information. Or, the communication complexity of bipartite function evaluation fits into this model. Suppose that Alice and Bob receive x and u, respectively, from a referee Charlie. Alice must send a message to Bob such that Bob may evaluate some function $g(x, u)$. To fit this into the model of fig. 5.1, imagine that Charlie first chooses a value i for g, then sends Alice and Bob some x, u pair such that $g(x, u) = i$. From the perspective of Alice and Bob, determining i is equivalent to evaluating $g(x, u)$. One may ask how many bits Alice needs to send to Bob to accomplish this.

In general, Alice's strategy is to encode her input x using some function $f : X \to S$ before sending it through the channel (a randomized strategy never helps when zero-error is required). As before, Bob receives a value v with probability $\mathcal{N}(v|s)$. The values u and v must be sufficient for Bob to compute i. For a given u, Bob knows Alice's input comes from the set $\{x : \exists i \text{ such that } P(x, u|i) \neq 0\}$. Bob only needs to distinguish between the values of x corresponding to different i, since his goal is to determine i. Define a graph G with vertices $V(G) = X$ and with edges between Alice inputs that Bob sometimes needs to distinguish:

$$x \sim_G y \iff \exists u, \exists i \neq j \text{ such that } P(x, u|i)P(y, u|j) \neq 0. \tag{5.2}$$

This is the *characteristic graph* of the source P. If Bob must sometimes distinguish x from y then Alice's encoding must ensure that x and y never get mapped to the same output by the noisy channel. In other words, her encoding must satisfy $f(x) \sim_H f(y)$ whenever $x \sim_G y$. By definition, this is possible precisely when G is homomorphic to H.

Definition 5.1. *Let G and H be graphs without loops.* G is homomorphic to H, *written $G \to H$, if there is a function $f : V(G) \to V(H)$ such that $x \sim y \implies f(x) \sim f(y)$. The function f is said to be a* homomorphism from G to H.

Graph homomorphisms are examined in great detail in [HT97, HN04]. We state here some basic facts that can be immediately verified.

Proposition 5.2. *Let F, G, H be graphs without loops.*

1. *If $F \to G$ and $G \to H$ then $F \to H$.*
2. *If G is a subgraph of H then $G \to H$.*
3. *The clique number $\omega(H)$ is the largest n such that $K_n \to H$.*

4. *The chromatic number $\chi(G)$ is the smallest n such that $G \to K_n$.*

The above arguments can be summarized as follows.

Proposition 5.3. *There exists a zero-error source-channel coding protocol for source $P(x,u|i)$ and channel $\mathcal{N}(v|s)$ if and only if $G \to H$ where G is the characteristic graph of the source, (5.2), and H is the distinguishability graph of the channel, (5.1).*

As required by definition 5.1, neither G nor H have loops. More precisely, G has a loop if and only if there is an x, u that can occur for two different inputs by Charlie. In this case it is impossible for Alice and Bob to recover Charlie's input, no matter how much communication is allowed.

We emphasize that, although we refer to source-channel coding and use the associated terminology, we are actually considering something a bit more general since we use a source $P(x,u|i)$, with Bob answering i, rather than a source $P(x,u)$, with Bob answering x. Standard source-channel coding, which can be recovered by setting $P(x,u|i) \neq 0 \iff x = i$, was characterized in terms of graph homomorphisms in [NTR06]. Our generalization does not substantially change the theory,[2] and will allow a smoother transition to the quantum version (in the next section).

The Lovász number of the complementary graph, $\bar{\vartheta}(G) = \vartheta(\overline{G})$, is given by the following dual (and equivalent) semidefinite programs: [Lov79, Lov03][3]

$$\begin{aligned} \bar{\vartheta}(G) = \max\{&\|I+T\| : I+T \succeq 0, \\ &T_{ij} = 0 \text{ for } i \not\sim j\}, \end{aligned} \tag{5.3}$$

$$\begin{aligned} \bar{\vartheta}(G) = \min\{&\lambda : \exists Z \succeq J, Z_{ii} = \lambda, \\ &Z_{ij} = 0 \text{ for } i \sim j\}, \end{aligned} \tag{5.4}$$

where we assume that G has no loops. The norm here is the operator norm (equal to the largest singular value), J is the matrix with every entry equal to 1, and $Z \succeq J$ means that $Z - J$ is positive semidefinite. This quantity is a homomorphism monotone in the sense that [dCST13]

$$G \to H \implies \bar{\vartheta}(G) \leq \bar{\vartheta}(H). \tag{5.5}$$

Consequently (see proposition 5.2) we have the Lovász sandwich theorem

$$\omega(G) \leq \bar{\vartheta}(G) \leq \chi(G). \tag{5.6}$$

Since source-channel coding is only possible when $G \to H$, it follows that $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$ is a necessary condition. Two related quantities, Schrijver's $\bar{\vartheta}'$ and Szegedy's $\bar{\vartheta}^+$, which will be defined in section 5.8, have similar monotonicity properties [dCST13] so they provide similar bounds.

Proposition 5.4. *One-shot source-channel coding is possible only if $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$, $\bar{\vartheta}'(G) \leq \bar{\vartheta}'(H)$, and $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$, where G is the characteristic graph of the source, (5.2), and H is the distinguishability graph of the channel, (5.1).*

Traditionally, source-channel coding has been studied in the case where $P(x,u|i) \neq 0$ only when $x = i$. In this case, the following bound holds [NTR06]:[4]

[2] Although, for our generalization extra care needs to be taken when considering parallel repetitions. This will be discussed in section 5.7.

[3] The first of these follows from theorem 6 of [Lov79] by setting $T = A/|\lambda_n(A)|$ (note that in [Lov79] vertices are considered adjacent to themselves). The second comes from page 167 of [Lov03], or from theorem 3 of [Lov79] by taking $Z = \lambda I - A + J$ with λ being the maximum eigenvalue of A.

[4] Actually, [NTR06] seems to have stopped just short of stating such a bound, although they lay all the necessary foundation.

Proposition 5.5. *Suppose $P(x, u|i) \neq 0$ only when $x = i$ and let graphs G and H be given by (5.2) and (5.1). Then m parallel instances of the source can be sent using n parallel instances of the channel only if*

$$\frac{n}{m} \geq \frac{\log \bar{\vartheta}(G)}{\log \bar{\vartheta}(H)}.$$

We will always take logarithms to be base 2. The infimum of n/m (equivalent to the limit as $m \to \infty$) is known as the *cost rate*; proposition 5.5 can be interpreted as an upper bound on the cost rate. This bound relies on the fact that $\bar{\vartheta}$ is multiplicative under various graph products, a property not shared by $\bar{\vartheta}'$ or $\bar{\vartheta}^+$. Propositions 5.4 and 5.5 apply also to the case of entanglement assisted source-channel coding, still with classical inputs and a classical channel [CMR+13]. We will later show (proposition 5.21) that the condition $P(x, u|i) \neq 0$ only when $x = i$ is not necessary in proposition 5.5.

With some interesting caveats, these two theorems in fact also apply to a generalization of source-channel coding in which the source produces bipartite entangled states and in which the channel is quantum. The rest of this paper is devoted to development of this theory.

5.4 Non-commutative graph theory

Given a graph G on vertices $V(G) = \{1, \ldots, n\}$ we may define the operator space

$$S = \operatorname{span}\{|x\rangle \langle y| : x \sim y\} \subseteq \mathcal{L}(\mathbb{C}^n) \tag{5.7}$$

where $|x\rangle$ and $|y\rangle$ are basis vectors from the standard basis. Because we consider symmetric rather than directed graphs, this space is Hermitian: $A \in S \iff A^\dagger \in S$ (more succinctly, $S = S^\dagger$). If G has no loops, S is trace-free (it consists only of trace-free operators). If G has loops on all vertices, S contains the identity.

Concepts from graph theory can be rephrased in terms of such operator spaces. For example, for trace-free spaces the clique number can be defined as the size of the largest set of nonzero vectors $\{|\psi_i\rangle\}$ such that $|\psi_i\rangle \langle\psi_j| \in S$ for all $i \neq j$. Note that since S is trace-free, these vectors must be orthogonal. Although not immediately obvious, this is indeed equivalent to $\omega(G)$ when S is defined as in (5.7).

Having defined clique number in terms of operator spaces, one can drop the requirement that S be of the form (5.7) and can speak of the clique number of an arbitrary Hermitian subspace. Such subspaces, thought of in this way, are called *non-commutative graphs* [DSW13]. Note that [DSW13] requires S to contain the identity, but we drop this requirement and insist only that $S = S^\dagger$. Such a generalization is analogous to allowing the vertices of a graph to not have loops. Dropping also the condition $S = S^\dagger$ would give structures analogous to directed graphs, however we will not have occasion to consider this.

To draw clear distinction between non-commutative graphs and the traditional kind, we will often refer to the latter as *classical graphs*. We will say S *derives from a classical graph* if S is of the form (5.7).

The distinguishability graph of a quantum channel $\mathcal{N} : \mathcal{L}(A) \to \mathcal{L}(B)$ with Kraus operators $\{N_i\}$ can be defined as

$$T = (\operatorname{span}\{N_i^\dagger N_j : \forall i, j\})^\perp \subseteq \mathcal{L}(A) \tag{5.8}$$

where $\perp$ denotes the perpendicular subspace under the Hilbert–Schmidt inner product $\langle X, Y\rangle = \operatorname{Tr}(X^\dagger Y)$. For a classical channel this is equal to (5.7) with G given by (5.1). The space $\operatorname{span}\{N_i^\dagger N_j\}$ (the confusability graph) was considered in [BS08, Dua09, FS13, DSW13]; however, we consider the perpendicular space for the same reason that we considered the distinguishability rather than

the confusability graph in section 5.3: it leads to simpler notation especially when discussing homomorphisms. It will be convenient to use the notation

$$N := \text{span}\{N_i\},$$

and likewise for other sets of Kraus operators so that (5.8) becomes simply

$$T = (N^\dagger N)^\perp, \tag{5.9}$$

with the multiplication of two operator spaces defined to be the linear span of the products of operators from the two spaces. Note that the closure condition for Kraus operators gives $\sum_i N_i^\dagger N_i = I \implies I \in N^\dagger N \implies I \perp T$. Therefore T is trace-free.

In [DSW13] a generalization of the Lovász $\vartheta(G)$ number was provided for non-commutative graphs, which they called $\tilde{\theta}(S)$. We present the definition in terms of $\overline{\vartheta}(S) := \tilde{\vartheta}(S^\perp)$, which should be thought of as a generalization of $\overline{\vartheta}(G) = \vartheta(\overline{G})$.

Definition 5.6 ([DSW13]). *Let $S \subseteq \mathcal{L}(A)$ be a trace-free non-commutative graph. Let A' be an ancillary system of the same dimension as A, and define the vector $|\Phi\rangle = \sum_i |i\rangle_A \otimes |i\rangle_{A'}$. Then $\overline{\vartheta}(S)$ is defined by the following dual (and equivalent) programs:*

$$\overline{\vartheta}(S) = \max\{\|I + X\| : X \in S \otimes \mathcal{L}(A'), I + X \succeq 0\}, \tag{5.10}$$

$$\overline{\vartheta}(S) = \min\{\|Tr_A Y\| : Y \in S^\perp \otimes \mathcal{L}(A'), Y \succeq |\Phi\rangle\langle\Phi|\}. \tag{5.11}$$

We will use the notation $\tilde{\vartheta}(S^\perp) = \overline{\vartheta}(S)$.

When S derives from loop-free graph G via (5.7), this reduces to the standard Lovász number: $\overline{\vartheta}(S) = \vartheta(G)$. Similarly, when S derives from a graph G having loops on all vertices, $\tilde{\vartheta}(S) = \vartheta(G)$. Analogous to the classical case, $\overline{\vartheta}(S)$ gives an upper bound on the zero-error capacity of a quantum channel. In fact, it even gives an upper bound on the zero-error entanglement assisted capacity [DSW13].

Independence number for non-commutative graphs has been investigated in [BS08, Dua09, FS13, DSW13], and in [DSW13] the authors posed the question of whether further concepts from graph theory can be generalized as well. We carry out this program by generalizing graph homomorphisms, which will in turn lead to a chromatic number for non-commutative graphs. These generalized graph homomorphisms will characterize quantum source-channel coding in analogy to proposition 5.3. In fact, one could *define* non-commutative graph homomorphisms as being the relation that gives a generalization of proposition 5.3, but we choose instead to provide more direct justification for our definition.

We begin by describing ordinary graph homomorphisms in terms of operator spaces of the form (5.7); this will lead to a natural generalization to non-commutative graphs. Suppose that $S \subseteq \mathcal{L}(A)$ and $T \subseteq \mathcal{L}(B)$ are derived from graphs G and H via (5.7), and consider a function $f : V(G) \to V(H)$. In terms of S and T, the homomorphism condition $x \sim_G y \implies f(x) \sim_H f(y)$ becomes

$$|x\rangle\langle y| \in S \implies |f(x)\rangle\langle f(y)| \in T, \tag{5.12}$$

where $|x\rangle$ and $|y\rangle$ are vectors from the standard basis. Consider the classical channel that maps $x \to f(x)$. Viewed as a quantum channel, this can be written as the superoperator $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ with the action $\mathcal{E}(|x\rangle\langle x|) = |f(x)\rangle\langle f(x)|$. The Kraus operators of this channel are $E_x = |f(x)\rangle\langle x|$. Again using the notation $E = \text{span}\{E_i\}$, (5.12) can be written $ESE^\dagger \subseteq T$. The generalization to non-commutative graphs is obtained by dropping the condition that $\mathcal{E}$ be a classical channel, allowing instead arbitrary completely positive trace preserving (CPTP) maps.

Definition 5.7. *Let $S \subseteq \mathcal{L}(A)$ and $T \subseteq \mathcal{L}(B)$ be trace-free non-commutative graphs. We write $S \to T$ if there exists a completely positive trace preserving (CPTP) map $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ with Kraus operators $\{E_i\}$ such that*

$$ESE^\dagger \subseteq T \text{ or, equivalently,} \tag{5.13}$$

$$E^\dagger T^\perp E \subseteq S^\perp. \tag{5.14}$$

Equivalently, $S \to T$ if and only if there is a Hilbert space C and an isometry $J : A \to B \otimes C$ such that

$$JSJ^\dagger \subseteq T \otimes \mathcal{L}(C) \text{ or, equivalently,} \tag{5.15}$$

$$J^\dagger(T^\perp \otimes \mathcal{L}(C))J \subseteq S^\perp. \tag{5.16}$$

We will say that the subspace E, or the Kraus operators $\{E_i\}$, or the isometry J, is a homomorphism from S to T.

That (5.13)-(5.16) are equivalent can be seen as follows. (5.13) $\iff$ ($\mathrm{Tr}\{ese'^\dagger t'\} = 0 \ \forall e, e' \in E, s \in S, t' \in T^\perp$) $\iff$ (5.14). Similar reasoning shows (5.15) $\iff$ (5.16), using $(T \otimes \mathcal{L}(C))^\perp = T^\perp \otimes \mathcal{L}(C)$. Equivalence of (5.14) and (5.16) follows from the fact that $E = \mathrm{span}_{|\phi\rangle}\{(I \otimes \langle\phi|)J\}$ where J is related to $\mathcal{E}$ by Stinespring's dilation theorem.

When S and T derive from classical graphs definition 5.7 is equivalent to definition 5.1, as we will now show.

Theorem 5.8. *For non-commutative graphs that derive from classical graphs, definitions 5.1 and 5.7 coincide. In other words, if S and T derive from graphs G and H according to the recipe (5.7) then $G \to H \iff S \to T$.*

Proof. Let S and T be non-commutative graphs deriving from classical graphs G and H.

($\implies$) Suppose $G \to H$. By definition 5.1 there is an $f : G \to H$ such that $x \sim_G y \implies f(x) \sim_H f(y)$. Consider the set of Kraus operators $E_x = |f(x)\rangle\langle x|$. Then,

$$\begin{aligned} ESE^\dagger &= \mathrm{span}\{E_i |x\rangle\langle y| E_j^\dagger : i, j, x \sim_G y\} \\ &= \mathrm{span}\{|f(x)\rangle\langle f(y)| : x \sim_G y\} \\ &\subseteq T. \end{aligned}$$

($\impliedby$) Suppose $S \to T$. By definition 5.7 there is a channel $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ such that $ESE^\dagger \subseteq T$. For each vertex x of G, there is an $i(x)$ such that $E_{i(x)}|x\rangle$ does not vanish. Pick an arbitrary nonvanishing index of the vector $E_{i(x)}|x\rangle$ and call this $f(x)$ so that $\langle f(x)| E_{i(x)} |x\rangle \neq 0$.

Now consider any edge $x \sim_G y$. We have

$$\begin{aligned} |x\rangle\langle y| \in S &\implies E|x\rangle\langle y|E^\dagger \in T \\ &\implies E_{i(x)}|x\rangle\langle y|E_{i(y)}^\dagger \in T. \end{aligned}$$

Define $\tau := E_{i(x)}|x\rangle\langle y|E_{i(y)}^\dagger$. Then $\tau \in T$ and

$$\begin{aligned} \langle f(x)|\tau|f(y)\rangle \neq 0 &\implies \mathrm{Tr}\{\tau |f(y)\rangle\langle f(x)|\} \neq 0 \\ &\implies |f(x)\rangle\langle f(y)| \notin T^\perp \\ &\implies |f(x)\rangle\langle f(y)| \in T \\ &\implies f(x) \sim_H f(y). \end{aligned}$$

Therefore $x \sim_G y \implies f(x) \sim_H f(y)$. □

Definition 5.7 could be loosened to require only that $\sum_i E_i^\dagger E_i$ be invertible (equivalently $E|\psi\rangle \neq \{0\}$ for all $|\psi\rangle$, equivalently $J^\dagger J$ invertible) rather than $\mathcal{E}$ being trace preserving. Theorem 5.8 would still hold; however, definition 5.7 as currently stated has an operational interpretation in terms of quantum source-channel coding (which we will introduce in section 5.5) and satisfies the monotonicity relation $S \to T \implies \overline{\vartheta}(S) \leq \overline{\vartheta}(T)$ (which we will show in section 5.6). Hilbert space structure seems to be important for non-commutative graphs, so it is reasonable to require that J preserve this structure (i.e. J should be an isometry).

As a guide to the intuition, one should not think of $ESE^\dagger$ in (5.13) as density operators $\rho \in S$ going into a channel, like $\sum_i E_i \rho E_i^\dagger$, but rather as a mechanism for comparing the action of the channel on two different states, something like $\{E_i|\psi\rangle\langle\phi|E_j^\dagger : \forall i,j\}$ with $|\psi\rangle\langle\phi| \in S$. But this is only a rough intuition, as S might not necessary be composed of dyads $|\psi\rangle\langle\phi|$. The two copies of E here are analogous to the two Kraus operators appearing in the Knill–Laflamme condition, which we will explore in section 5.5. Note that $E|\psi\rangle$ is equal to the support of $\mathcal{E}(|\psi\rangle\langle\psi|)$.

The non-commutative graph homomorphism of definition 5.7 satisfies properties analogous to those of proposition 5.2.

Proposition 5.9. *Let R, S, T be trace-free non-commutative graphs.*

1. *If $R \to S$ and $S \to T$ then $R \to T$.*
2. *If $S \subseteq T$ then $S \to T$. More generally, if J is an isometry and $JSJ^\dagger \subseteq T$ then $S \to T$.*

Proof. Item 1 follows from considering the composition of channels associated with the homomorphisms $R \to S$ and $S \to T$. Item 2 follows trivially from (5.15), taking space C to be trivial (one dimensional). □

(The condition that appears above, $JSJ^\dagger \subseteq T$ with J an isometry, seems to be a reasonable generalization of the notion of subgraphs for non-commutative graphs, although we won't be making use of this concept. Note that [DSW13] defined *induced subgraphs* as $J^\dagger SJ$. It appears that these two definitions are somewhat incompatible.)

For classical graphs the clique number is the greatest n such that $K_n \to G$ and the chromatic number is the least n such that $G \to K_n$. We use this to extend these concepts to non-commutative graphs. In the previous section, the complete graph K_n was defined to have no loops. The corresponding non-commutative graph, defined via (5.7), is $\text{span}\{|x\rangle\langle y| : x \neq y\}$, the space of matrices with zeros on the diagonal. However, it is reasonable to also consider $(\mathbb{C}I)^\perp$, the space of trace-free operators. We consider both.

Definition 5.10. *For $n \geq 1$ define the* classical and quantum complete graphs

$$K_n = \text{span}\{|x\rangle\langle y| : x \neq y\} \subseteq \mathcal{L}(\mathbb{C}^n),$$
$$Q_n = (\mathbb{C}I)^\perp \subseteq \mathcal{L}(\mathbb{C}^n).$$

One can think of K_n as consisting of the operators orthogonal to the "classical loops" $|x\rangle\langle x|$ and Q_n as consisting of the operators orthogonal to the "coherent loop" I. We use these to define clique, independence, and chromatic numbers for non-commutative graphs. In section 5.5 we will see that all of these quantities have operational interpretations in the context of communication problems. These quantities, and others, are summarized in table 5.1.

Definition 5.11. *Let S be a trace-free non-commutative graph. We define the following quantities.*

1. *$\omega(S)$ is the greatest n such that $K_n \to S$*
2. *$\omega_q(S)$ is the greatest n such that $Q_n \to S$*
3. *$\alpha(S^\perp) = \omega(S)$ and $\alpha_q(S^\perp) = \omega_q(S)$. Note that $I \in S^\perp$.*

4. $\chi(S)$ *is the least* n *such that* $S \to K_n$*, or* ∞ *if* $S \not\to K_n$ *for all* n

5. $\chi_q(S)$ *is the least* n *such that* $S \to Q_n$

The quantities ω_q *and* χ_q *are not to be confused with the quantities of similar name that are discussed in the context of Bell-like nonlocal games [AHKS06, CMN+07, RM12].*

When S derives from a classical graph G, our ω and χ correspond to the ordinary definitions of clique number and chromatic number and our χ_q corresponds to the orthogonal rank $\xi(G)$.[5] This will be proved shortly. For non-commutative graphs with $I \in S$, our definition of $\alpha(S)$ and $\alpha_q(S)$ corresponds to that of [DSW13, Dua09, FS13, BS08], as we will show in theorem 5.13. In other words, when $S = N^\dagger N$ is the confusability graph of a channel $\mathcal{N}$, $\alpha(S)$ and $\alpha_q(S)$ correspond to the one-shot classical and quantum capacities; when $S = (N^\dagger N)^\perp$ the same can be said for $\omega(S)$ and $\omega_q(S)$.

Theorem 5.12. *Let* S *be the non-commutative graph associated with a classical loop-free graph* G*. Then* $\omega(S) = \omega(G)$*,* $\chi(S) = \chi(G)$*,* $\chi_q(S) = \xi(G)$*, and* $\omega_q(S) = 1$*.*

Proof. $\omega(S) = \omega(G)$ and $\chi(S) = \chi(G)$ follow directly from definition 5.11 and proposition 5.2 and theorem 5.8.

An *orthogonal representation* of G is a map from vertices to nonzero vectors such that adjacent vertices correspond to orthogonal vectors. The *orthogonal rank* $\xi(G)$ is defined to be the smallest possible dimension of an orthogonal representation. Let $\{|\psi_x\rangle\}_{x \in V(G)} \subseteq \mathcal{L}(\mathbb{C}^n)$ be an orthogonal representation of G. Without loss of generality assume these vectors to be normalized. The Kraus operators $E_x = |\psi_x\rangle \langle i|$ provide a homomorphism $S \to Q_n$. So $\chi_q(S) \leq \xi(G)$.

Conversely, suppose a set of Kraus operators $\{E_i\}$ provides a homomorphism $S \to Q_n$ with $n = \chi_q(S)$. Because $\sum_i E_i^\dagger E_i = I$, for each $x \in G$ there is an $i(x)$ such that $E_{i(x)} |x\rangle$ does not vanish. Define $|\psi_x\rangle = E_{i(x)} |x\rangle$. For any edge $x \sim y$ of G we have

$$\begin{aligned} |x\rangle \langle y| \in S &\implies E |x\rangle \langle y| E^\dagger \in Q_n \\ &\implies |\psi_x\rangle \langle \psi_y| \in Q_n \\ &\implies \langle \psi_x | \psi_y \rangle = 0. \end{aligned}$$

So $\{|\psi_x\rangle\}_{x \in V(G)}$ is an orthogonal representation of G of dimension n, giving $\xi(G) \leq \chi_q(S)$.

$\omega_q(S) = 1$ because it is not possible to have $Q_n \to S$ if $n > 1$. For, suppose that such a homomorphism E existed. There must be some $x \in V(G)$ and some i such that $\langle x| E_i \neq 0$. Since G is loop free, $|x\rangle \langle x| \in S^\perp$ so $E_i^\dagger |x\rangle \langle x| E_i \in E^\dagger S^\perp E$. But $Q_n^\perp = \mathbb{C}I$ contains no rank-1 operators so $E^\dagger S^\perp E \not\subseteq Q_n^\perp$ and E cannot be a homomorphism from Q_n to S. □

Theorem 5.13. *Let* $S \subseteq \mathcal{L}(A)$ *be a non-commutative graph with* $I \in S$*. Then our* $\alpha(S)$ *and* $\alpha_q(S)$ *are equivalent to the independence number and quantum independence number of [DSW13, Dua09, FS13, BS08].*

Proof. This is a consequence of the operational interpretation of non-commutative graph homomorphisms which we will prove in section 5.5; however, we give here a direct proof. The independence number of [DSW13] is the largest number of nonzero vectors $\{|\psi_i\rangle\}_i$ such that

$$|\psi_i\rangle \langle \psi_j| \in S^\perp \text{ when } i \neq j. \tag{5.17}$$

Given such a collection of n vectors one can define $E_i : \mathbb{C}^n \to A$ as $E_i = |\psi_i\rangle \langle i|$. Since $I \in S$, (5.17) requires orthogonal vectors; thus $\sum_i E_i^\dagger E_i = I$ so these $\{E_i\}$ are indeed Kraus operators. Now,

$$\begin{aligned} E K_n E^\dagger &= \text{span}\{E_{i'} |i\rangle \langle j| E_{j'} : i \neq j\} \\ &= \text{span}\{|\psi_i\rangle \langle \psi_j| : i \neq j\} \subseteq S^\perp, \end{aligned}$$

[5] The *orthogonal rank* of a graph is the smallest dimension of a vector space such that each vertex may be assigned a nonzero vector, with the vectors of adjacent vertices being orthogonal.

giving $K_n \to S^\perp$, or $\alpha(S) \geq n$.

Conversely, take $n = \alpha(S)$. By the definition of $\alpha(S)$, we have $K_n \to S^\perp$. Let $\{E_i\}$ be the Kraus operators that satisfy $EK_nE^\dagger \subseteq S^\perp$, as per definition 5.7. Since $\sum_k E_k^\dagger E_k = I$, for each $i \in \{1, \ldots, n\}$ there must be some $k(i)$ such that $E_{k(i)} |i\rangle \neq 0$. Define $|\psi_i\rangle = E_{k(i)} |i\rangle$. Then for $i \neq j$, $EK_nE^\dagger \subseteq S^\perp \implies |\psi_i\rangle \langle\psi_j| \in S^\perp$.

The quantum independence number is the largest rank projector P such that $PSP = \mathbb{C}P$. Suppose we have such a projector. Let $n = \text{rank}(P)$ and let $J : \mathbb{C}^n \to A$ be an isometry such that $JJ^\dagger = P$. Then $J^\dagger SJ = J^\dagger PSPJ = \mathbb{C}J^\dagger PJ = \mathbb{C}I = Q_n^\perp$. By (5.16), taking C to be the trivial (one-dimensional) space, this gives $Q_n \to S^\perp$, or $\alpha_q(S) \geq n$.

Conversely, take $n = \alpha_q(S)$. Since $Q_n \to S^\perp$, there are Kraus operators $\{E_i\}$ such that $E^\dagger SE \subseteq Q_n^\perp = \mathbb{C}I$, as per (5.14). At least one of these Kraus operators, call it E_0, must satisfy $E_0^\dagger E_0 \neq 0$. Since $I \in S$, $E_0^\dagger I E_0 \in E^\dagger SE \subseteq \mathbb{C}I$, so $E_0^\dagger E_0 = \alpha I$ with $\alpha \neq 0$. Then $J := E_0/\sqrt{\alpha}$ is an isometry and $P := JJ^\dagger$ is a rank n projector. Furthermore, $PSP = JE_0^\dagger SE_0 J^\dagger \subseteq \mathbb{C}JIJ^\dagger = \mathbb{C}P$. □

5.5 Quantum source-channel coding

We construct a quantum version of source-channel coding, as depicted in fig. 5.2. The channel $\mathcal{N}$ from Alice to Bob is now a quantum channel. Instead of classical inputs x and u, Alice and Bob receive a bipartite quantum state. One may imagine that a referee Charlie chooses a bipartite mixed state $\rho_i \in \mathcal{L}(A) \otimes \mathcal{L}(B)$ from some finite collection and sends the A subsystem to Alice and the B subsystem to Bob. The details of the collection $\{\rho_i\}$ are known ahead of time to Alice and Bob. Bob must determine i, with zero chance of error, using Alice's message and his share of ρ_i. We call this *discrete quantum source-channel coding* (discrete QSSC). Here "discrete" refers to i; we will later quantize even this. Discrete QSSC reduces to classical source-channel coding (section 5.3) by taking $\mathcal{N}$ to be a classical channel and the source to be of the form $\rho_i = \sum_{xu} P(x, u|i)|x\rangle \langle x| \otimes |u\rangle \langle u|$.

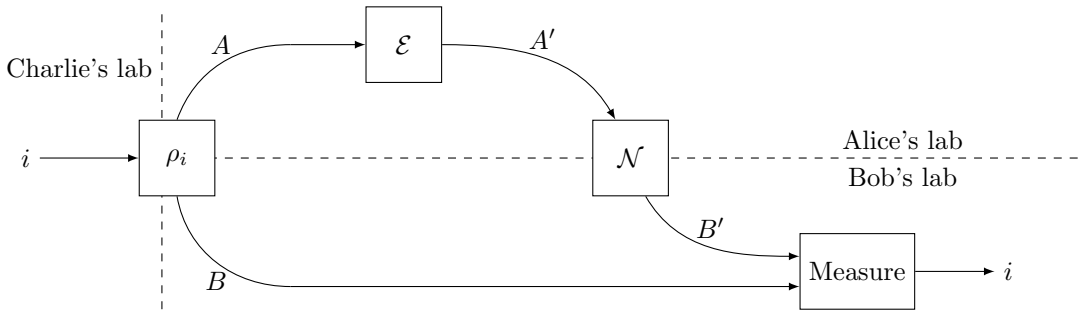


Figure 5.2: Discrete quantum source-channel coding (discrete QSSC).

The most general strategy is for Alice to encode her portion of ρ_i using some quantum operation (some CPTP map) $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(A')$ before sending it through $\mathcal{N}$ to Bob, and for Bob to perform a POVM measurement on the joint state consisting of his portion of ρ_i and the message received from Alice. After receiving Alice's message, Bob is in possession of the mixed state

$$\begin{aligned}\sigma_i &= \mathcal{N}(\mathcal{E}(\rho_i)) \\ &= \sum_{jk} (N_k E_j \otimes I)\rho_i (E_j^\dagger N_k^\dagger \otimes I) \\ &= \sum_{jkl} (N_k E_j \otimes I) |\psi_{il}\rangle \langle\psi_{il}| (E_j^\dagger N_k^\dagger \otimes I),\end{aligned} \tag{5.18}$$

where the unnormalized vectors $|\psi_{il}\rangle$ are defined according to $\rho_i = \sum_l |\psi_{il}\rangle\langle\psi_{il}|$. There is a measurement that can produce the value i with zero error if and only if the states σ_i and $\sigma_{i'}$ are orthogonal whenever $i \neq i'$. Since each term of (5.18) is positive semidefinite we have, with brackets denoting the Hilbert–Schmidt inner product,

$$\begin{aligned}
\langle \sigma_i, \sigma_{i'} \rangle = 0 &\iff \left\langle \psi_{il} \middle| (E_j^\dagger N_k^\dagger \otimes I)(N_{k'} E_{j'} \otimes I) \middle| \psi_{i'l'} \right\rangle = 0 \quad \forall j, j', k, k', l, l' \\
&\iff \langle E_{j'} \mathrm{Tr}_B\{|\psi_{i'l'}\rangle\langle\psi_{il}|\} E_j^\dagger, N_{k'}^\dagger N_k \rangle = 0 \quad \forall j, j', k, k', l, l' \\
&\iff E \cdot \mathrm{Tr}_B\{|\psi_{i'l'}\rangle\langle\psi_{il}|\} \cdot E^\dagger \perp N^\dagger N \quad \forall l, l'.
\end{aligned}$$

By definition 5.7, such an encoding $\mathcal{E}$ exists if and only if $\mathrm{span}\{\mathrm{Tr}_B\{|\psi_{i'l'}\rangle\langle\psi_{il}|\} : \forall i \neq i', \forall l, l'\} \to (N^\dagger N)^\perp$. This immediately leads to the following theorem.

Theorem 5.14. *Consider discrete QSSC (fig. 5.2) with $i \in \{1, \ldots, n\}$. For each i, let $|\psi_i\rangle \in A \otimes B \otimes C$ be a purification of $\rho_i \in \mathcal{L}(A) \otimes \mathcal{L}(B)$. Define the isometry $J = \sum_{i=1}^n |\psi_i\rangle\langle i|$. There is a winning strategy if and only if $S \to T$ where T is the distinguishability graph of $\mathcal{N}$, given by (5.9), and S is the characteristic graph of the source, given by*

$$S = Tr_{BC}\{\mathcal{L}(C) J K_n J^\dagger\}. \tag{5.19}$$

Suppose that Alice and Bob also share an entanglement resource $|\lambda\rangle \in A'' \otimes B''$. This can be absorbed into the source, considering the source to be $\rho_i \otimes |\lambda\rangle\langle\lambda|$. Then (5.19) becomes $\mathrm{Tr}_{BC}\{\mathcal{L}(C) J K_n J^\dagger\} \otimes \Lambda$ where $\Lambda = \mathrm{Tr}_{B''}\{|\lambda\rangle\langle\lambda|\}$. This motivates the following definition:

Definition 5.15. *Let S and T be trace-free non-commutative graphs. We say there is an* entanglement assisted homomorphism *$S \overset{*}{\to} T$ if there exists an operator $\Lambda \succ 0$ such that $S \otimes \Lambda \to T$. The entanglement assisted quantities $\alpha_*(S)$, $\alpha_{q*}(S)$, $\omega_*(S)$, $\omega_{q*}(S)$, $\chi_*(S)$, and $\chi_{q*}(S)$ are defined by using $\overset{*}{\to}$ rather than $\to$ in definition 5.11.*

If S and T are induced by classical graphs G and H then $S \overset{*}{\to} T$ if and only if $G \overset{*}{\to} H$ as defined in [BBL+13, CMR+13]. This equivalence follows from the fact that $S \overset{*}{\to} T$ and $G \overset{*}{\to} H$ have identical operational interpretation in terms of entanglement assisted source-channel coding. Our $\alpha_*(S)$ corresponds to the entanglement assisted independence number of [DSW13] and if S derives from a classical graph our $\chi_*(S)$ corresponds to the entangled chromatic number of [BBL+13, CMR+13]. These quantities, and others, are summarized in table 5.1.

We give some examples.

- Dense coding. Let $\rho_i = |i\rangle\langle i|_{A_1} \otimes |\lambda\rangle\langle\lambda|_{A_2 B}$ where $i \in \{1, \ldots, m\}$ represents the codeword to be transmitted and $|\lambda\rangle\langle\lambda|_{A_2 B}$ is an entanglement resource shared by Alice and Bob. Take $\mathcal{N}$ to be a noiseless quantum channel of dimension n (i.e. a channel of $\log n$ qubits). By theorem 5.14, dense coding is possible if and only if $K_m \otimes \mathrm{Tr}_B\{|\lambda\rangle\langle\lambda|\} \to Q_n$. The well known bound $m \leq n^2$ for dense coding gives ($K_m \overset{*}{\to} Q_n \iff m \leq n^2$). In other words, $\omega_*(Q_n) = n^2$ and $\chi_{q*}(K_{n^2}) = n$.

- Entanglement assisted zero-error communication of n different codewords through a noisy channel $\mathcal{N}$ is possible if and only if $K_n \overset{*}{\to} (N^\dagger N)^\perp$. So the one-shot entanglement assisted classical capacity is $\log \alpha_*(N^\dagger N)$.

- Classical or quantum one-way communication complexity of a function. Suppose the referee sends Alice a classical message x and sends Bob a classical message y, with $(x, y) \in R$. How large of a message must Alice send to Bob such that Bob may compute some function $f(x, y)$? The set R and function f are known ahead of time to all parties.

  Take $\rho_i = \sum_{(x,y) \in R \cap f^{-1}(i)} |x\rangle\langle x| \otimes |y\rangle\langle y|$. Let $S = \mathrm{Tr}_{BC}\{\mathcal{L}(C) J K_n J^\dagger\}$ from (5.19). Then S derives (via (5.7)) from the graph G with edges $x \sim x' \iff \exists y$ s.t. $f(x, y) \neq f(x', y)$. A

classical channel of size n suffices iff $S \to K_n$, and a quantum channel suffices iff $S \to Q_n$. So the smallest sufficient n for a classical channel is $\chi(S)$ and for a quantum channel is $\chi_q(S)$. Since S derives from a classical graph, $\chi(S)$ and $\chi_q(S)$ are just the chromatic number and orthogonal rank of G. This reproduces the result of [Wit76] and theorem 8.5.2 of [dW01].

If Alice and Bob can share an entangled state the condition becomes $S \overset{*}{\to} K_n$ or $S \overset{*}{\to} Q_n$ and the smallest n is $\chi_*(S)$ or $\chi_{q*}(S)$.

- One-way communication complexity of nonlocal measurement. Alice and Bob each receive half of a bipartite state $|\psi_i\rangle \in \mathcal{L}(A) \otimes \mathcal{L}(B)$ drawn from some finite collection agreed to ahead of time. What is the smallest message that must be sent from Alice to Bob so that Bob can determine i? Defining $S = \text{span}\{\text{Tr}_B\{|\psi_i\rangle \langle\psi_j|\} : i \neq j\}$, a quantum message of dimension n suffices if and only if $S \to Q_n$. So the message from Alice to Bob must be at least $\log \chi_q(S)$ qubits or $\log \chi(S)$ bits. If the states $\{|\psi_i\rangle\}$ are not distinguishable via one-way local operations and classical communication (LOCC-1) then $\chi(S) = \infty$.

We further generalize by replacing the index i with a quantum state. Instead of the referee sending ρ_i, we imagine an isometry $J : R \to A \otimes B \otimes C$ into which the referee passes a quantum state $|\psi\rangle \in R$. Alice receives subsystem A, Bob receives B, and C is dumped to the environment. One may think of J as the Stinespring isometry for a channel $\mathcal{J} : \mathcal{L}(R) \to \mathcal{L}(A \otimes B)$. We call this *coherent QSSC*; the setup is depicted in fig. 5.3. The goal is for Bob to reproduce the state $|\psi\rangle$, with perfect fidelity. Discrete QSSC is recovered by taking $J = \sum_i |\psi_i\rangle \langle i|$ where $|\psi_i\rangle \in A \otimes B \otimes C$ is a purification of $\rho_i \in \mathcal{L}(A) \otimes \mathcal{L}(B)$, and requiring that the input state be a basis state.

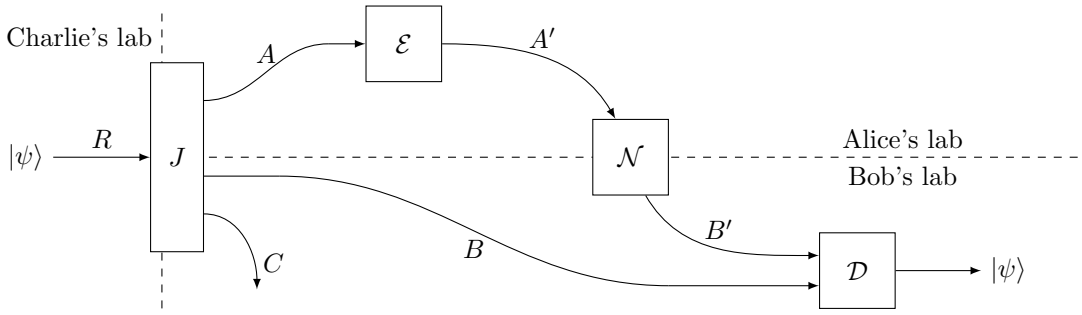


Figure 5.3: Coherent quantum source-channel coding (coherent QSSC).

After Alice's transmission, Bob is in possession of the state $\mathcal{N}(\mathcal{E}(\text{Tr}_C(J|\psi\rangle \langle\psi|J^\dagger)))$. In order to recover $|\psi\rangle$, Bob must perform some operation that converts the channel $\rho \to \mathcal{N}(\mathcal{E}(\text{Tr}_C(J\rho J^\dagger)))$ into the identity channel. The Kraus operators of this channel are $\{(N_k E_j \otimes I_B \otimes \langle l|_C)J\}_{jkl} \subseteq \mathcal{L}(R \to B \otimes B')$. By the Knill–Laflamme error correction condition [KL97], recovery of $|\psi\rangle$ is possible if and only if, $\forall j, j', k, k', l, l'$,

$$J^\dagger \left(E_j^\dagger N_k^\dagger N_{k'} E_{j'} \otimes I_B \otimes |l\rangle \langle l'|_C \right) J \in \mathbb{C}I. \tag{5.20}$$

An operator is proportional to I if and only if it is orthogonal to all trace free operators, so this becomes

$$\begin{aligned} &\left\langle J^\dagger \left(E_j^\dagger N_k^\dagger N_{k'} E_{j'} \otimes I_B \otimes |l\rangle \langle l'|_C \right) J, X \right\rangle = 0 \quad \forall j, j', k, k', l, l', \forall X \in Q_n \\ \iff &\left\langle E_{j'} \text{Tr}_B\{\langle l'|_C JXJ^\dagger |l\rangle_C\} E_j^\dagger, N_{k'}^\dagger N_k \right\rangle = 0 \quad \forall j, j', k, k', l, l', \forall X \in Q_n \\ \iff &E \cdot \text{Tr}_{BC}\{\mathcal{L}(C) J Q_n J^\dagger\} \cdot E^\dagger \subseteq (N^\dagger N)^\perp. \end{aligned}$$

Or, using the terminology of homomorphisms,

Theorem 5.16. *There is a winning strategy for coherent QSSC (fig. 5.3) if and only if $S \to T$ where T is the distinguishability graph of $\mathcal{N}$, given by (5.9), and S is the characteristic graph of the source, given by*

$$S = Tr_{BC}\{\mathcal{L}(C)JQ_nJ^\dagger\} \tag{5.21}$$

where $n = \dim(R)$.

This differs from theorem 5.14 only in the replacement of K_n by Q_n. As before, if Alice and Bob are allowed to make use of an entanglement resource the condition becomes $S \overset{*}{\to} T$ rather than $S \to T$.

We give some examples.

- Teleportation. Take $J = I_{A_1} \otimes |\lambda\rangle_{A_2B}$ where I is the identity operator (i.e. the referee directly gives $|\psi\rangle$ to Alice) and $|\lambda\rangle$ is an entanglement resource. Take $\mathcal{N}$ to be a perfect classical channel. By theorem 5.16 teleportation is possible if and only if $Q_m \otimes \mathrm{Tr}_B\{|\lambda\rangle\langle\lambda|\} \to K_n$ where m is the dimension of the state to be teleported and n is the dimension of the classical channel. The well known bound $m^2 \leq n$ for teleportation gives $(Q_m \overset{*}{\to} K_n \iff m^2 \leq n)$. In other words, $\omega_{q*}(K_{m^2}) = m$ and $\chi_*(Q_m) = m^2$.

- Zero-error one-shot quantum communication capacity. Take $\mathcal{N}$ to be a noisy channel, and take $J : R \to A$ to be the identity operator (i.e. the referee gives $|\psi\rangle$ directly to Alice, and Bob gets no input). It is possible to send $\log m$ error-free qubits though $\mathcal{N}$ if and only if $Q_m \to (N^\dagger N)^\perp$. By definition, $m \leq \alpha_q(N^\dagger N)$. If Alice and Bob can use an entangled state, these conditions become $Q_m \overset{*}{\to} (N^\dagger N)^\perp$ and $m \leq \alpha_{q*}(N^\dagger N)$.

- Suppose Alice and Bob each have a share of a quantum state that has been cloned in the standard basis. That is to say, suppose $J = \sum_{x=1}^{n} |xx\rangle_{AB}\langle x|_R$. Can Alice send a classical message to Bob such that Bob may reconstruct the original quantum state? The characteristic graph of this source (call it S) is the space of trace-free diagonal matrices. Conjugating by the Fourier matrix yields a subspace of K_n. So the Fourier transform is a homomorphism $S \to K_n$; indeed a classical message does suffice.

- Imagine that Alice tries to send a quantum message to Bob, but part of the signal bounces back. This can be modeled by a channel $\mathcal{J} : \mathcal{L}(R) \to \mathcal{L}(A) \otimes \mathcal{L}(B)$. Alice must now send a second message though a second channel $\mathcal{N}$ in order to allow Bob to reconstruct the original message. This is exactly the setup depicted in fig. 5.3, with Charlie being Alice and J being the Stinespring isometry of $\mathcal{J}$.

- Correction of algebras. Suppose instead of transmitting $|\psi\rangle$ perfectly, one needs only that some C^*-algebra of observables $\mathcal{A}$ be preserved (i.e. the receiver can do any POVM measurement with elements from $\mathcal{A}$). This reduces to discrete QSSC when $\mathcal{A}$ consists of the diagonal operators. By theorem 2 of [BKK07], this problem is analyzed via a straightforward modification of the Knill–Laflamme condition: $\mathbb{C}I$ in (5.20) should be replaced by the space of operators that commute with everything in $\mathcal{A}$ (the *commutant* of $\mathcal{A}$); theorem 5.16 is modified by replacing Q_n with the space perpendicular to the commutant of $\mathcal{A}$. Theorem 5.14 is recovered by taking $\mathcal{A}$ to consist of the diagonal operators.

- Consider discrete QSSC with the inputs $\rho_1, \rho_2, \rho_3, \rho_4$ being the four Bell states (or even three of the four). The characteristic graph is Q_2. This is the same as the graph for coherent QSSC with the goal being for Alice to transmit an arbitrary qubit to Bob ($J : R \to A$ is the identity operator). Since the characteristic graphs are the same for the two problems, they require the same communication resources.

Quantity	Interpretation
$K_n = \{M \in \mathcal{L}(\mathbb{C}^n) : M_{ii} = 0\}$	Classical complete graph. The set of $n \times n$ matrices with zeros down the diagonal.
$Q_n = (\mathbb{C}I_n)^\perp$	Quantum complete graph. The set of trace-free $n \times n$ matrices.
$N = \text{span}\{N_i\}$	Span of Kraus operators for channel $\mathcal{N}$.
$N^\dagger N = \text{span}\{N_i^\dagger N_j\}$	Confusability graph of channel $\mathcal{N}$.
$(N^\dagger N)^\perp$	Distinguishability graph of channel $\mathcal{N}$.
$S \to T \iff ESE^\dagger \subseteq T$ with E span of Kraus operators	Graph homomorphism. Source with characteristic graph S can be transmitted using channel with distinguishability graph T.
$S \overset{*}{\to} T \iff (\exists \Lambda \succ 0$ s.t. $S \otimes \Lambda \to T)$	Entanglement assisted homomorphism. As before, but sender and receiver share an entanglement resource.
$\omega(S) = \max\{n : K_n \to S\}$	Clique number. One-shot classical capacity of channel with distinguishability graph S is $\log \omega(S)$.
$\omega_q(S) = \max\{n : Q_n \to S\}$	Quantum clique number. One-shot quantum capacity of channel with distinguishability graph S is $\log \omega(S)$.
$\alpha(S) = \omega(S^\perp)$	Independence number. One-shot classical capacity of channel with confusability graph S is $\log \alpha(S)$.
$\alpha_q(S) = \omega_q(S^\perp)$	Quantum independence number. One-shot quantum capacity of channel with confusability graph S is $\log \alpha(S)$.
$\chi(S) = \min\{n : S \to K_n\}$	Chromatic number. Source with characteristic graph S can be transmitted using $\log \chi(S)$ classical bits.
$\chi_q(S) = \min\{n : S \to Q_n\}$	Quantum chromatic number. Source with characteristic graph S can be transmitted using $\log \chi_q(S)$ qubits. For classical graphs this equals the orthogonal rank.
ω_*, ω_{q*}, α_*, α_{q*}, χ_*, χ_{q*}	Entanglement assisted quantities. Replace $\to$ with $\overset{*}{\to}$ in above definitions. Relevant when sender and receiver share an entanglement resource.

Table 5.1: Basic definitions used in this paper, and their interpretations. See definition 5.7 for the full definition of $S \to T$. See theorems 5.14 and 5.16 for the definition of characteristic graph.

Lemma 2 of [Dua09] states that every non-commutative graph containing the identity is the confusability graph of some channel (equivalently, every trace-free non-commutative graph is the distinguishability graph of some channel). A similar statement holds for sources.

Theorem 5.17. *Every non-commutative graph S is the characteristic graph for discrete QSSC with only two inputs (i.e. ρ_0 and ρ_1).*

Proof. Let $S \in \mathcal{L}(A)$ be a non-commutative graph and let $\{S_x\}_{x\in X}$ be a basis of S, with each S_x being Hermitian. That such a Hermitian basis always exists is shown in [Dua09]. Without loss of generality, assume that each S_x is normalized under the Frobenius norm. Let $(S_x)_{ij}$ be the entries of matrix S_x and define $|S_x\rangle = \sum_{ij}(S_x)_{ij}\,|i\rangle_A\,|j\rangle_B$. Also define $|\Phi\rangle = \dim(A)^{-1/2}\sum_i |i\rangle_A \otimes |i\rangle_B$. Consider discrete QSSC with sources $\rho_i = \mathrm{Tr}_C\{|\psi_i\rangle\langle\psi_i|\}$ for $i \in \{0,1\}$ with $|\psi_i\rangle \in A\otimes B\otimes B'\otimes C$ defined by

$$|\psi_0\rangle = |X|^{-1/2}\sum_x |\Phi\rangle \otimes |x\rangle_{B'} \otimes |x\rangle_C$$
$$|\psi_1\rangle = |X|^{-1/2}\sum_x |S_x\rangle \otimes |x\rangle_{B'} \otimes |x\rangle_C$$

Alice receives subsystem A and Bob receives subsystems $B\otimes B'$. Subsystem C goes to the environment. As per theorem 5.14, the characteristic graph is

$$\begin{aligned} S &= \mathrm{Tr}_{BB'C}\{\mathcal{L}(C)\,|\psi_1\rangle\langle\psi_0|\} + \text{h.c.} \\ &= \mathrm{span}_x\{\mathrm{Tr}_B(|S_x\rangle\langle\Phi|)\} + \text{h.c.} \\ &= \mathrm{span}_x\{S_x\} + \text{h.c.} = S \end{aligned}$$

where "+h.c." means that the adjoints of the operators are also included in the subspace. □

Note that we didn't require S to be trace-free in theorem 5.17; however, if S is not trace-free then source-channel coding will be impossible: ρ_0 and ρ_1 would be non-orthogonal and so would not be distinguishable by any measurement.

5.6 $\overline{\vartheta}$ is a homomorphism monotone

We will show that $\overline{\vartheta}$ is monotone under entanglement assisted homomorphisms of non-commutative graphs. This leads to a Lovász sandwich theorem for non-commutative graphs, and a bound on quantum source-channel coding. We begin by showing $\overline{\vartheta}$ to be insensitive to entanglement. Recall that a source having non-commutative graph S, combined with an entanglement resource $|\lambda\rangle \in A''\otimes B''$, yields a composite source with non-commutative graph $S\otimes\Lambda$ where $\Lambda = \mathrm{Tr}_{B''}\{|\lambda\rangle\langle\lambda|\}$.

Lemma 5.18. *Let S be a trace-free non-commutative graph. Let Λ be a positive operator. Then $\overline{\vartheta}(S) = \overline{\vartheta}(S\otimes\Lambda)$.*

Proof. Suppose $S \subseteq \mathcal{L}(A)$ and $\Lambda \in \mathcal{L}(B)$. By (5.10) we have

$$\overline{\vartheta}(S) = \max\{\|I+T\| : T \in S\otimes\mathcal{L}(\mathbb{C}^m), I+T \succeq 0\} \tag{5.22}$$
$$\overline{\vartheta}(S\otimes\Lambda) = \max\{\|I+T\| : T \in S\otimes\Lambda\otimes\mathcal{L}(\mathbb{C}^n), I+T \succeq 0\}. \tag{5.23}$$

In [DSW13] is it shown that the ancillary space can be taken to be any dimension at least as large as $\dim(A)$, so in (5.22)-(5.23) we may take any values $m \geq \dim(A)$ and $n \geq \dim(A\otimes B)$.

Take $n = \dim(A\otimes B)$ and $m = n\dim(B)$. Any T feasible for (5.23) is also feasible for (5.22) since $\Lambda\otimes\mathcal{L}(\mathbb{C}^n) \subseteq \mathcal{L}(\mathbb{C}^m)$. So $\overline{\vartheta}(S) \geq \overline{\vartheta}(S\otimes\Lambda)$.

Now take $m = n = \dim(A\otimes B)$. Let T be feasible for (5.22). Without loss of generality, assume $\|\Lambda\| = 1$. Then $T' := T\otimes\Lambda$ is feasible for (5.23). Indeed, $T \succeq -I \implies T' \succeq -I$ since $\Lambda \succ 0$ and $\|\Lambda\| \leq 1$. Also, $\|I+T'\| \geq \|I+T\|$ since $\Lambda \succ 0$ and $\|\Lambda\| \geq 1$. So $\overline{\vartheta}(S\otimes\Lambda) \geq \overline{\vartheta}(S)$. □

Before we prove the main theorem, we introduce some notation that will also be used in section 5.8. For any (finite dimensional) Hilbert space A, define the state

$$|\Phi\rangle_A = \sum_i |i\rangle_A \otimes |i\rangle_{A'} . \tag{5.24}$$

where A' is another Hilbert space of the same dimension as A. Note that this provides an isomorphism between A and the dual space of A' via the action $|\psi\rangle_A \to \langle\Phi|(|\psi\rangle_A \otimes I_{A'})$. A bar over an operator denotes entrywise complex conjugation in the standard basis (i.e. the basis used in (5.24)). Additionally, the bar will be understood to move an operator to the primed spaces (or from primed to unprimed). For example, if $J : A \to B \otimes C$ then $\overline{J} : A' \to B' \otimes C'$ is equal to

$$\overline{J} = \mathrm{Tr}_{BC}\{|\Phi\rangle_{BC} \langle\Phi|_A J^\dagger\}. \tag{5.25}$$

We now prove the main theorem of this section: that $\overline{\vartheta}$ is monotone under entanglement assisted homomorphisms. Such an inequality was already known for classical graphs [CMR+13].

Theorem 5.19. *Let S and T be trace-free non-commutative graphs. If $S \to T$ or $S \overset{*}{\to} T$ then $\overline{\vartheta}(S) \leq \overline{\vartheta}(T)$.*

Proof. If we prove monotonicity under $S \to T$ then monotonicity under $S \overset{*}{\to} T$ follows. Indeed, if $S \overset{*}{\to} T$ then there is a $\Lambda \succ 0$ such that $S \otimes \Lambda \to T$. Supposing that $\overline{\vartheta}$ is monotone under (non-entanglement assisted) homomorphisms, we have $\overline{\vartheta}(S \otimes \Lambda) \leq \overline{\vartheta}(T)$. Lemma 5.18 then gives $\overline{\vartheta}(S) \leq \overline{\vartheta}(T)$.

We now show that $S \to T$ implies $\overline{\vartheta}(S) \leq \overline{\vartheta}(T)$. This can be seen as a consequence of corollaries from [DSW13]. Let $S \subseteq \mathcal{L}(A)$ and $T \subseteq \mathcal{L}(B)$ be trace-free non-commutative graphs with $S \to T$. By definition 5.7 there is a Hilbert space C and an isometry $J : A \to B \otimes C$ such that $J^\dagger(T^\perp \otimes \mathcal{L}(C))J \subseteq S^\perp$. Then

$$\begin{aligned}
\tilde{\vartheta}(T^\perp) &= \tilde{\vartheta}(T^\perp)\tilde{\vartheta}(\mathcal{L}(C)) && \text{(Since } \tilde{\vartheta}(\mathcal{L}(C)) = 1\text{)} \\
&= \tilde{\vartheta}(T^\perp \otimes \mathcal{L}(C)) && \text{(Corollary 10 of [DSW13])} \\
&\geq \tilde{\vartheta}(J^\dagger(T^\perp \otimes \mathcal{L}(C))J) && \text{(Corollary 11 of [DSW13])} \\
&\geq \tilde{\vartheta}(S^\perp). && \text{(Corollary 11 of [DSW13])}
\end{aligned}$$

We present also a more direct proof, since this can later be generalized for the $\overline{\vartheta}'_{\mathcal{C}}$ and $\overline{\vartheta}^+_{\mathcal{C}}$ quantities of section 5.8. Let $S \subseteq \mathcal{L}(A)$ and $T \subseteq \mathcal{L}(B)$ be trace-free non-commutative graphs with $S \to T$. By definition there is a CPTP map $\mathcal{E} : \mathcal{L}(A) \to \mathcal{L}(B)$ with Kraus operators $\{E_i\}$ such that $E^\dagger T^\perp E \subseteq S^\perp$ where $E = \mathrm{span}\{E_i\}$. Let $J : A \to B \otimes C$ be the Stinespring isometry for the channel $\mathcal{E}$, so that $J = \sum_i |i\rangle_C E_i$. Let $\overline{J} : A' \to B' \otimes C'$ be the entrywise complex conjugate of J. Recall that $\overline{J}$ takes the form (5.25).

Let $Y' \subseteq \mathcal{L}(B) \otimes \mathcal{L}(B')$ be an optimal solution for (5.11) for $\overline{\vartheta}(T)$. Define $Y \subseteq \mathcal{L}(A) \otimes \mathcal{L}(A')$ as

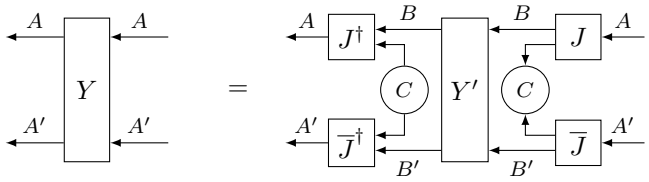
$$\begin{aligned}
Y &= \sum_{ij} (E_i \otimes \overline{E}_i)^\dagger Y' (E_j \otimes \overline{E}_j) \\
&= (J \otimes \overline{J})^\dagger (|\Phi\rangle_C \otimes Y' \otimes \langle\Phi|_C)(J \otimes \overline{J}).
\end{aligned} \tag{5.26}$$

We have that

$$\begin{aligned}
Y' \in T^\perp \otimes \mathcal{L}(B') &\implies Y \in E^\dagger T^\perp E \otimes \overline{E}^\dagger \mathcal{L}(B') \overline{E} \\
&\implies Y \in S^\perp \otimes \mathcal{L}(A').
\end{aligned} \tag{5.27}$$

Dirac notation becomes unworkable with so many Hilbert spaces, so we will make use of diagrams similar to those of [GWYC06] (such diagrams have been used for example in [SG11,BO11]). Operators and states are denoted by labeled boxes and multiplication (or traces or tensor contraction) by wires. This is very much like standard quantum circuits, except that the diagram has no interpretation of time ordering. Wires have an arrow pointing away from the ket space and towards the bra space, and are labeled with the corresponding Hilbert space. Labeled circles represent the transposer operators (5.24):

With this notation, (5.26) becomes

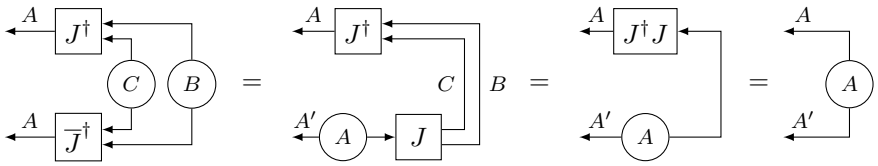


(5.28)

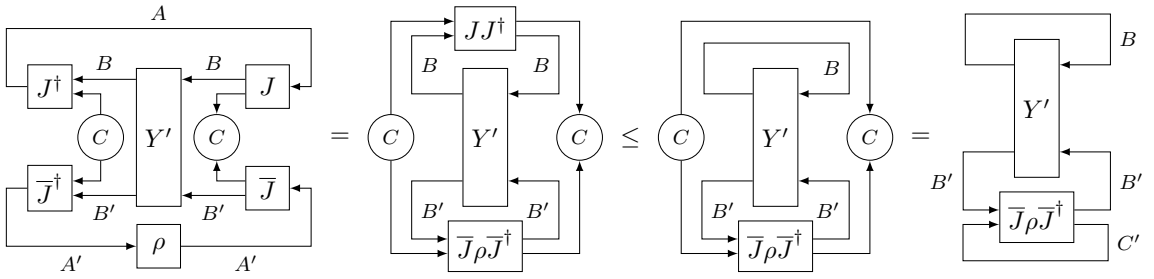
The operator $(J \otimes \overline{J})^\dagger(|\Phi\rangle_C \otimes I_{BB'})$ turns the transposer $|\Phi\rangle_B$ into $|\Phi\rangle_A$:

$$\begin{aligned}(J \otimes \overline{J})^\dagger(|\Phi\rangle_C \otimes |\Phi\rangle_B) &= (J^\dagger \otimes \overline{J}^\dagger)(|\Phi\rangle_C \otimes |\Phi\rangle_B) \\ &= (J^\dagger J \otimes I)\,|\Phi\rangle_A = |\Phi\rangle_A\,.\end{aligned}$$

With diagrams, the same derivation is written as



Consequently we have $Y' \succeq |\Phi\rangle\langle\Phi|_B \implies Y \succeq |\Phi\rangle\langle\Phi|_A$ so Y is feasible for (5.11) for $\overline{\vartheta}(S)$. To get $\overline{\vartheta}(S) \leq \overline{\vartheta}(T)$ it remains only to show $\|\mathrm{Tr}_A Y\| \leq \|\mathrm{Tr}_B Y'\|$. Let $\rho \in \mathcal{L}(A')$ be a density operator achieving $\mathrm{Tr}\{(I_A \otimes \rho)Y\} = \|\mathrm{Tr}_A Y\|$. Plugging (5.28) (the definition of Y) into $\mathrm{Tr}\{(I_A \otimes \rho)Y\}$ gives



The first equality involves only a rearrangement of the diagram. The inequality follows from the fact that $JJ^\dagger \preceq I_{BC}$ (since $JJ^\dagger$ is a projector) and the rest of the diagram represents a positive semidefinite operator.[6] The last equality uses $\mathrm{Tr}_C\{|\Phi\rangle\langle\Phi|_C\} = I_{C'}$. The same derivation can be

[6] If a portion of a diagram has reflection symmetry, with the operators located on the reflection axis being positive semidefinite, then that portion of the diagram is positive semidefinite [GWYC06]. This follows from the fact that $M \succeq 0 \implies N^\dagger M N \succeq 0$.

written in equations as

$$\begin{aligned}\mathrm{Tr}_{AA'}\{(I_A \otimes \rho)Y\} &= \mathrm{Tr}_{AA'}\{(I_A \otimes \rho)(J \otimes \overline{J})^\dagger(|\Phi\rangle_C \otimes Y' \otimes \langle\Phi|_C)(J \otimes \overline{J})\} \\ &= \mathrm{Tr}_{BB'CC'}\{(JJ^\dagger \otimes \overline{J}\rho\overline{J}^\dagger)(|\Phi\rangle_C \otimes Y' \otimes \langle\Phi|_C)\} \\ &\leq \mathrm{Tr}_{BB'CC'}\{(I_{BC} \otimes \overline{J}\rho\overline{J}^\dagger)(|\Phi\rangle_C \otimes Y' \otimes \langle\Phi|_C)\} \\ &= \mathrm{Tr}_{BB'C'}\{(I_B \otimes \overline{J}\rho\overline{J}^\dagger)(Y' \otimes I_{C'})\}. \end{aligned} \tag{5.29}$$

Since $\overline{J}$ is an isometry, $\mathrm{Tr}_{C'}\{\overline{J}\rho\overline{J}^\dagger\}$ is a density operator. So (5.29) is bounded from above by $\|\mathrm{Tr}_B Y'\|$ and we have $\|\mathrm{Tr}_A Y\| \leq \|\mathrm{Tr}_B Y'\|$ as desired. □

An immediate corollary is that $\overline{\vartheta}$ satisfies a sandwich theorem analogous to (5.6).

Corollary 5.20. *Let S be a trace-free non-commutative graph. Then*

1. $\omega_*(S) \leq \overline{\vartheta}(S) \leq \chi_*(S)$.
2. $\omega_{q*}(S) \leq \sqrt{\overline{\vartheta}(S)} \leq \chi_{q*}(S)$.

Proof. This follows from applying theorem 5.19 to definition 5.15 and using the fact that $\overline{\vartheta}(K_n) = n$ and $\overline{\vartheta}(Q_n) = n^2$. That $\overline{\vartheta}(K_n)$ and $\overline{\vartheta}(Q_n)$ take these values is not hard to see, and is also proved in [DSW13]. □

The bound on $\omega_*(S)$ and $\omega_{q*}(S)$ was shown already in [DSW13], although it was stated in terms of α_* and α_{q*}. The bound on $\chi_*(S)$ and $\chi_{q*}(S)$ is new, although the bound on $\chi_*(S)$ was known already for classical graphs [BBL+13]. Note that, since for example $\omega(S) \leq \omega_*(S)$, we also have the weaker sandwich theorem $\omega(S) \leq \overline{\vartheta}(S) \leq \chi(S)$ and $\omega_q(S) \leq \sqrt{\overline{\vartheta}(S)} \leq \chi_q(S)$.

This bound $\sqrt{\overline{\vartheta}(S)} \leq \chi_q(S)$ is not particularly tight when S corresponds to a classical graph G, for the following reason. In such cases $\chi_q(S) = \xi(G)$, the orthogonal rank. But it is known that $\overline{\vartheta}(S) = \bar{\vartheta}(G) \leq \xi(G)$ [Lov79], so in this case the square root over $\overline{\vartheta}$ is unnecessary. The necessity of the square root arises from the possibility of dense coding, since we are bounding the entanglement assisted quantities in corollary 5.20. Notice that $\omega_{q*}(S) = \left\lfloor \sqrt{\omega_*(S)} \right\rfloor$ and $\chi_{q*}(S) = \left\lceil \sqrt{\chi_*(S)} \right\rceil$ since a quantum channel of dimension n can simulate a classical channel of dimension n^2, and teleportation can do the reverse.

The square root in corollary 5.20 could be eliminated by defining a different generalization of Lovász's $\bar{\vartheta}$ that is monotone under homomorphisms and which takes the value n on the graph Q_n. Such a quantity would necessarily not be monotone under entanglement assisted homomorphisms, since $Q_2 \overset{*}{\to} K_4$. Finding such a quantity is left as an open question.

Theorem 5.19 can be applied to give bounds for all of the examples in section 5.5. Two are especially noteworthy: theorem 5.19 gives the well known bounds $n \leq m^2$ for dense coding and $n^2 \leq m$ for teleportation (where n is the dimension of the source and m is the dimension of the channel).

5.7 Graph products and parallel repetition

Consider the problem of sending several parallel sources using several parallel channels. In general these several sources (as well as the channels) could all be distinct, and we will in fact consider this. In the special case where the sources are identical, as well as the channels, one may ask how many channel uses are required for each instance of the source. This is known as the cost rate. For classical sources and channels, we saw already (proposition 5.5) that a bound on cost rate is given in terms of

the Lovász $\bar{\vartheta}$ number. The goal of this section is to prove an analogous bound in the case of quantum sources and channels. To build this theory, we begin with an investigation of the classical case.

Consider two channels $\mathcal{N}(v|s)$ and $\mathcal{N}'(v'|s')$ having distinguishability graphs H and H'. It is not hard to see that the composite channel $\mathcal{N}''(v, v'|s, s') = \mathcal{N}(v|s)\mathcal{N}'(v'|s')$ has a distinguishability graph with vertices $V(H) \times V(H')$ and edges

$$(x, x') \sim (y, y') \iff (x \sim y) \text{ or } (x' \sim y'). \tag{5.30}$$

This is known as the *disjunctive product*, denoted $H * H'$. If n identical copies of $\mathcal{N}$ are used in parallel, the resulting composite channel will have distinguishability graph $H^{*n} = H * H * \ldots * H$. Since the one-shot capacity of a channel is $\log \omega(H)$ bits, the capacity (per-channel use) of n parallel channels is $\frac{1}{n} \log \omega(H^{*n})$. The capacity in the limit $n \to \infty$ is known as the *Shannon capacity* of the channel,

$$C_0(\overline{H}) = \lim_{n\to\infty} \frac{1}{n} \log \omega(H^{*n}). \tag{5.31}$$

The complement in the argument of $C_0(\overline{H})$ is because we consider the distinguishability graph rather than the confusability graph, in terms of which C_0 is typically defined. Since $\bar{\vartheta}(H^{*n}) = \bar{\vartheta}(H)^n$, it holds that $C_0(\overline{H}) \leq \log \bar{\vartheta}(H)$ [Lov79]. In fact this was the original motivation for defining the ϑ number.

Now consider parallel sources. Recall from section 5.3 that the sources $P(x, u|i)$ we consider are somewhat generalized from what is traditionally considered. The traditional definition is obtained by requiring $P(x, u|i) \neq 0$ only when $x = i$. In this case, the characteristic graphs of parallel sources combine by the *strong product* [Wit76] which has vertices $V(G) \times V(G')$ and edges

$$\begin{aligned}(x, x') \sim (y, y') \iff &(x \sim y \text{ and } x' \sim y') \text{ or} \\ &(x = y \text{ and } x' \sim y') \text{ or} \\ &(x \sim y \text{ and } x' = y').\end{aligned} \tag{5.32}$$

Adapting this to non-commutative graphs is problematic because there is no clear analogue of the condition $x = y$. But already for our generalized sources, which can have $P(x, u|i) \neq 0$ when $x \neq i$, the product rule needs modification.

Consider two parallel sources $P(x, u|i)$ and $P(x', u'|i')$ (these can be over different alphabets) with characteristic graphs G and G'. Call the combined source $P''(x, x', u, u'|i, i') := P(x, u|i)P(x', u'|i')$. This has characteristic graph G'' with vertex set $V(G) \times V(G')$ and edges given by a generalization of (5.32). To this end, we introduce a graph G_0 having the same vertices as G but with edges

$$x \sim_{G_0} y \iff \exists u, \exists i \text{ s.t. } P(x, u|i)P(y, u|i) \neq 0. \tag{5.33}$$

G'_0 is defined similarly. If $P(x, u|i) \neq 0$ only when $x = i$ then G_0 has edges $x \sim_{G_0} y \iff x = y$. In other words G_0 consists only of loops. So (5.33) can be regarded as a set of generalized loops. We will call the pair (G, G_0) a *graph with generalized loops*. We can now compute G'', the characteristic graph for the composite source:

$$\begin{aligned}(x, x') \sim_{G''} (y, y') \iff &\exists u, u', \exists (i, i') \neq (j, j') \text{ s.t.} \\ &P''(x, x', u, u'|i, i')P''(y, y', u, u'|j, j') \neq 0 \\ \iff &(x \sim_G y \text{ and } x' \sim_{G'} y') \text{ or} \\ &(x \sim_{G_0} y \text{ and } x' \sim_{G'} y') \text{ or} \\ &(x \sim_G y \text{ and } x' \sim_{G'_0} y').\end{aligned} \tag{5.34}$$

And the graph G_0'', defined analogously to (5.33), has edges

$$\begin{aligned}(x,x') \sim_{G_0''} (y,y') &\iff \exists u,u',\exists i,i' \text{ s.t.} \\ &\quad P''(x,x',u,u'|i,i')P''(y,y',u,u'|i,i') \neq 0 \\ &\iff x \sim_{G_0} y \text{ and } x' \sim_{G_0'} y'. \end{aligned} \tag{5.35}$$

We introduce the notation $(G'',G_0'') = (G,G_0)\boxtimes(G',G_0')$ as shorthand for (5.34)-(5.35). By induction, m parallel instances of a source yields a characteristic graph $(G,G_0)^{\boxtimes m} := (G,G_0)\boxtimes(G,G_0)\boxtimes\cdots\boxtimes(G,G_0)$.

For convenience we will abuse notation by treating these ordered pairs as being graphs themselves. For instance, $(G,G_0)^{\boxtimes m} \to H$ will be taken to mean $G' \to H$ where $(G',G_0') = (G,G_0)^{\boxtimes m}$; similarly $\bar{\vartheta}((G,G_0)^{\boxtimes m})$ will be taken to mean $\bar{\vartheta}(G')$ and $(G,G_0)^{\boxtimes m} \supseteq G^{\boxtimes m}$ to mean $G' \supseteq G^{\boxtimes m}$.

Now we can show that the condition $P(x,u|i) \neq 0$ only when $x = i$ can be dropped in proposition 5.5. We will later generalize this to quantum sources and quantum channels.

Proposition 5.21. *Let $P(x,u|i)$ be a classical source and $\mathcal{N}(v|s)$ a classical channel. Let graphs G and H be given by (5.2) and (5.1). Then m parallel instances of the source can be sent using n parallel instances of the channel only if*

$$\frac{n}{m} \geq \frac{\log \bar{\vartheta}(G)}{\log \bar{\vartheta}(H)}.$$

Proof. Without loss of generality assume that each x is possible. In other words, assume that for each x there is an i and u such that $P(x,u|i) \neq 0$. Generality is not lost because one can decrease the alphabet associated with x, removing values that can never occur. Reducing this alphabet only removes isolated vertices from G, and so doesn't affect the value of $\bar{\vartheta}(G)$. Let G_0 be defined as in (5.33). Since each x is possible, this graph has loops on all vertices: $x \sim_{G_0} x$ for all x.

As per the above discussion, the composite source (consisting of m parallel instances of $P(x,u|i)$) will have characteristic graph $(G,G_0)^{\boxtimes m}$. Since G_0 has loops on all vertices, our generalized strong product (5.34) has at least as many edges as the standard strong product (5.32). Since $\bar{\vartheta}$ is monotone increasing under addition of edges and is multiplicative under the strong product [Knu94] we have $\bar{\vartheta}((G,G_0)^{\boxtimes m}) \geq \bar{\vartheta}(G^{\boxtimes m}) = \bar{\vartheta}(G)^m$.

The distinguishability graph of n parallel instances of the channel $\mathcal{N}(v|s)$ is H^{*n}. Since $\bar{\vartheta}$ is multiplicative under the disjunctive product [Lov79] we have $\bar{\vartheta}(H^{*n}) = \bar{\vartheta}(H)^n$. If m parallel sources can be sent via n parallel channels then $(G,G_0)^{\boxtimes m} \to H^{*n}$. Since $\bar{\vartheta}$ is monotone under homomorphisms,

$$\begin{aligned}(G,G_0)^{\boxtimes m} \to H^{*n} &\implies \bar{\vartheta}((G,G_0)^{\boxtimes m}) \leq \bar{\vartheta}(H^{*n}) \\ &\implies \bar{\vartheta}(G)^m \leq \bar{\vartheta}(H)^n \\ &\implies \frac{n}{m} \geq \frac{\log \bar{\vartheta}(G)}{\log \bar{\vartheta}(H)}.\end{aligned}$$

□

Similar arguments apply for quantum source-channel coding. It is easy to see that the confusability graphs for parallel channels should combine by tensor product since the Kraus operators combine by tensor product. We have been using instead the distinguishability graph, which then combines as $(S^\perp \otimes T^\perp)^\perp$. We take this as the definition of disjunctive product:

Definition 5.22. *Let $S \subseteq \mathcal{L}(A)$ and $T \subseteq \mathcal{L}(B)$ be non-commutative graphs. Their* disjunctive product *is $S * T = S \otimes \mathcal{L}(B) + \mathcal{L}(A) \otimes T = (S^\perp \otimes T^\perp)^\perp$.*

When S and T derive from classical graphs this definition is equivalent to (5.30). We will use the notation $S^{*n} := S * S * \ldots * S$. Analogous to (5.31), the Shannon capacity of a quantum channel with distinguishability graph T is

$$C_0(T^\perp) = \lim_{n\to\infty} \frac{1}{n} \log \omega(T^{*n})$$

It is known that $\overline{\vartheta}(T)$ is an upper bound on $C_0(T^\perp)$, since $\overline{\vartheta}(T^{*n}) = \tilde{\vartheta}((T^\perp)^{\otimes n}) = \overline{\vartheta}(T)^n$ [DSW13].

Consider now two parallel sources, with characteristic graphs S and S'. Analogous to (5.33) we define (S, S_0), a *non-commutative graph with generalized loops*. For discrete QSSC, the subject of theorem 5.14, define

$$\begin{aligned} S &= \mathrm{Tr}_{BC}\{\mathcal{L}(C) J K_r J^\dagger\} \\ S_0 &= \mathrm{Tr}_{BC}\{\mathcal{L}(C) J K_r^\perp J^\dagger\} \end{aligned} \tag{5.36}$$

and for coherent QSSC, the subject of theorem 5.16, define

$$\begin{aligned} S &= \mathrm{Tr}_{BC}\{\mathcal{L}(C) J Q_r J^\dagger\} \\ S_0 &= \mathrm{Tr}_{BC}\{\mathcal{L}(C) J Q_r^\perp J^\dagger\} \\ &= \mathrm{Tr}_{BC}\{\mathcal{L}(C) J J^\dagger\}. \end{aligned} \tag{5.37}$$

Analogous to (5.34)-(5.35) define the strong product $(S'', S_0'') = (S, S_0) \boxtimes (S', S_0')$ where

$$\begin{aligned} S'' &= (S \otimes S') + (S_0 \otimes S') + (S \otimes S_0') \\ S_0'' &= S_0 \otimes S_0'. \end{aligned} \tag{5.38}$$

If S, S_0, S', S_0' correspond to classical graphs G, G_0, G', G_0' then this product corresponds to the classical graph $(G, G_0) \boxtimes (G', G_0')$. If G_0 and G_0' consist of only loops on each vertex (i.e. $S_0 = \mathrm{span}\{|x\rangle\langle x|\}$ and similarly for S_0') then this corresponds to $G \boxtimes G'$. Define the graph power $(S, S_0)^{\boxtimes m}$ to be repeated application of (5.38).

Other graph products could be defined similarly. For example, the Cartesian product of graphs, $G \,\square\, G'$ is defined to have edges $(x, x') \sim (y, y') \iff (x = y \wedge x' \sim y') \vee (x \sim y \wedge x' = y')$, so for non-commutative graphs one could define $(S'', S_0'') = (S, S_0) \,\square\, (S', S_0')$ with $S'' = (S_0 \otimes S') + (S \otimes S_0')$ and $S_0'' = S_0 \otimes S_0'$. The complement of a graph has edges $x \not\sim y \wedge x \neq y$, which would have non-commutative analogue $\overline{(S, S_0)} = (S^\perp \setminus S_0, S_0)$, assuming $S_0 \subseteq S^\perp$. We will not have occasion to consider such constructions, but mention it as a starting point for possible development of a richer theory of non-commutative graphs. A similar concept was explored in [DSW10]; however, they suggested a specific form of S_0 in terms of the multiplicative domain of a channel whereas we leave the form of S_0 to be determined by the application at hand.

As before, we abuse notation and take $(S, S_0)^{\boxtimes m} \to T$ to mean $S' \to T$ where $(S', S_0') = (S, S_0)^{\boxtimes m}$, and $\bar{\vartheta}((S, S_0)^{\boxtimes m})$ to mean $\bar{\vartheta}(S')$. The strong product (5.38) indeed corresponds to the characteristic graph of parallel sources:

Theorem 5.23. *Consider discrete QSSC with two parallel sources $\{|\psi_i\rangle\}_i$ and $\{|\psi'_{i'}\rangle\}_{i'}$. Let $J : R \to A \otimes B \otimes C$ and $J' : R' \to A' \otimes B' \otimes C'$ be the isometries corresponding to these sources, as in theorem 5.14. Let (S, S_0) and (S', S_0') be the characteristic graphs (with generalized loops) for these two sources, as defined by (5.36), and similarly (S'', S_0'') for the joint source $\{|\psi_i\rangle \otimes |\psi'_{i'}\rangle\}_{i,i'}$. Then it holds that $(S'', S_0'') = (S, S_0) \boxtimes (S', S_0')$. These two sources can be sent using one copy of the channel $\mathcal{N}$ iff*

$$(S, S_0) \boxtimes (S', S_0') \to T \tag{5.39}$$

where $T = (N^\dagger N)^\perp$.

The analogous statement holds for coherent QSSC, where now (S, S_0), (S', S_0'), *and* (S'', S_0'') *are defined using* (5.37) *rather than* (5.36).

In either case (discrete or coherent QSSC), it is possible to send m *copies of a source using* n *copies of a channel iff*

$$(S, S_0)^{\boxtimes m} \to T^{*n}. \tag{5.40}$$

Proof. We give the proof only for discrete QSSC; the proof for coherent QSSC follows from similar arguments. A state from the joint source will be of the form $|\psi''_{ii'}\rangle = |\psi_i\rangle \otimes |\psi'_{i'}\rangle$ and the corresponding isometry will be $J'' = J \otimes J'$, so we have (according to (5.36))

$$S'' = \mathrm{Tr}_{BB'CC'}\{\mathcal{L}(C \otimes C')J''K_{r''}J''^{\dagger}\}$$
$$S_0'' = \mathrm{Tr}_{BB'CC'}\{\mathcal{L}(C \otimes C')J''K_{r''}^{\perp}J''^{\dagger}\}$$

where $r'' = rr' = \dim(R)\dim(R')$. It is readily verified that $(S'', S_0'') = (S, S_0) \boxtimes (S', S_0')$, since

$$K_{r''} = K_r \otimes K_{r'} + K_r^{\perp} \otimes K_{r'} + K_r \otimes K_{r'}^{\perp},$$
$$K_{r''}^{\perp} = K_r^{\perp} \otimes K_{r'}^{\perp}.$$

By theorem 5.14, the joint source can be sent using a single use of channel $\mathcal{N}$ iff $S'' \to T$, that is to say iff condition (5.39) holds.

By induction, m instances of a source can be sent with a single channel use iff $(S, S_0)^{\boxtimes m} \to T$. Since the distinguishability graph of n copies of the channel is T^{*n}, it is possible to send m instances of the source using n instances of the channel iff $(S, S_0)^{\boxtimes m} \to T^{*n}$. □

For classical source-channel coding the amount of communication needed to transmit a joint source is at least as much as is needed for each individual source, since the second source can always be simulated: Alice and Bob can just agree ahead of time on some x' and u' that can be emitted be the second source. Somewhat surprisingly, this does not necessarily hold for quantum source-channel coding. For example, consider the following two sources. The first source is some classical source for which an entanglement resource $|\lambda\rangle\langle\lambda|$ would allow for more efficient transmission. In other words, $\chi(S)$ is large and $\chi_*(S)$ is small. Examples of such graphs are given in, e.g. [AHKS06]. The second source consists of only a single possible input: $|\lambda\rangle\langle\lambda|$. So $S' = \varnothing$ and $S_0' = \mathbb{C}\Lambda$ where $\Lambda = \mathrm{Tr}_B\{|\lambda\rangle\langle\lambda|\}$. Then the first source requires an amount of communication $\chi(S)$, the second requires no communication (i.e. $\chi(S') = 1$), but the joint source requires communication $\chi_*(S) < \max\{\chi(S), \chi(S')\}$.

Entanglement assisted chromatic number does not exhibit this same anomaly. Indeed, the joint source can never be easier to transmit than either of the individual sources since Alice and Bob can always simulate (some particular input from) the second source, by choosing said state ahead of time and adding this to their entanglement resource. For a similar reason, even without entanglement assistance a joint source is not easier to transmit than the individual sources in the case where the individual sources are each capable of producing a product state: Alice and Bob can simulate any of these sources by producing the product state themselves, in order to turn a single source into (a subset of) the joint source.

For classical source-channel coding, we defined the cost rate as the infimum of n/m such that m instances of the source can be transmitted using n instances of the channel. As per the above discussion, this can be achieved iff $(G, G_0)^{\boxtimes m} \to H^{*n}$, so the cost rate is

$$\lim_{m\to\infty} \frac{1}{m} \min\left\{n : (G, G_0)^{\boxtimes m} \to H^{*n}\right\}.$$

Cost rate for quantum source-channel coding can be defined similarly,

$$\lim_{m\to\infty} \frac{1}{m} \min\left\{n : (S, S_0)^{\boxtimes m} \to T^{*n}\right\}, \tag{5.41}$$

where (S, S_0) is the characteristic graph of the source (as per (5.36) or (5.37)) and T is the distinguishability graph of the channel. Similarly, the entanglement assisted cost rate is

$$\lim_{m\to\infty} \frac{1}{m} \min\left\{n : (S, S_0)^{\boxtimes m} \xrightarrow{*} T^{*n}\right\}. \tag{5.42}$$

Clearly (5.42) $\leq$ (5.41). For the classical case, the Lovász $\bar{\vartheta}$ number is multiplicative under the relevant graph products and is a homomorphism monotone, so it leads to a lower bound on the cost rate, proposition 5.21. A similar bound applies for quantum source-channel coding, with a caveat. The $\overline{\vartheta}$ quantity is not multiplicative under strong product in general; however, it is when S_0 and S_0' contain the identity. So our generalization of proposition 5.21 will require $I \in S_0$. This happens for example when the states emitted by the source include a maximally entangled state, or product states with Alice's shares forming a complete orthonormal basis (such as is the case with classical source-channel coding). We have then the following bound on cost rate.

Theorem 5.24. *Consider a source with characteristic graph (S, S_0), defined as in (5.36) for discrete QSSC or as in (5.37) for coherent QSSC. Consider a noisy quantum channel $\mathcal{N}$ with distinguishability graph $T = (N^\dagger N)^\perp$. If $I \in S_0$ then the entanglement assisted cost rate (5.42) is lower bounded by $\log \overline{\vartheta}(S)/\log \overline{\vartheta}(T)$.*

Proof. Since $I \in S_0$, the $\overline{\vartheta}$ quantity is multiplicative under both strong and disjunctive graph powers, by lemma 5.25. Using this fact, and the fact that $\overline{\vartheta}$ is monotone under entanglement assisted homomorphisms, we have

$$\begin{aligned}
(5.42) &\geq \lim_{m\to\infty} \frac{1}{m} \min\left\{n : \overline{\vartheta}((S, S_0)^{\boxtimes m}) \leq \overline{\vartheta}(T^{*n})\right\} \\
&= \lim_{m\to\infty} \frac{1}{m} \min\left\{n : \overline{\vartheta}(S)^m \leq \overline{\vartheta}(T)^n\right\} \\
&= \lim_{m\to\infty} \frac{1}{m} \min\left\{n : \log \overline{\vartheta}(S)/\log \overline{\vartheta}(T) \leq n/m\right\} \\
&= \log \overline{\vartheta}(S)/\log \overline{\vartheta}(T).
\end{aligned}$$

□

We now prove the lemma used in the preceding proof.

Lemma 5.25. *Let S and S' be trace-free non-commutative graphs. Then,*

- $\overline{\vartheta}(S * S') = \overline{\vartheta}(S)\overline{\vartheta}(S')$
- $\overline{\vartheta}((S, S_0) \boxtimes (S', S_0')) = \overline{\vartheta}(S)\overline{\vartheta}(S')$ *if $I \in S_0$ and $I \in S_0'$*

Proof. From [DSW13] we have $\tilde{\vartheta}(S^\perp \otimes S'^\perp) = \tilde{\vartheta}(S^\perp)\tilde{\vartheta}(S'^\perp)$. But $(S^\perp \otimes S'^\perp)^\perp = S * S'$, so $\overline{\vartheta}(S * S') = \overline{\vartheta}(S)\overline{\vartheta}(S')$. Since $(S, S_0) \boxtimes (S', S_0') \subseteq S * S'$ and since $\overline{\vartheta}$ is monotone decreasing under subsets, we have

$$\overline{\vartheta}((S, S_0) \boxtimes (S', S_0')) \leq \overline{\vartheta}(S * S') = \overline{\vartheta}(S)\overline{\vartheta}(S').$$

Let X be an optimal solution to (5.10) for $\overline{\vartheta}(S)$, from definition 5.6. Then $X \in S \otimes \mathcal{L}(B)$ (for some Hilbert space B), $I + X \succeq 0$, and $\|I + X\| = \overline{\vartheta}(S)$. Similarly, there is an $X' \in S' \otimes \mathcal{L}(B')$, $I + X' \succeq 0$, and $\|I + X'\| = \overline{\vartheta}(S')$. Define

$$X'' = (I_{AB} + X) \otimes (I_{A'B'} + X') - I_{AA'BB'}.$$

Clearly $I + X'' \succeq 0$. Also,

$$\begin{aligned} X'' &= X \otimes X' + I_{AB} \otimes X' + X \otimes I_{A'B'} \\ &\in [S \otimes S' + I_A \otimes S' + S \otimes I_{A'}] \otimes \mathcal{L}(BB') \\ &\subseteq [S \otimes S' + S_0 \otimes S' + S \otimes S_0'] \otimes \mathcal{L}(BB') \\ &= [(S, S_0) \boxtimes (S', S_0')] \otimes \mathcal{L}(BB'). \end{aligned}$$

So X'' is feasible for (5.10) for $\overline{\vartheta}((S, S_0) \boxtimes (S', S_0'))$. Therefore

$$\begin{aligned} \overline{\vartheta}((S, S_0) \boxtimes (S', S_0')) &\geq \|I + X''\| \\ &= \|(I + X) \otimes (I + X')\| \\ &= \overline{\vartheta}(S)\overline{\vartheta}(S'). \end{aligned}$$

□

5.8 Schrijver and Szegedy

In this section we will provide a generalization to non-commutative graphs for two quantities related to Lovász's ϑ: Schrijver's ϑ' and Szegedy's ϑ^+. Schrijver's number comes from adding extra constraints to the maximization program for ϑ, yielding a smaller value; Szegedy's number comes from adding extra constraints to the minimization (dual) program for ϑ, yielding a larger value. We will consider the complimentary quantities $\bar{\vartheta}'(G) = \vartheta'(\overline{G})$ and $\bar{\vartheta}^+(G) = \vartheta^+(\overline{G})$. These are homomorphism monotones in the same sense that $\bar{\vartheta}$ is [dCST13]; therefore they satisfy the sandwich theorem

$$\omega(G) \leq \bar{\vartheta}'(G) \leq \bar{\vartheta}(G) \leq \bar{\vartheta}^+(G) \leq \chi(G).$$

These quantities are not suitable for bounding asymptotic channel capacity or cost rate for source-channel coding because they are not multiplicative under the appropriate graph products [CMR+13].

For classical graphs these quantities have been shown to be monotone under entanglement assisted homomorphisms [CMR+13]. Strangely enough, our generalization to non-commutative graphs will yield quantities monotone under homomorphisms but not under entanglement assisted homomorphisms. For classical graphs the gap between $\bar{\vartheta}'(G)$ and $\bar{\vartheta}^+(G)$ tends to be small or, often, zero. For non-commutative graphs the gap tends to be much more extreme, sometimes infinite, even for random graphs of small dimension. After developing basic properties of these quantities we will show how they can be used to reproduce some results from the literature regarding entanglement assisted activation of one-shot channel capacity and impossibility of one-way LOCC measurement of entangled states. Also we will provide a channel for which maximally entangled states are not sufficient for achieving the entanglement assisted one-shot capacity.

The classical quantities are defined as follows [Lov79, Knu94, Sch79, MRRJ78, Sze94].

Definition 5.26. *The Lovász, Schrijver, and Szegedy numbers of the complement of a graph, $\bar{\vartheta}(G)$, $\bar{\vartheta}'(G)$, and $\bar{\vartheta}^+(G)$, are defined by the following dual (and equivalent) semidefinite programs. All matrices are either real or complex (it doesn't matter), J is the all-ones matrix, and $\mathcal{N}$ is the cone of symmetric entrywise non-negative matrices. Take $S = span\{|x\rangle\langle y| : x \sim y\}$ and $S_0 = span\{|x\rangle\langle x| : x \in V(G)\}$ (the diagonal matrices).*

$$\bar{\vartheta}(G) = \max\langle B, J\rangle \text{ s.t. } B \succeq 0, TrB = 1, B \in S + S_0 \tag{5.43}$$

$$\bar{\vartheta}'(G) = \max\langle B, J\rangle \text{ s.t. } B \succeq 0, TrB = 1, B \in S + S_0, B \in \mathcal{N} \tag{5.44}$$

$$\bar{\vartheta}^+(G) = \max\langle B, J\rangle \text{ s.t. } B \succeq 0, TrB = 1, B + L \in S + S_0, L \in \mathcal{N} \tag{5.45}$$

$$\bar{\vartheta}(G) = \min \lambda \text{ s.t. } Z \succeq J, (Z_{ii} = \lambda \text{ for all } i), Z \in S^\perp \tag{5.46}$$

$$\bar{\vartheta}'(G) = \min \lambda \text{ s.t. } Z \succeq J, (Z_{ii} = \lambda \text{ for all } i), Z + L \in S^\perp, L \in \mathcal{N} \tag{5.47}$$

$$\bar{\vartheta}^+(G) = \min \lambda \text{ s.t. } Z \succeq J, (Z_{ii} = \lambda \text{ for all } i), Z \in S^\perp, Z \in \mathcal{N} \tag{5.48}$$

The constraint $B \in \mathcal{N}$ that is added to (5.43) to yield (5.44) has the following justification. Suppose that $W \subseteq V(G)$ is a clique. Then the matrix

$$B_{ij} = \begin{cases} 1/|W| & \text{if } i,j \in W \\ 0 & \text{otherwise}\} \end{cases}$$

is a feasible solution to (5.43) with value $|W|$. So $\omega(G) \leq \bar{\vartheta}(G)$. But $B \in \mathcal{N}$, so this condition can be added to the maximization program to yield a potentially smaller quantity $\bar{\vartheta}'(G)$ that still upper bounds $\omega(G)$.[7] Similarly, if $f : V(G) \to \{1, \ldots, m\}$ is a proper coloring of G then

$$Z_{ij} = \begin{cases} m & \text{if } f(i) = f(j) \\ 0 & \text{otherwise} \end{cases}$$

is feasible for (5.46) with value $\chi(G)$, so $\bar{\vartheta}(G) \geq \chi(G)$. Since this satisfies $Z \in \mathcal{N}$, adding this condition to the minimization program gives a quantity $\bar{\vartheta}^+(G)$ still lower bounding $\chi(G)$. We will follow this sort of strategy to create analogues of $\bar{\vartheta}'$ and $\bar{\vartheta}^+$ for non-commutative graphs.

The primal program for $\overline{\vartheta}$ can be written [DSW13]

$$\begin{aligned} \overline{\vartheta}(S) = \max \ & \langle \Phi | I \otimes \rho + T | \Phi \rangle \, 0 \\ \text{s.t.} \ & \rho \succeq 0, \mathrm{Tr}\rho = 1, \\ & I \otimes \rho + T \succeq 0, \\ & T \in S \otimes \mathcal{L}(A'), \end{aligned} \tag{5.49}$$

where A' is an ancillary system of the same dimension as A and $|\Phi\rangle = \sum_i |i\rangle_A \otimes |i\rangle_{A'}$. With this definition it is easy to see that $\omega(S) \leq \overline{\vartheta}(S)$: since $\omega(S)$ is the classical communication capacity of the distinguishability graph S, there are $m = \omega(S)$ vectors $|\psi_1\rangle, \ldots, |\psi_m\rangle \in A$ such that $|\psi_i\rangle \langle\psi_j| \in S$ for $i \neq j$. Define

$$\begin{aligned} T &= \frac{1}{m} \sum_{i \neq j} |\psi_i\rangle \langle\psi_j|_A \otimes |\overline{\psi_i}\rangle \langle\overline{\psi_j}|_{A'} \ \text{ and} \\ \rho &= \frac{1}{m} \sum_i |\psi_i\rangle \langle\psi_i|_{A'} , \end{aligned} \tag{5.50}$$

where a bar over a vector represents complex conjugation in the computational basis. This is readily verified to be a feasible solution to (5.49) with value m. A tighter upper bound on $\omega(S)$ can be obtained by adding constraints to (5.49). As long as (5.50) remains feasible under these new constraints, the modified program will remain an upper bound on $\omega(S)$.

To this end, consider the "rotated transpose" linear superoperator $\mathcal{R} : \mathcal{L}(A) \otimes \mathcal{L}(A') \to \mathcal{L}(A) \otimes \mathcal{L}(A')$ with action

$$\begin{aligned} \mathcal{R}\left(|i\rangle \langle j|_A \otimes |k\rangle \langle l|_{A'}\right) &= |i\rangle \langle k|_A \otimes |j\rangle \langle l|_{A'} && \text{(on standard basis states)} \\ \mathcal{R}\left(|\psi\rangle \langle\phi|_A \otimes |\chi\rangle \langle\xi|_{A'}\right) &= |\psi\rangle \langle\overline{\chi}|_A \otimes |\overline{\phi}\rangle \langle\xi|_{A'} . && \text{(in general)} \end{aligned}$$

Note that $\mathcal{R}$ is an involution (it is its own inverse). Define the double-dagger operation

$$X^{\ddagger} = \mathcal{R}(\mathcal{R}(X)^{\dagger}).$$

We have $\mathcal{R}(I_A \otimes I_{A'}) = |\Phi\rangle \langle\Phi|$. The T from (5.50) transforms as

$$\mathcal{R}(T) = \frac{1}{m} \sum_{i \neq j} |\psi_i\rangle \langle\psi_i| \otimes |\overline{\psi_j}\rangle \langle\overline{\psi_j}|.$$

[7] An even tighter constraint, requiring B to be completely positive, yields $\omega(G)$ exactly [dKP02].

Since $\mathcal{R}(T)$ is a separable operator, we may add this as an extra constraint in (5.49) to get a tighter bound on $\omega(S)$.

In general, consider some closed convex cone $\mathcal{C} \subseteq \mathcal{L}(A) \otimes \mathcal{L}(A')$ and a trace-free non-commutative graph S. We consider only cones over the real inner product space of Hermitian matrices. For $S \in \mathcal{L}(A)$, we use the notation $\overline{S} := \{\overline{M} : M \in S\} \subseteq \mathcal{L}(A')$, where a bar over an operator denotes entrywise complex conjugate, with the conjugated operator moved into the primed space (as discussed in section 5.6). Define the semidefinite program

$$\begin{aligned} \overline{\vartheta}'_{\mathcal{C}}(S) = \max \ & \langle \Phi | I \otimes \rho + T | \Phi \rangle \\ \text{s.t.} \ & \rho \succeq 0, \mathrm{Tr}\rho = 1, \\ & I \otimes \rho + T \succeq 0, \\ & T \in S \otimes \overline{S}, \\ & \mathcal{R}(T) \in \mathcal{C}. \end{aligned} \tag{5.51}$$

Note that $T \in S \otimes \mathcal{L}(A')$ and $\mathcal{R}(T) \in \mathcal{C}$ implies $T \in S \otimes \overline{S}$, since $\mathcal{C}$ contains only Hermitian operators. We choose to explicitly state the condition $T \in S \otimes \overline{S}$ in (5.51).

Since linear programming duality turns constraints into variables, the dual of this program is similar to (5.11) but with an extra variable that runs over the dual cone $\mathcal{C}^*$. In appendix 5.A we show that strong duality holds, so that primal and dual have equal value. The dual program is

$$\begin{aligned} \overline{\vartheta}'_{\mathcal{C}}(S) = \min \ & \|\mathrm{Tr}_A Y\| \\ \text{s.t.} \ & Y + (L + L^\dagger) \in S^\perp * \overline{S}^\perp = (S \otimes \overline{S})^\perp, \\ & \mathcal{R}(L) + \mathcal{R}(L)^\dagger \in \mathcal{C}^*, \\ & Y \succeq |\Phi\rangle \langle \Phi|, \\ & L \in \mathcal{L}(A) \otimes \mathcal{L}(A'). \end{aligned} \tag{5.52}$$

Recall that "$*$" denotes the disjunctive product from definition 5.22. The point $\rho = I/\dim(A)$, $T = 0$ is feasible for (5.51), giving $\overline{\vartheta}'_{\mathcal{C}}(S) \geq 1$. In appendix 5.A we provide a feasible point for (5.52), giving $\overline{\vartheta}'_{\mathcal{C}}(S) < \infty$.

Denote by SEP the cone of separable operators in $\mathcal{L}(A) \otimes \mathcal{L}(A')$. Since (5.50) satisfies $\mathcal{R}(T) \in \mathrm{SEP}$, it is feasible for (5.51) for $\overline{\vartheta}'_{\mathrm{SEP}}$. Therefore $\omega(S) \leq \overline{\vartheta}'_{\mathrm{SEP}}(S)$. One can also show $\omega_q(S)^2 \leq \overline{\vartheta}'_{\mathrm{SEP}}(S)$ by similar means, but we will eventually obtain this result by showing $\overline{\vartheta}'_{\mathrm{SEP}}$ to be a homomorphism monotone in the same sense that $\overline{\vartheta}$ is.

From a computational perspective $\overline{\vartheta}'_{\mathrm{SEP}}(S)$ is not the most convenient because there is no efficient way to determine whether an operator is in SEP. Fortunately there are closed convex cones containing SEP that are efficiently optimized over and that give good bounds on $\omega(S)$ and $\omega_q(S)$. Namely, consider $\mathcal{S}^+$, the cone of positive semidefinite matrices, PPT, the cone of matrices with positive semidefinite partial transpose, or even $\mathcal{S}^+ \cap \mathrm{PPT}$. Note that $\mathcal{S}^+$ and PPT are self-dual and the dual of $\mathcal{S}^+ \cap \mathrm{PPT}$ is $\mathcal{S}^+ + \mathrm{PPT}$. The dual of SEP is the set of entanglement witnesses: $\mathrm{SEP}^* = \{W : \langle W, M \rangle \geq 0 \text{ for all } M \in \mathrm{SEP}\}$. We have

$$\omega(S) \leq \overline{\vartheta}'_{\mathrm{SEP}}(S) \leq \overline{\vartheta}'_{\mathcal{S}^+ \cap \mathrm{PPT}}(S) \leq \overline{\vartheta}'_{\mathcal{S}^+}(S) \leq \overline{\vartheta}(S). \tag{5.53}$$

This sequence of refinements is reminiscent of the approximations to the copositive cone that yield the Lovász and Schrijver numbers for classical graphs [dKP02, BFL10]. In fact the middle three values in the above chain of inequalities collapse to Schrijver's number when S derives from a classical graph.

Theorem 5.27. *Let G be a classical loop-free graph and $S = \mathrm{span}\{|i\rangle \langle j| : i \sim j\}$. Then for any closed convex cone $\mathcal{C}$ satisfying $\mathrm{SEP} \subseteq \mathcal{C} \subseteq \mathrm{SEP}^*$, it holds that $\overline{\vartheta}'_{\mathcal{C}}(S) = \overline{\vartheta}'(G)$.*

Proof. Define the isometry $V = \sum_i |ii\rangle \langle i|$. Let T and ρ be an optimal solution for (5.51) for $\overline{\vartheta}'_{\mathrm{SEP}^*}(S)$. We will show that $B = V^\dagger(I \otimes \rho + T)V$ is feasible for (5.44). This matrix has coefficients

$$\begin{aligned} B_{ij} &= \rho_{ii}\delta_{ij} + \langle ii|T|jj\rangle \\ &= \rho_{ii}\delta_{ij} + \langle ij|\mathcal{R}(T)|ij\rangle\,. \end{aligned}$$

Since $\rho_{ii} \geq 0$, $\mathcal{R}(T) \in \mathrm{SEP}^*$, and $|ij\rangle\langle ij| \in \mathrm{SEP}$, it holds that $B_{ij} \geq 0$ for all i, j. So $B \in \mathcal{N}$. We have $I \otimes \rho + T \succeq 0 \implies B \succeq 0$. Since $T \in S \otimes \overline{S}$ we have $\langle ii|T|jj\rangle = 0$ when $i \not\sim j$. In particular, $B_{ii} = \rho_{ii}$ and $B_{ij} = 0$ when $i \not\sim j$, $i \neq j$. Since $\mathrm{Tr}\rho = 1$, also $\mathrm{Tr}B = 1$. So B is feasible for (5.44). Its value is

$$\langle B, J\rangle = \sum_{ij} B_{ij} = \sum_{ij} \langle ii|I \otimes \rho + T|jj\rangle = \langle \Phi|I \otimes \rho + T|\Phi\rangle = \overline{\vartheta}'_{\mathrm{SEP}^*}(S).$$

Therefore $\bar{\vartheta}'(G) \geq \overline{\vartheta}'_{\mathrm{SEP}^*}(S)$.

Now let B be an optimal solution for (5.44). Decompose this into diagonal and off-diagonal components: $B = \rho + T'$. Define $T = VT'V^\dagger$. We will show these to be feasible for (5.51) for $\overline{\vartheta}'_{\mathrm{SEP}}(S)$. For any vector $|\psi\rangle \in A \otimes A'$ we have

$$\begin{aligned} \langle \psi|I \otimes \rho + T|\psi\rangle &= \sum_{ij} \left|\psi_{ij}^2\right| \rho_{jj} + \sum_{i \neq j} \psi_{ii}^* T'_{ij} \psi_{jj} \\ &= \sum_{i \neq j} \left|\psi_{ij}^2\right| \rho_{jj} + \sum_{ij} \psi_{ii}^* B_{ij} \psi_{jj} \geq 0, \end{aligned}$$

where the last inequality follows from $\rho_{jj} \geq 0$ and $B \succeq 0$. Therefore $I \otimes \rho + T \succeq 0$. We have

$$\begin{aligned} \mathcal{R}(T) &= \sum_{ij} T'_{ij} \mathcal{R}(|ii\rangle\langle jj|) \\ &= \sum_{ij} T'_{ij} |ij\rangle\langle ij| \in \mathrm{SEP}, \end{aligned}$$

where the last relation requires $T'_{ij} \geq 0$. Clearly $\rho \succeq 0$ and $\mathrm{Tr}\rho = \mathrm{Tr}B = 1$. For $i \not\sim j$ we have $T'_{ij} = 0 \implies (\langle i| \otimes I)T(|j\rangle \otimes I) = 0$, giving $T \in S \otimes \mathcal{L}(A')$. Similarly, $T \in \mathcal{L}(A) \otimes \overline{S}$. So in fact $T \in (S \otimes \mathcal{L}(A')) \cap (\mathcal{L}(A) \otimes \overline{S}) = S \otimes \overline{S}$. Therefore ρ and T are feasible for (5.51) for $\overline{\vartheta}'_{\mathrm{SEP}}(S)$. This solution has value

$$\begin{aligned} \langle \Phi|I \otimes \rho + T|\Phi\rangle &= \sum_{ij} \langle ii|I \otimes \rho + T|jj\rangle \\ &= \sum_i \rho_{ii} + \sum_{ij} T'_{ij} = \langle B, J\rangle = \bar{\vartheta}'(G), \end{aligned}$$

giving $\overline{\vartheta}'_{\mathrm{SEP}}(S) \geq \bar{\vartheta}'(G)$.

Clearly $\mathrm{SEP} \subseteq \mathcal{C} \subseteq \mathrm{SEP}^* \implies \overline{\vartheta}'_{\mathrm{SEP}}(S) \leq \overline{\vartheta}'_{\mathcal{C}}(S) \leq \overline{\vartheta}'_{\mathrm{SEP}^*}(S)$ since maximization programs have nondecreasing value as constraints are loosened. Combining this with the above two inequalities gives the desired equality result. □

A generalization of Szegedy's number to non-commutative graphs follows similarly, now adding extra constraints to the dual program (5.11). Extra constraints on the dual become extra variables in the primal. For a closed convex cone $\mathcal{C}$ of operators in $\mathcal{L}(A) \otimes \mathcal{L}(A')$ and for a trace-free

non-commutative graph S, the primal and dual take the form

$$\begin{aligned}
\overline{\vartheta}^+_{\mathcal{C}}(S) = \max\ & \langle\Phi|I\otimes\rho + T|\Phi\rangle \\
\text{s.t.}\ & \rho \succeq 0, \mathrm{Tr}\rho = 1, \\
& I\otimes\rho + T \succeq 0, \\
& T + (L + L^\dagger) \in S * \overline{S} = (S^\perp \otimes \overline{S}^\perp)^\perp, \\
& \mathcal{R}(L) + \mathcal{R}(L)^\dagger \in \mathcal{C}^*, \\
& L \in \mathcal{L}(A)\otimes\mathcal{L}(A'),
\end{aligned} \tag{5.54}$$

$$\begin{aligned}
\overline{\vartheta}^+_{\mathcal{C}}(S) = \min\ & \|\mathrm{Tr}_A Y\| \\
\text{s.t.}\ & Y \in S^\perp \otimes \overline{S}^\perp, \\
& \mathcal{R}(Y) \in \mathcal{C}, \\
& Y \succeq |\Phi\rangle\langle\Phi|.
\end{aligned} \tag{5.55}$$

That these two programs take the same value is shown in appendix 5.A. The point $\rho = I/\dim(A)$, $T = 0, L = 0$ is feasible for (5.54), giving $\overline{\vartheta}^+_{\mathcal{C}}(S) \geq 1$. Although (5.52) is always feasible, in some cases (5.55) is not feasible so $\overline{\vartheta}^+_{\mathcal{C}}(S)$ can be infinite; see lemma 5.30 for an example.

Similar to (5.53), we have the chain of inequalities

$$\overline{\vartheta}(S) \leq \overline{\vartheta}^+_{\mathcal{S}^+}(S) \leq \overline{\vartheta}^+_{\mathcal{S}^+\cap\mathrm{PPT}}(S) \leq \overline{\vartheta}^+_{\mathrm{SEP}}(S) \leq \chi(S). \tag{5.56}$$

The last inequality will be proved in corollary 5.31, and the others follow from the fact that (5.55) has nondecreasing value as constraints are tightened. Note, however, that the last two values can be ∞ and, unlike $\overline{\vartheta}(S)$, don't provide a bound on $\chi_q(S)^2$. As was the case with our Schrijver generalization, this generalized Szegedy quantity matches the classical value when S derives from a classical graph.

Theorem 5.28. *Let G be a classical loop-free graph and $S = \mathrm{span}\{|i\rangle\langle j| : i \sim j\}$. Then for any closed convex cone $\mathcal{C}$ satisfying $\mathrm{SEP} \subseteq \mathcal{C} \subseteq \mathrm{SEP}^*$, it holds that $\overline{\vartheta}^+_{\mathcal{C}}(S) = \bar{\vartheta}^+(G)$.*

Proof. Define the isometry $V = \sum_i |ii\rangle\langle i|$. Let Z be an optimal solution for (5.48). Define $Y = VZV^\dagger$. We have $Z \succeq J \implies Y \succeq VJV^\dagger = |\Phi\rangle\langle\Phi|$. Since $Z \in S^\perp$ we have $Y = \sum_{i\not\sim j} Z_{ij}|ii\rangle\langle jj|$; this is an element of $S^\perp\otimes\overline{S}^\perp$. Z being entrywise nonnegative ensures that $\mathcal{R}(Y) = \sum_{ij} Z_{ij}|i\rangle\langle i|\otimes|j\rangle\langle j| \in \mathrm{SEP}$. So Y is feasible for (5.55) for $\overline{\vartheta}^+_{\mathrm{SEP}}(S)$. Its value is $\|\mathrm{Tr}_A Y\| = \|\sum_i Z_{ii}|i\rangle\langle i|\| = \bar{\vartheta}^+(G)$. Therefore $\overline{\vartheta}^+_{\mathrm{SEP}}(S) \leq \bar{\vartheta}^+(G)$.

Now let B, L' be an optimal solution for (5.45) for $\bar{\vartheta}^+(G)$. Without loss of generality, assume that L' is Hermitian (any feasible solution for (5.45) can be averaged with its adjoint). Also, assume that L' vanishes on the diagonal since zeroing the diagonal entries of L' doesn't affect feasibility for (5.45). Decompose B into diagonal and off-diagonal components: $B = \rho + T'$. Define $T = VT'V^\dagger$ and $L = VL'V^\dagger/2$. We will show this to be feasible for (5.54) for $\overline{\vartheta}^+_{\mathrm{SEP}^*}(S)$. By the arguments of theorem 5.27, $\rho \succeq 0$, $\mathrm{Tr}\rho = 1$, and $I\otimes\rho + T \succeq 0$. For $i \not\sim j$ we have $(T'+L')_{ij} = 0$ since $B + L' \in S + S_0$ and $T' + L'$ vanishes on the diagonal. So $T + L + L^\dagger = V(T' + L')V^\dagger = \sum_{i\sim j}(T' + L')_{ij}|ii\rangle\langle jj|$, which is an element of $S \otimes \overline{S}$. We have

$$\begin{aligned}
\mathcal{R}(L) &= \sum_{ij} \frac{1}{2} L'_{ij} \mathcal{R}(|ii\rangle\langle jj|) \\
&= \sum_{ij} \frac{1}{2} L'_{ij} |ij\rangle\langle ij| \in \mathrm{SEP},
\end{aligned}$$

where the last line relies on $L'_{ij} \geq 0$. Similarly $\mathcal{R}(L)^\dagger \in \text{SEP}$; therefore $\mathcal{R}(L)+\mathcal{R}(L)^\dagger \in \text{SEP} = \text{SEP}^{**}$. So ρ, T, and L are feasible for (5.54) for $\overline{\vartheta}^+_{\text{SEP}^*}(S)$. By the arguments of theorem 5.27, the value of this solution is $\bar{\vartheta}^+(G)$; therefore $\overline{\vartheta}^+_{\text{SEP}^*}(S) \geq \bar{\vartheta}^+(G)$.

Clearly $\text{SEP} \subseteq \mathcal{C} \subseteq \text{SEP}^* \implies \overline{\vartheta}^+_{\text{SEP}}(S) \geq \overline{\vartheta}^+_{\mathcal{C}}(S) \geq \overline{\vartheta}^+_{\text{SEP}^*}(S)$ since maximization programs have nonincreasing values as constraints are tightened. Combining this with the above two inequalities gives the desired equality result. □

Theorem 5.29. *Suppose a closed convex cone $\mathcal{C}$ is closed under the action of maps of the form $\mathcal{E} \otimes \overline{\mathcal{E}}$ where $\mathcal{E}$ is a completely positive trace preserving map and $\overline{\mathcal{E}}$ is the entrywise complex conjugate of $\mathcal{E}$.*[8] *In particular, the cones $\{\text{SEP}, \mathcal{S}^+, \text{PPT}, \mathcal{S}^+ \cap \text{PPT}, \text{SEP}^*\}$ satisfy this requirement. Then $\overline{\vartheta}'_{\mathcal{C}}$ and $\overline{\vartheta}^+_{\mathcal{C}}$ are homomorphism monotones in the sense that for trace-free non-commutative graphs S and T we have*

$$S \to T \implies \overline{\vartheta}'_{\mathcal{C}}(S) \leq \overline{\vartheta}'_{\mathcal{C}}(T), \tag{5.57}$$

$$\overline{\vartheta}^+_{\mathcal{C}}(S) \leq \overline{\vartheta}^+_{\mathcal{C}}(T). \tag{5.58}$$

Proof. The proof is similar to that of theorem 5.19, so we only describe the needed modifications. To prove (5.57), let $Y', L' \subseteq \mathcal{L}(B) \otimes \mathcal{L}(B')$ be a feasible solution for (5.52) for $\overline{\vartheta}'_{\mathcal{C}}(T)$. As was done in theorem 5.19, define $Y \subseteq \mathcal{L}(A) \otimes \mathcal{L}(A')$ as $Y = \sum_{ij}(E_i \otimes \overline{E}_i)^\dagger Y'(E_j \otimes \overline{E}_j)$ where the Kraus operators $\{E_i\}$ are a homomorphism $S \to T$. Similarly, define $L = \sum_{ij}(E_i \otimes \overline{E}_i)^\dagger L'(E_j \otimes \overline{E}_j)$. We will show this to be a feasible solution for (5.52) for $\overline{\vartheta}'_{\mathcal{C}}(S)$ with value at most $\overline{\vartheta}'_{\mathcal{C}}(T)$. The arguments in the proof of theorem 5.19 apply directly to show $Y \succeq |\Phi\rangle\langle\Phi|$ and $\|\text{Tr}_A Y\| \leq \|\text{Tr}_B Y'\|$.

Since Y', L' are feasible for (5.52) for $\overline{\vartheta}'_{\mathcal{C}}(T)$ we have that $Y'+L'+L'^\dagger \in T^\perp \otimes \mathcal{L}(B') + \mathcal{L}(B) \otimes \overline{T}^\perp$, giving

$$\begin{aligned} Y + L + L^\dagger &= \sum_{ij}(E_i \otimes \overline{E}_i)^\dagger (Y' + L' + L'^\dagger)(E_j \otimes \overline{E}_j) \\ &\in E^\dagger T^\perp E \otimes \overline{E}^\dagger \mathcal{L}(B')\overline{E} + E^\dagger \mathcal{L}(B) E \otimes \overline{E}^\dagger \overline{T}^\perp \overline{E} \\ &\subseteq S^\perp \otimes \mathcal{L}(A') + \mathcal{L}(A) \otimes \overline{S}^\perp. \end{aligned}$$

All that remains is to show $\mathcal{R}(L) + \mathcal{R}(L)^\dagger \in \mathcal{C}^*$. We have

$$\begin{aligned} \mathcal{R}(L) &= \sum_{ij} \mathcal{R}((E_i \otimes \overline{E}_i)^\dagger L'(E_j \otimes \overline{E}_j)) \\ &= \sum_{ij}(E_i \otimes \overline{E}_j)^\dagger \mathcal{R}(L')(E_i \otimes \overline{E}_j) \\ &= (\widehat{\mathcal{E}}^* \otimes \overline{\mathcal{E}^*})(\mathcal{R}(L')). \end{aligned} \tag{5.59}$$

Since completely positive maps commute with the taking of adjoints we also have $\mathcal{R}(L)^\dagger = (\widehat{\mathcal{E}}^* \otimes \overline{\mathcal{E}^*})(\mathcal{R}(L')^\dagger)$. Consequently, $\mathcal{R}(L) + \mathcal{R}(L)^\dagger = (\widehat{\mathcal{E}}^* \otimes \overline{\mathcal{E}^*})(\mathcal{R}(L') + \mathcal{R}(L')^\dagger)$. But $\mathcal{R}(L') + \mathcal{R}(L')^\dagger \in \mathcal{C}$ and this cone is assumed to be closed under such product maps, so $\mathcal{R}(L) + \mathcal{R}(L)^\dagger \in \mathcal{C}$.

To prove (5.58), let $Y' \subseteq \mathcal{L}(B) \otimes \mathcal{L}(B')$ be a feasible solution for (5.55) for $\overline{\vartheta}^+_{\mathcal{C}}(T)$ and define Y as in the previous paragraph. We will show this to be a feasible solution for (5.55) for $\overline{\vartheta}'_{\mathcal{C}}(S)$ with value at most $\overline{\vartheta}'_{\mathcal{C}}(T)$. Again the arguments in the proof of theorem 5.19 apply directly to show $Y \succeq |\Phi\rangle\langle\Phi|$ and $\|\text{Tr}_A Y\| \leq \|\text{Tr}_B Y'\|$. A straightforward modification of (5.27) yields $Y \in S^\perp \otimes \overline{S}^\perp$.

[8] Note that $(\mathcal{E} \otimes \overline{\mathcal{E}})(X)$ can be on a different Hilbert space than X. So, technically, one must consider a collection of cones, one for each Hilbert space. For example, SEP is such a collection.

All that remains is to show that $\mathcal{R}(Y) \in \mathcal{C}$. Similar to (5.59), we have $\mathcal{R}(Y) = (\widehat{\mathcal{E}^*} \otimes \overline{\mathcal{E}^*})(\mathcal{R}(Y'))$. But $\mathcal{R}(Y') \in \mathcal{C}$ and this cone is assumed to be closed under such product maps, so $\mathcal{R}(Y)^\dagger \in \mathcal{C}$. □

Lemma 5.30. *Let $\mathcal{C}$ be a closed convex cone. Then,*

1. $\overline{\vartheta}'_{\mathcal{C}}(K_n) = \overline{\vartheta}^+_{\mathcal{C}}(K_n) = n$ *if* $\mathcal{C} \supseteq \mathrm{SEP}$
2. $\overline{\vartheta}'_{\mathcal{C}}(Q_n) = n^2$ *if* $\mathcal{C} \supseteq \mathrm{SEP}$
3. $\overline{\vartheta}^+_{\mathcal{C}}(Q_n) = n^2$ *if* $|\Phi\rangle\langle\Phi| \in \mathcal{C}$ *(e.g. if* $\mathcal{C} \supseteq \mathcal{S}^+$*)*
4. $\overline{\vartheta}^+_{\mathcal{C}}(Q_n) = \infty$ *if* $|\Phi\rangle\langle\Phi| \notin \mathcal{C}$ *(e.g. if* $\mathcal{C} \subseteq \mathrm{PPT}$*)*

Proof. For $\mathcal{C} \supseteq \mathrm{SEP}$ we have $\overline{\vartheta}'_{\mathrm{SEP}}(K_n) \leq \overline{\vartheta}'_{\mathcal{C}}(K_n) \leq \overline{\vartheta}(K_n) \leq \overline{\vartheta}^+_{\mathcal{C}}(K_n) \leq \overline{\vartheta}^+_{\mathrm{SEP}}(K_n)$. By theorems 5.27 and 5.28 $\overline{\vartheta}'_{\mathrm{SEP}}(K_n) = \overline{\vartheta}^+_{\mathrm{SEP}}(K_n) = n$, since $\bar{\vartheta}'(K_n) = \bar{\vartheta}^+(K_n) = n$.

A feasible solution for (5.51) for $\overline{\vartheta}'_{\mathrm{SEP}}(Q_n)$ is given by $\rho = I/n$ and $T = |\Phi\rangle\langle\Phi| - I \otimes I/n$. The operator $\mathcal{R}(T) = I \otimes I - |\Phi\rangle\langle\Phi|/n$ is separable [GB02]. The value of this solution is n^2, so $\overline{\vartheta}'_{\mathrm{SEP}}(Q_n) \geq n^2$. For $\mathcal{C} \supseteq \mathrm{SEP}$ we have $\overline{\vartheta}'_{\mathrm{SEP}}(Q_n) \leq \overline{\vartheta}'_{\mathcal{C}}(Q_n) \leq \overline{\vartheta}(Q_n) = n^2$, so in fact $\overline{\vartheta}'_{\mathcal{C}}(Q_n) = n^2$.

Suppose $|\Phi\rangle\langle\Phi| \in \mathcal{C}$. Then $\mathcal{R}(I \otimes I) \in \mathcal{C}$ so a feasible solution for (5.55) for $\overline{\vartheta}^+_{\mathcal{C}}(Q_n)$ with value n^2 is given by $Y = nI \otimes I$; therefore $\overline{\vartheta}^+_{\mathcal{C}}(Q_n) \leq n^2$. But also $\overline{\vartheta}^+_{\mathcal{C}}(Q_n) \geq \overline{\vartheta}(Q_n) = n^2$, so in fact $\overline{\vartheta}^+_{\mathcal{C}}(Q_n) = n^2$.

Suppose $|\Phi\rangle\langle\Phi| \notin \mathcal{C}$. Any feasible solution for (5.55) for $\overline{\vartheta}^+_{\mathcal{C}}(Q_n)$ requires $Y \in Q_n^\perp \otimes \overline{Q}_n^\perp = \mathrm{span}\{I \otimes I\}$. In other words, $Y = cI \otimes I$ for some $c > 0$. But then $\mathcal{R}(Y) = c|\Phi\rangle\langle\Phi|$. Since $|\Phi\rangle\langle\Phi| \notin \mathcal{C}$, there can be no feasible solution. So $\overline{\vartheta}^+_{\mathcal{C}}(Q_n) = \infty$. □

Corollary 5.31. *Let S be a trace-free non-commutative graph. For $\mathcal{C} \in \{\mathrm{SEP}, \mathcal{S}^+, \mathrm{PPT}, \mathcal{S}^+ \cap \mathrm{PPT}, \mathrm{SEP}^*\}$, it holds that $\omega(S) \leq \overline{\vartheta}'_{\mathcal{C}}(S) \leq \overline{\vartheta}(S) \leq \overline{\vartheta}^+_{\mathcal{C}}(S) \leq \chi(S)$ and $[\omega_q(S)]^2 \leq \overline{\vartheta}'_{\mathcal{C}}(S)$. For $\mathcal{C} \in \{\mathcal{S}^+, \mathrm{SEP}^*\}$, $\overline{\vartheta}^+_{\mathcal{C}}(S) \leq [\chi_q(S)]^2$.*

Proof. The corollary follows from application of theorem 5.29 to the definition of $\omega(S)$, $\omega_q(S)$, $\chi(S)$, and $\chi_q(S)$, and using the values from lemma 5.30. Note that for $\mathcal{C} \subseteq \mathrm{PPT}$, in particular, the bound $\overline{\vartheta}^+_{\mathcal{C}}(S) \leq [\chi_q(S)]^2$ does not hold since $\chi_q(Q_n) = n$ but $\overline{\vartheta}^+_{\mathcal{C}}(Q_n) = \infty$. □

Having developed the basic theory of Schrijver and Szegedy numbers for non-commutative graphs, we turn now to commentary and applications. It is interesting to note that a gap between $\bar{\vartheta}'$, $\bar{\vartheta}$, and $\bar{\vartheta}^+$ for classical graphs is somewhat difficult to find and the gaps are often small. The smallest classical graph displaying a gap between any of these three quantities has 8 vertices.[9] The gap is much more pronounced for non-commutative graphs, showing up already for graphs in $\mathcal{L}(\mathbb{C}^2)$. Indeed, by lemma 5.30, $\overline{\vartheta}(Q_2) = 4$ but $\overline{\vartheta}^+_{\mathrm{PPT}}(Q_2) = \infty$. Numerical results on 10000 random graphs $S \in \mathcal{L}(\mathbb{C}^3)$ with $\dim(S) = 4$ yielded $\overline{\vartheta}'_{\mathrm{PPT}}(S) = 1$ for all test cases and $\overline{\vartheta}^+_{\mathrm{PPT}}(S) = \infty$ for 93% of test cases (with the solver failing to converge in one case).

An extreme gap between $\overline{\vartheta}$ and $\overline{\vartheta}'_{\mathrm{PPT}}$ appears for $S = \mathbb{C}\Delta$ with $\Delta = \mathrm{diag}\{d-1, -1, \ldots, -1\} \subseteq \mathcal{L}(\mathbb{C}^d)$. In this case, $\overline{\vartheta}(S) = d$ [DSW10], but $\overline{\vartheta}'_{\mathrm{PPT}}(S) = 1$. This can be seen as follows. For $\overline{\vartheta}(S)$, the feasible solution $T = \Delta \otimes |0\rangle\langle 0|$, $\rho = |0\rangle\langle 0|$ allows $\overline{\vartheta}(S) = d$. For $\overline{\vartheta}'_{\mathrm{PPT}}(S)$ it is required first of all that $T \in S \otimes \overline{S}$. The only feasible solutions are then of the form $T = c\Delta \otimes \overline{\Delta}$ for some constant c. But $\mathcal{R}(c\Delta \otimes \overline{\Delta}) \in \mathrm{PPT}$ requires $c = 0$. Therefore the only feasible solution for $\overline{\vartheta}'_{\mathrm{PPT}}(S)$ is $T = 0$, giving $\overline{\vartheta}'_{\mathrm{PPT}}(S) = 1$. So in this case $\overline{\vartheta}'_{\mathrm{PPT}}(S) = 1$ exactly matches the clique number $\omega(S)$, since $1 \leq \omega(S) \leq \overline{\vartheta}'_{\mathrm{PPT}}(S) = 1$.

[9] Verified numerically. The graph with graph6 code `GRddY{` has $\bar{\vartheta}' = 3.236$, $\bar{\vartheta} = 3.302$, $\bar{\vartheta}^+ = 3.338$.

Note, however, that the entanglement assisted clique number of $S = \mathbb{C}\Delta$ is $\omega_*(S) = 2$ [DSW10]. So, in this case, $\overline{\vartheta}'_{\mathrm{PPT}}(S)$ is *not* an upper bound on one-shot entanglement assisted capacity. This is a bit of a surprise, since for classical graphs and for any cone $\mathrm{SEP} \subseteq \mathcal{C} \subseteq \mathrm{SEP}^*$ our $\overline{\vartheta}'_{\mathcal{C}}$ and $\overline{\vartheta}^+_{\mathcal{C}}$ reduce to $\bar{\vartheta}'$ and $\bar{\vartheta}^+$ (by theorems 5.27 and 5.28), and these are known to be monotone under entanglement assisted homomorphisms [CMR+13]. In particular, for classical graphs, $\bar{\vartheta}'(G)$ upper bounds one-shot entanglement assisted capacity.

The failure of $\overline{\vartheta}'_{\mathrm{PPT}}(S)$ to bound entanglement assisted one-shot capacity ω_* can be understood as follows. This capacity is the largest n such that $K_n \overset{*}{\to} S$. By definition 5.15 this means there is some $\Lambda \succ 0$ such that $K_n \otimes \Lambda \to S$. By theorem 5.19 we have $\overline{\vartheta}(K_n \otimes \Lambda) \leq \overline{\vartheta}(S)$ and by lemma 5.18 $\overline{\vartheta}(K_n \otimes \Lambda) = n$, so $n \leq \overline{\vartheta}(S)$. Thus $\omega_*(S) \leq \overline{\vartheta}(S)$. It is this last step that breaks down for $\overline{\vartheta}'_{\mathrm{PPT}}$. By theorem 5.29 we have $\overline{\vartheta}'_{\mathrm{PPT}}(K_n \otimes \Lambda) \leq \overline{\vartheta}'_{\mathrm{PPT}}(S)$. But, as we will show in lemma 5.32, $\overline{\vartheta}'_{\mathrm{PPT}}(K_n \otimes \Lambda) = 1$, so this is a trivial bound that says nothing about n.

Although $\mathcal{C} = \mathrm{PPT}$ is therefore unsuitable for bounding entanglement assisted clique number, all is not lost. In lemma 5.33 we will show $\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes I) = \overline{\vartheta}'_{\mathcal{S}^+}(S)$. So $\overline{\vartheta}'_{\mathcal{S}^+}$ indeed provides a bound on entanglement assisted one-shot capacity, when sender and receiver share a maximally entangled state (i.e. $\Lambda = I$). For general Λ this does not hold: $\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes \Lambda)$ can be smaller than $\overline{\vartheta}'_{\mathcal{S}^+}(S)$.

Lemma 5.32. *Let S be a trace-free non-commutative graph and $\Lambda \succeq 0$ with rank$(\Lambda) > 1$. Then $\overline{\vartheta}'_{\mathrm{PPT}}(S \otimes \Lambda) = 1$.*

Proof. We will show that the only possible feasible solutions for (5.51) are those with $T = 0$. Indeed, suppose that $T \neq 0$. It is required that $T \in (S \otimes \Lambda) \otimes (\overline{S} \otimes \overline{\Lambda})$, so T must be of the form $T = T' \otimes \Lambda \otimes \overline{\Lambda}$ where $T' \in S \otimes \overline{S}$. Then $\mathcal{R}(T) = \mathcal{R}(T') \otimes \mathcal{R}(\Lambda \otimes \overline{\Lambda}) \in \mathrm{PPT}$ requires that $\mathcal{R}(\Lambda \otimes \overline{\Lambda}) \in \mathrm{PPT}$. But $\mathcal{R}(\Lambda \otimes \overline{\Lambda}) = |\psi\rangle\langle\psi|$, where $|\psi\rangle = \sum_{ij} \Lambda_{ij} |ij\rangle$, is an entangled state since $\mathrm{rank}(\Lambda) > 1$. Entangled pure states are not in PPT. □

Lemma 5.33. *Let S be a trace-free non-commutative graph and let $\Lambda \succeq 0$, $\Lambda \neq 0$. Then*

$$\frac{\overline{\vartheta}'_{\mathcal{S}^+}(S) - 1}{\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes \Lambda) - 1} = \frac{\|\Lambda\| \, Tr(\Lambda)}{Tr(\Lambda^2)}. \tag{5.60}$$

In particular, $\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes I) = \overline{\vartheta}'_{\mathcal{S}^+}(S)$.

Proof. Work in a basis in which Λ is diagonal: $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\|\Lambda\| = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

($\geq$): Say $S \subseteq \mathcal{L}(A)$ and $\Lambda \in \mathcal{L}(B)$. Let $T \in \mathcal{L}(A \otimes B \otimes A' \otimes B')$ and $\rho \in \mathcal{L}(A' \otimes B')$ be an optimal solution for (5.51) for $\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes \Lambda)$. Since $T \in (S \otimes \Lambda) \otimes (\overline{S} \otimes \overline{\Lambda})$ it must be that $T = T' \otimes (\Lambda \otimes \overline{\Lambda})$ for some $T' \in S \otimes \overline{S}$. So T is block diagonal:

$$T = \sum_{ij} \lambda_i \lambda_j T' \otimes |i\rangle\langle i|_B \otimes |j\rangle\langle j|_{B'}.$$

Without loss of generality ρ is also block diagonal: $\rho = \sum_j \rho_j \otimes |j\rangle\langle j|_{B'}$. This can be assumed since the off diagonal components of ρ can be zeroed out without affecting its trace or the relation $I_{AB} \otimes \rho + T \succeq 0$. Since $I_{AB} \otimes \rho + T$ is block diagonal and positive semidefinite, each block must be positive semidefinite: $I_A \otimes \rho_j + \lambda_i \lambda_j T' \succeq 0$ or, equivalently,

$$I_A \otimes \frac{\rho_j}{\lambda_i \lambda_j} + T' \succeq 0.$$

Let σ be the member of $\{\rho_j / \lambda_1 \lambda_j\}_j$ with the least trace. We have

$$\mathrm{Tr}(\Lambda)\mathrm{Tr}(\sigma) = \sum_j \lambda_j \mathrm{Tr}(\sigma) \leq \sum_j \mathrm{Tr}(\rho_j)/\lambda_1 = \mathrm{Tr}(\rho)/\|\Lambda\|.$$

But $\operatorname{Tr}(\rho) = 1$ so $c := \operatorname{Tr}(\sigma)^{-1} \geq \operatorname{Tr}(\Lambda)\,\|\Lambda\|$. We have $\operatorname{Tr}(c\sigma) = 1$ and $I_A \otimes \sigma + T' \succeq 0 \implies I_A \otimes c\sigma + cT' \succeq 0$. Also

$$\begin{aligned}
\mathcal{R}(T) \succeq 0 &\implies \mathcal{R}(T') \otimes \mathcal{R}(\sum_{ij} \lambda_i \lambda_j |i\rangle\langle i|_B \otimes |j\rangle\langle j|_{B'}) \succeq 0 \\
&\implies \mathcal{R}(T') \otimes (\sum_i \lambda_i |ii\rangle_{BB'})(\sum_j \lambda_j \langle jj|_{BB'}) \succeq 0 \\
&\implies \mathcal{R}(T') \succeq 0. \qquad (5.61)
\end{aligned}$$

So $c\sigma$ and cT' are feasible for (5.51) for $\overline{\vartheta}'_{\mathcal{S}^+}(S)$ with value

$$\begin{aligned}
\langle \Phi_A | I_A \otimes c\sigma + cT' | \Phi_A \rangle &= \operatorname{Tr}(c\sigma) + c\langle \Phi_A | T' | \Phi_A \rangle \\
&= 1 + \frac{c}{\operatorname{Tr}(\Lambda^2)} \langle \Phi_A | T' | \Phi_A \rangle \langle \Phi_B | \Lambda \otimes \overline{\Lambda} | \Phi_B \rangle \\
&= 1 + \frac{c}{\operatorname{Tr}(\Lambda^2)} \langle \Phi_{AB} | T | \Phi_{AB} \rangle \\
&\geq 1 + \frac{\operatorname{Tr}(\Lambda)\,\|\Lambda\|}{\operatorname{Tr}(\Lambda^2)} (\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes \Lambda) - 1). \qquad (5.62)
\end{aligned}$$

Therefore $\overline{\vartheta}'_{\mathcal{S}^+}(S) \geq$ (5.62) and the left side of (5.60) is at least as great as the right side.

($\leq$): Let ρ' and T' be an optimal solution for (5.51) for $\overline{\vartheta}'_{\mathcal{S}^+}(S)$. Define $\rho = \rho' \otimes \overline{\Lambda}/\operatorname{Tr}(\Lambda)$ and $T = T' \otimes \Lambda \otimes \overline{\Lambda}/(\|\Lambda\| \operatorname{Tr}(\Lambda))$. Then $\operatorname{Tr}(\rho) = 1$, $T \in (S \otimes \Lambda) \otimes (\overline{S} \otimes \overline{\Lambda})$, and

$$\begin{aligned}
I_{AB} \otimes \rho + T &= \frac{1}{\operatorname{Tr}(\Lambda)} \left(I_A \otimes \rho' \otimes I_B + T' \otimes \frac{\Lambda}{\|\Lambda\|} \right) \otimes \overline{\Lambda} \\
&\succeq \frac{1}{\operatorname{Tr}(\Lambda)} \left(I_A \otimes \rho' \otimes \frac{\Lambda}{\|\Lambda\|} + T' \otimes \frac{\Lambda}{\|\Lambda\|} \right) \otimes \overline{\Lambda} \\
&= \frac{1}{\operatorname{Tr}(\Lambda)\,\|\Lambda\|} (I_A \otimes \rho' + T') \otimes \Lambda \otimes \overline{\Lambda} \succeq 0.
\end{aligned}$$

And $\mathcal{R}(T) \succeq 0$ by following the logic of (5.61) in reverse. So ρ and T are feasible for (5.51) for $\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes \Lambda)$. The objective value is

$$\begin{aligned}
\langle \Phi_{AB} | I_{AB} \otimes \rho + T | \Phi_{AB} \rangle &= 1 + \langle \Phi_{AB} | T | \Phi_{AB} \rangle \\
&= 1 + \langle \Phi_A | T' | \Phi_A \rangle \langle \Phi_B | \Lambda \otimes \overline{\Lambda} | \Phi_B \rangle / (\|\Lambda\| \operatorname{Tr}(\Lambda)) \\
&= 1 + (\overline{\vartheta}'_{\mathcal{S}^+}(S) - 1) \frac{\operatorname{Tr}(\Lambda^2)}{\|\Lambda\| \operatorname{Tr}(\Lambda)}. \qquad (5.63)
\end{aligned}$$

So $\overline{\vartheta}'_{\mathcal{S}^+}(S \otimes \Lambda) \geq$ (5.63) and the left side of (5.60) is no greater than the right side. □

An extreme example of the difference between unassisted capacity and entanglement assisted capacity is given in theorem 3 of [Dua09]: a channel is defined having distinguishability graph $S = Q_n \otimes I_2$, where I_2 is the 2×2 identity operator. In [Dua09] it is shown that this channel has no unassisted zero-error classical capacity (even with many uses of the channel) but has one-shot entanglement assisted quantum capacity $\log n$. In other words, $\omega_{q*}(S) = n$. Our techniques show this result to be "obvious in retrospect". Indeed, trivially $Q_n \overset{*}{\to} S$ since $Q_n \otimes I_2 \to Q_n \otimes I_2$. So $\omega_{q*}(S) \geq n$; the channel has one-shot entanglement assisted capacity of at least $\log n$ qubits. And by lemma 5.32, $\overline{\vartheta}'_{\mathrm{PPT}}(S) = 1$ so $\omega(S) = 1$; the channel has no one-shot capacity in the absence of entanglement. Unfortunately we cannot use these techniques to bound the asymptotic capacity

$\lim_{m\to\infty} \frac{1}{m} \log \omega(S^{*m})$ since $\overline{\vartheta}'_{\mathrm{PPT}}$ is not in general multiplicative under powers S^{*m} (even for classical graphs [CMR+13]). We conjecture, however, that $\overline{\vartheta}'_{\mathcal{C}}$ (for certain cones $\mathcal{C}$) is multiplicative when $\overline{\vartheta}'_{\mathcal{C}}(S) = 1$.

Inspired by this $S = Q_n \otimes I_2$ example, we construct a channel that has no one-shot capacity when assisted by a maximally entangled state of arbitrary dimension, but does have one-shot capacity when assisted by a non-maximally entangled state. To our knowledge this is a new result. We note that the possibility of such behavior for a classical channel is still an open problem [RM12, CMR+13]. This example nicely illustrates the utility of these semidefinite programming bounds which, at least for small dimensions, are very computationally tractable. The following example was found and verified numerically before lemma 5.33 was discovered; the latter was inspired by the former.

Theorem 5.34. *There is a channel that can transmit an error-free quantum state of dimension n (i.e. $\log n$ qubits) using entanglement between sender and receiver, but that cannot transmit even a single error-free classical bit if the sender and receiver only share a maximally entangled state.*

Proof. Let $T = Q_n \otimes \Lambda$ where Λ satisfies $c := \frac{\|\Lambda\| \mathrm{Tr}(\Lambda)}{\mathrm{Tr}(\Lambda^2)} > n^2 - 1$. For instance, take $\Lambda = \mathrm{diag}(1, \alpha, \ldots, \alpha) \in \mathcal{L}(\mathbb{C}^m)$ where $\alpha = (\sqrt{m} - 1)/(m - 1)$. This maximizes c for a given m, achieving $c = (m-1)/2(\sqrt{m} - 1)$. So if $n = 2$ we can take $m = 26$ to get $c > 3$.

By lemma 2 of [Dua09], T is the distinguishability graph of some quantum channel. $Q_n \otimes \Lambda \to T$ (there is always a homomorphism from a graph to itself), so a quantum state of dimension n can be sent using an entanglement resource $|\lambda\rangle$ with reduced density operator Λ. In fact, the encoding is trivial: Alice simply puts her state to be transmitted, along with her half of the entanglement resource, directly into the channel.

On the other hand, by lemma 5.33, $\overline{\vartheta}'_{\mathcal{S}^+}(K_2 \otimes I) = \overline{\vartheta}'_{\mathcal{S}^+}(K_2) = 2$ (with I being identity on a space of arbitrary finite dimension) whereas $\overline{\vartheta}'_{\mathcal{S}^+}(T) = 1 + (\overline{\vartheta}'_{\mathcal{S}^+}(Q_n) - 1)/c = 1 + (n^2 - 1)/c < 2$. Since $\overline{\vartheta}'_{\mathcal{S}^+}$ is a homomorphism monotone, $K_2 \otimes I \not\to T$; it is not possible to transmit an error-free classical bit using a maximally entangled resource. □

As mentioned above, we conjecture that $\overline{\vartheta}'_{\mathcal{C}}$ (for certain cones $\mathcal{C}$) is multiplicative when $\overline{\vartheta}'_{\mathcal{C}}(S) = 1$. If this were the case, then $\overline{\vartheta}'_{\mathcal{C}}(S) = 1$ would be enough to guarantee that a channel has no zero-error asymptotic capacity without entanglement assistance. We might as well focus on $\mathcal{C} = \mathrm{SEP}$ since this is the smallest of the cones we have considered, and so gives the strongest bound. When is $\overline{\vartheta}'_{\mathrm{SEP}}(S) = 1$? Below we present a characterization, but leave the interpretation open.

Theorem 5.35. *Let S be a trace-free non-commutative graph. $\overline{\vartheta}'_{\mathrm{SEP}}(S) = 1$ iff there is an $M \in (S \otimes \overline{S})^{\perp}$ such that $\mathcal{R}(M) - I \in \mathrm{SEP}^*$ (i.e. is an entanglement witness). Such channels have no unassisted one-shot capacity.*

Proof. ($\implies$): Let $S \subseteq \mathcal{L}(A)$ be a trace-free non-commutative graph with $\overline{\vartheta}'_{\mathcal{C}}(S) = 1$ Let Y, L be an optimal solution for (5.52) for $\overline{\vartheta}'_{\mathcal{C}}(S)$. We have

$$\begin{aligned} \|\mathrm{Tr}_A Y\| = \overline{\vartheta}'_{\mathcal{C}}(S) = 1 &\implies \mathrm{Tr}_A Y \preceq I = \mathrm{Tr}_A(|\Phi\rangle\langle\Phi|) \\ &\implies \mathrm{Tr}_A(Y - |\Phi\rangle\langle\Phi|) \preceq 0 \\ &\implies \mathrm{Tr}(Y - |\Phi\rangle\langle\Phi|) \leq 0 \end{aligned}$$

But $Y - |\Phi\rangle\langle\Phi| \succeq 0$ so in fact $Y = |\Phi\rangle\langle\Phi|$.

Notice that $Y = |\Phi\rangle\langle\Phi|$ is symmetric under $\dagger$ and $\ddagger$ (i.e. $Y = Y^{\dagger} = Y^{\ddagger}$). The subspace $(S \otimes \overline{S})^{\perp}$ is also symmetric under these operations, as is the cone SEP^*. So we can assume without loss of generality that L is invariant under $\dagger$ and $\ddagger$. Indeed, any general L could be replaced with $(L + L^{\dagger} + L^{\ddagger} + L^{\dagger\ddagger})/4$. Then $Y + 2L \in (S \otimes \overline{S})^{\perp}$ and $\mathcal{R}(L) \in \mathrm{SEP}^*$. Define $M = Y + 2L$. Then $M \in (S \otimes \overline{S})^{\perp}$ and $\mathcal{R}(M) - I = \mathcal{R}(|\Phi\rangle\langle\Phi|) + 2\mathcal{R}(L) - I = I + 2\mathcal{R}(L) - I = 2\mathcal{R}(L) \in \mathrm{SEP}^*$.

($\Longleftarrow$): Suppose $M \in (S \otimes \overline{S})^{\perp}$ and $\mathcal{R}(M) - I \in \mathrm{SEP}^*$. By the same logic as the first part of the proof, we can assume that M is invariant under $\dagger$ and $\ddagger$, so that $M = M^{\dagger}$ and $\mathcal{R}(M) = \mathcal{R}(M)^{\dagger}$. Define $Y = |\Phi\rangle\langle\Phi|$ and $L = (M - Y)/2$. Then $Y + L + L^{\dagger} = M \in (S \otimes \overline{S})^{\perp}$ and $\mathcal{R}(L) + \mathcal{R}(L)^{\dagger} = \mathcal{R}(M) - \mathcal{R}(Y) = \mathcal{R}(M) - I \in \mathrm{SEP}^*$, so this is a feasible solution for (5.52) for $\overline{\vartheta}'_{\mathcal{C}}(S)$. Its value is $\|\mathrm{Tr}_A Y\| = \|I_{A'}\| = 1$, so $\overline{\vartheta}'_{\mathcal{C}}(S) \leq 1$. But any feasible solution has $Y \geq |\Phi\rangle\langle\Phi|$ and so must have value at least $\|\mathrm{Tr}_A\{|\Phi\rangle\langle\Phi|\}\| = 1$. Therefore also $\overline{\vartheta}'_{\mathcal{C}}(S) \geq 1$. □

We now turn our attention to $\overline{\vartheta}^{+}_{\mathcal{C}}$. Whereas $\overline{\vartheta}'_{\mathcal{C}}(S) = 1$, for any cone $\mathcal{C} \supseteq \mathrm{SEP}$, certifies that a channel has no one-shot capacity (without entanglement assistance), $\overline{\vartheta}^{+}_{\mathcal{C}}(S) = \infty$ certifies that a source cannot be transmitted using local operations and one-way classical communication (LOCC-1). This is because

$$\mathcal{C} \supseteq \mathrm{SEP} \implies \overline{\vartheta}^{+}_{\mathcal{C}}(S) \leq \overline{\vartheta}^{+}_{\mathrm{SEP}}(S) \leq \chi(S).$$

So if $\overline{\vartheta}^{+}_{\mathcal{C}}(S) = \infty$ then $\chi(S) = \infty$ and no amount of classical communication from Alice to Bob can transmit the source.

As an example, [Nat13] provides a set of three maximally entangled states that are LOCC-1 indistinguishable:

$$|\psi_0\rangle = \frac{1}{2}(|00\rangle + |11\rangle)_{A_1B_1} \otimes (|00\rangle + |11\rangle)_{A_2B_2}$$
$$|\psi_1\rangle = \frac{1}{2}(\omega\,|00\rangle + |11\rangle)_{A_1B_1} \otimes (|01\rangle + |10\rangle)_{A_2B_2}$$
$$|\psi_2\rangle = \frac{1}{2}(\gamma\,|00\rangle + |11\rangle)_{A_1B_1} \otimes (|00\rangle - |11\rangle)_{A_2B_2}$$

where ω and γ are phases in general position. The characteristic graph for this source is $\mathrm{span}\{I, Z\} \otimes Q_2$. The quantity $\overline{\vartheta}^{+}_{\mathrm{PPT}}(S)$ is efficiently computable numerically (at least for spaces this small), and immediately provides a certificate that these states are LOCC-1 indistinguishable, with no manual computation needed. In the case of this example there is in fact an alternate proof of this result. If the three states defined above were LOCC-1 distinguishable then there would be an n such that $\mathrm{span}\{I, Z\} \otimes Q_2 \to K_n$. But then

$$Q_2 \to \mathrm{diag}(1, 0) \otimes Q_2 \to \mathrm{span}\{I, Z\} \otimes Q_2 \to K_n$$

where the second follows from $\mathrm{diag}(1, 0) \otimes Q_2 \subseteq \mathrm{span}\{I, Z\} \otimes Q_2$. By transitivity of homomorphisms this yields $Q_2 \to K_n$. But a qubit cannot be transmitted through a classical channel so $Q_2 \not\to K_n$.

5.9 Conclusion

We have defined and investigated the problem of quantum zero-error source-channel coding. This broad class of problems includes dense coding, teleportation, channel capacity, and one-way LOCC state measurement. Whereas classical zero-error source-channel coding relies on graphs, the quantum version relies on non-commutative graphs. Central to this theory is a generalization of the notion of graph homomorphism to non-commutative graphs.

For classical graphs, it is known that the Lovász number is monotone under homomorphisms (and in fact even entanglement assisted homomorphisms). The Lovász number has been generalized to non-commutative graphs by [DSW13]; we showed this quantity to be monotone under entanglement assisted homomorphisms on non-commutative graphs.

We investigated the problem of sending many parallel source instances using many parallel channels and found that the Lovász number provides a bound on the cost rate, but only if the source

satisfies a particular condition. Classical sources, as well as sources that can produce a maximally entangled state, both satisfy this condition.

We defined Schrijver and Szegedy quantities for non-commutative graphs. These are monotone under non-commutative graph homomorphisms, but not entanglement assisted homomorphisms. In fact, we derived a sequence of such quantities that are all equal to the traditional Schrijver and Szegedy quantities for classical graphs but can take different values on general non-commutative graphs. These results were used to investigate some known examples from the literature regarding entanglement assisted communication over a noisy channel and one-way LOCC measurements. Strangely, one of the Schrijver variants, $\overline{\vartheta}'_{\mathcal{S}^+}$, scores non-maximally entangled states as more valuable a resource than maximally entangled states (which are not even visible to $\overline{\vartheta}'_{\mathcal{S}^+}$). Exploiting this oddity we constructed a channel that can transmit several zero-error qubits if sender and receiver can share an arbitrary entangled state, but cannot transmit even a single classical bit if only a maximally entangled resource is allowed. It is still an open question whether such behavior is possible for a classical channel.

Most of all, and more importantly than any specific bounds provided for the quantum source-channel coding problem, we have furthered the program of non-commutative graph theory set forth in [DSW13]. It is a curiosity that a field as discrete as graph theory can be "quantized" by replacing sets with Hilbert spaces and binary relations with operator subspaces. Non-commutative graphs offer the promise that some of the wealth of graph theory may be imported into the theory of operator subspaces. But actually this promise is more of a tease, as even the most basic facts from graph theory lead only to (interesting!) open questions in the theory of non-commutative graphs. We close by outlining some these questions.

- For classical graphs, $\chi(G)\omega(\overline{G}) \geq |V(G)|$. Does this hold also for non-commutative graphs, with an appropriate definition of graph complement? We propose the complement (for trace-free graphs) $S^c = (S + \mathbb{C}I)^\perp$, and conjecture that $\chi(S)\omega(S^c) \geq n$ where $S \subseteq \mathcal{L}(\mathbb{C}^n)$. Note that $\chi(S)$ and $\omega(S^c)$ are only defined when S and S^c are both trace free. Similarly, does it hold that $\overline{\vartheta}(S)\overline{\vartheta}(S^c) \geq n$?

- What is the analogue of vertex transitive for non-commutative graphs, and what are the properties of these graphs? We propose to define the automorphism group as $\mathrm{Aut}(S) = \{U \text{ unitary} : USU^\dagger = S\}$ and to call such a group vertex transitive if the only operators satisfying $U\rho U^\dagger = \rho$ for all $U \in \mathrm{Aut}(S)$ are those proportional to identity.

- A Hamiltonian path for a trace-free non-commutative graph $S \in \mathcal{L}(\mathbb{C}^n)$ can be taken to be a set of nonzero vectors such that $|\psi_i\rangle\langle\psi_{i+1}| \in S$ for $i \in \{1, \ldots, n-1\}$. Does the Lovász conjecture generalize? That is to say, does every connected trace-free vertex transitive non-commutative graph have a Hamiltonian path?

- Let S be a non-commutative graph associated with the classical graph G. We saw that $\chi(G)$ is the smallest n such that $S \to K_n$ and orthogonal rank $\xi(G)$ is the smallest n such that $S \to Q_n$. Projective rank ξ_f [RM12] is to ξ as fractional chromatic number χ_f is to χ. Since $\chi_f(G) = \min\{p/q : G \to K_{p:q}\}$ where $K_{p:q}$ is the Kneser graph [GR01], is it the case that $\xi_f(G) = \min\{p/q : S \to K'_{p:q}\}$ for some class of non-commutative graphs $K'_{p:q}$?

- How is the distinguishability graph of a channel related to that of the complementary channel? The same question can be asked for the source: swapping Alice and Bob's inputs defines a complementary source.

- For classical graphs, $\bar{\vartheta}'$ and $\bar{\vartheta}^+$ are monotone under entanglement assisted homomorphisms. For non-commutative graphs this does not always hold. Is there some insight here? Or does this mean there is some better generalization of $\bar{\vartheta}'$ and $\bar{\vartheta}^+$?

- Is it the case that $\overline{\vartheta}'_{\mathcal{C}}(S) = 1$ implies $\overline{\vartheta}'_{\mathcal{C}}(S^{*n}) = 1$, for some suitable choice of $\mathcal{C}$? If so, $\overline{\vartheta}'_{\mathcal{C}}(S) = 1$ would certify that a channel had no asymptotic zero-error capacity.
- Any trace-free non-commutative graph is both the characteristic graph of some source and the distinguishability graph of some channel. Is there something to be learned from this relation between sources and channels?
- It is known that two channels with no one-shot capacity, when put in parallel, may have positive one-shot capacity [CCH11, Dua09, CS12]. Is there a similar effect with sources? Are there two sources that are both one-way LOCC (LOCC-1) indistinguishable but in parallel are LOCC-1 distinguishable?
- The quantity $\|\Lambda\| \operatorname{Tr}(\Lambda)/\operatorname{Tr}(\Lambda^2)$, which shows up in lemma 5.33, is only greater than 1 for the reduced density operator of a non-maximally entangled state. Is this an ad hoc quantity, or is it a meaningful measure of entanglement?

5.10 Acknowledgments

The author would like to thank Simone Severini, David Roberson, and Vern Paulsen for many helpful discussions.

This research received financial support from the National Science Foundation through Grant PHY-1068331.

5.A Duality Proofs

We will derive the dual of (5.51), which we rewrite here for reference.

$$\begin{aligned} \overline{\vartheta}'_{\mathcal{C}}(S) = \max \; & \langle \Phi | I \otimes \rho + T | \Phi \rangle \\ \text{s.t.} \; & \rho \succeq 0, \operatorname{Tr}\rho = 1, \\ & I \otimes \rho + T \succeq 0, \\ & T \in S \otimes \overline{S}, \\ & \mathcal{R}(T) \in \mathcal{C}, \end{aligned} \tag{5.64}$$

Section 4.7 of [GM12] gives the following duality recipe for conic programming over real vectors, where $\mathcal{G}$ and $\mathcal{H}$ are closed convex cones:

$$\begin{aligned} \text{(Primal)} \quad & \max \langle \mathbf{c}, \boldsymbol{x} \rangle \\ & \text{s.t.} \; \mathbf{b} - A(\boldsymbol{x}) \in \mathcal{G}, \\ & \quad \boldsymbol{x} \in \mathcal{H} \end{aligned} \tag{5.65}$$

$$\begin{aligned} \text{(Dual)} \quad & \min \langle \mathbf{b}, \boldsymbol{y} \rangle \\ & \text{s.t.} \; A^T(\boldsymbol{y}) - \mathbf{c} \in \mathcal{H}^*, \\ & \quad \boldsymbol{y} \in \mathcal{G}^*. \end{aligned} \tag{5.66}$$

This nearly suffices for our purposes, since (5.64) can be viewed as a program over real vectors by considering the real inner product space of Hermitian matrices with the Hilbert–Schmidt inner product (cf. [Wat11] for the special case where the cones are $\mathcal{S}^+$). The difficulty is that the superoperator $\mathcal{R}$ is not Hermiticity-preserving, and so cannot be considered as a linear map on the space of Hermitian matrices. This is not hard to fix, as the condition $\mathcal{R}(T) \in \mathcal{C}$ requires $\mathcal{R}(T)$ to be Hermitian and so is equivalent to the pair of conditions $\mathcal{R}(T) - \mathcal{R}(T)^\dagger = 0$ and $\mathcal{R}(T) + \mathcal{R}(T)^\dagger \in \mathcal{C}$. The first of these can

also be written $T - T^\ddagger = 0$ (recall that we define $X^\ddagger = \mathcal{R}(\mathcal{R}(X)^\dagger)$). Note that the left-hand sides of these relations, seen as superoperators (e.g. $T \to T - T^\ddagger$), are not linear in the space $\mathcal{L}(A) \otimes \mathcal{L}(A')$ since they each contain an anti-linear term. They are, however, linear in the real inner product space of Hermitian matrices. Within this space, the map $T \to \mathcal{R}(T) + \mathcal{R}(T)^\dagger$ is self-adjoint. Indeed, for Hermitian L, T we have

$$\begin{aligned}
\langle L, \mathcal{R}(T) + \mathcal{R}(T)^\dagger \rangle &= \langle L, \mathcal{R}(T) \rangle + \langle L, \mathcal{R}(T)^\dagger \rangle \\
&= \langle L, \mathcal{R}(T) \rangle + \langle \mathcal{R}(T), L \rangle^* \\
&= \langle \mathcal{R}(L), T \rangle + \langle T, \mathcal{R}(L) \rangle^* \\
&= \langle \mathcal{R}(L), T \rangle + \langle \mathcal{R}(L)^\dagger, T \rangle \\
&= \langle \mathcal{R}(L) + \mathcal{R}(L)^\dagger, T \rangle.
\end{aligned}$$

The map $T \to T - T^\ddagger$ is also self-adjoint within the space of Hermitian matrices. The primal becomes

$$\begin{aligned}
\overline{\vartheta}'_{\mathcal{C}}(S) = \max \ & \langle \Phi | I \otimes \rho + T | \Phi \rangle & \\
\text{s.t.} \ & 1 - \mathrm{Tr}\rho = 0, & (5.67) \\
& I \otimes \rho + T \succeq 0, & (5.68) \\
& T - T^\ddagger = 0, & (5.69) \\
& \mathcal{R}(T) + \mathcal{R}(T)^\dagger \in \mathcal{C}, & (5.70) \\
& \rho \succeq 0, T \in S \otimes \overline{S}, & (5.71)
\end{aligned}$$

Applying the recipe (5.66) gives a dual formulation with a variable for each constraint in the primal: λ for (5.67), W for (5.68), X for (5.69), and L' for (5.70). In other words, $\boldsymbol{y} = \lambda \oplus W \oplus X \oplus L'$ (with these thought of as vectors in the inner product space of Hermitian matrices). The (5.71) constraints correspond to the $\boldsymbol{x} \in K$ constraint in (5.65), taking $\boldsymbol{x} = \rho \oplus T$. The dual will have a constraint for each variable of the primal: (5.72) for ρ and (5.73) for T. The dual is then

$$\begin{aligned}
&\min \lambda \\
&\text{s.t.} \ \ \lambda I - \mathrm{Tr}_A W - I \succeq 0, && (5.72) \\
&\qquad - (W + X - X^\ddagger + \mathcal{R}(L') + \mathcal{R}(L')^\dagger) - |\Phi\rangle\langle\Phi| \in (S \otimes \overline{S})^\perp, && (5.73) \\
&\qquad \lambda \in \mathbb{R}, W \succeq 0, X \text{ Hermitian}, L' \in \mathcal{C}^*. && (5.74)
\end{aligned}$$

Define $Y = W + |\Phi\rangle\langle\Phi|$ and $L = \mathcal{R}(L') + (X - X^\ddagger)/2$. Note that L is not necessarily Hermitian, but L' is since $L' \in \mathcal{C}^*$. We have $\mathcal{R}(L) + \mathcal{R}(L)^\dagger = L' + L'^\dagger + (\mathcal{R}(X) - \mathcal{R}(X^\ddagger) + \mathcal{R}(X)^\dagger - \mathcal{R}(X^\ddagger)^\dagger)/2 = 2L' \in \mathcal{C}^*$ since $\mathcal{R}(X^\ddagger) = \mathcal{R}(X)^\dagger$. So these give a solution to

$$\begin{aligned}
&\min \lambda \\
&\text{s.t.} \ \ \lambda I - \mathrm{Tr}_A Y \succeq 0 \\
&\qquad Y + (L + L^\dagger) \in (S \otimes \overline{S})^\perp, \\
&\qquad \mathcal{R}(L) + \mathcal{R}(L)^\dagger \in \mathcal{C}^*, \\
&\qquad Y \succeq |\Phi\rangle\langle\Phi|, \\
&\qquad \lambda \in \mathbb{R}, L \in \mathcal{L}(A) \otimes \mathcal{L}(A'). && (5.75)
\end{aligned}$$

Conversely, a solution to (5.75) gives a solution to (5.72)-(5.74) via $W = Y - |\Phi\rangle\langle\Phi|$, $L' = (\mathcal{R}(L) + \mathcal{R}(L)^\dagger)/2$, $X = [(L - L^\ddagger) + (L - L^\ddagger)^\dagger]/4$. The program (5.75) is equivalent to (5.52).

We now show the primal and dual to have equal and finite optimum values. Let L' be in the relative interior[10] of $\mathcal{C}^*$, and let $X = 0$. There is a $W \succ 0$ such that the left hand side of (5.73) is

[10] See [BV04] for the definition of relative interior.

proportional to negative identity, and so is in $(S \otimes \overline{S})^{\perp}$. For a large enough λ, (5.72) is satisfied with strict inequality. Thus the dual program (5.72)-(5.74) is strictly feasible. The dual has finite value because $W \succeq 0$ requires $\lambda \geq 1$ in (5.72). Therefore strong duality holds: the primal (5.67)-(5.71) and dual (5.72)-(5.74) are both feasible and take the same optimal value. Since these are equivalent to (5.64) and (5.75), these two are also feasible and take the same value.

We now compute the primal for (5.55), which we rewrite here for reference.

$$\begin{aligned} \overline{\vartheta}^{+}_{\mathcal{C}}(S) = \min \ & \|\mathrm{Tr}_A Y\| \\ \text{s.t.} \ & Y \in S^{\perp} \otimes \overline{S}^{\perp}, \\ & \mathcal{R}(Y) \in \mathcal{C}, \\ & Y \succeq |\Phi\rangle\langle\Phi| . \end{aligned} \tag{5.76}$$

As with $\overline{\vartheta}'_{\mathcal{C}}$, we can rewrite this using only Hermiticity preserving maps:

$$\begin{aligned} \overline{\vartheta}^{+}_{\mathcal{C}}(S) = \min \ & \lambda \\ \text{s.t.} \ & \lambda I - \mathrm{Tr}_A Y \succeq 0, && (5.77) \\ & Y - |\Phi\rangle\langle\Phi| \succeq 0, && (5.78) \\ & Y - Y^{\ddagger} = 0, && (5.79) \\ & \mathcal{R}(Y) + \mathcal{R}(Y)^{\dagger} \in \mathcal{C}, && (5.80) \\ & \lambda \in \mathbb{R}, Y \in S^{\perp} \otimes \overline{S}^{\perp}. && (5.81) \end{aligned}$$

The primal will have a variable for each constraint in the dual: ρ for (5.77), T' for (5.78), X for (5.79), and L' for (5.80). In other words, $\boldsymbol{x} = \rho \oplus T' \oplus X \oplus L'$. The (5.81) constraints correspond to $\boldsymbol{y} \in \mathcal{G}^*$ in (5.66). The dual will have a constraint for each variable of the primal: (5.82) for λ and (5.83) for Y. The primal is then

$$\begin{aligned} \max \ & \langle\Phi|T'|\Phi\rangle \\ \text{s.t.} \ & 1 - \mathrm{Tr}\rho = 0, && (5.82) \\ & I \otimes \rho - T' - X + X^{\ddagger} - \mathcal{R}(L') - \mathcal{R}(L')^{\dagger} \in (S^{\perp} \otimes \overline{S}^{\perp})^{\perp}, && (5.83) \\ & \rho \succeq 0, T' \succeq 0, X \text{ Hermitian}, L' \in \mathcal{C}^*. && (5.84) \end{aligned}$$

Define $T = T' - I \otimes \rho$ and $L = \mathcal{R}(L') + (X - X^{\ddagger})/2$. Note that L is not necessarily Hermitian, but L' is since $L' \in \mathcal{C}^*$. As before, we have $\mathcal{R}(L) + \mathcal{R}(L)^{\dagger} \in \mathcal{C}^*$. These give a solution to

$$\begin{aligned} \max \ & \langle\Phi|I \otimes \rho + T|\Phi\rangle \\ \text{s.t.} \ & \mathrm{Tr}\rho = 1, \\ & T + (L + L^{\dagger}) \in (S^{\perp} \otimes \overline{S}^{\perp})^{\perp} \\ & \mathcal{R}(L) + \mathcal{R}(L)^{\dagger} \in \mathcal{C}^*, \\ & \rho \succeq 0, I \otimes \rho + T \succeq 0, \\ & L \in \mathcal{L}(A) \otimes \mathcal{L}(A'). \end{aligned} \tag{5.85}$$

Conversely, a solution to (5.85) gives a solution to (5.82)-(5.84) via $T' = T + I \otimes \rho$, $L' = (\mathcal{R}(L) + \mathcal{R}(L)^{\dagger})/2$, $X = [(L - L^{\ddagger}) + (L - L^{\ddagger})^{\dagger}]/4$. The program (5.85) is equivalent to (5.54).

We now show the primal and dual to have equal, but not necessarily finite, optimum values. Let L'' be in the relative interior of $\mathcal{C}^*$; then for any $c > 0$, $L' = cL''$ is also in the relative interior of $\mathcal{C}^*$. Let $X = 0$ and $\rho = I/\dim(A)$. For sufficiently small c, there is a $T' \succ 0$ such that the left hand side of (5.83) vanishes. Thus the primal (5.82)-(5.84) is strictly feasible. If the primal takes

finite optimum value, then by strong duality the dual is feasible and takes the same value. On the other hand, if the primal is unbounded (has infinite optimal value) then by weak duality the dual is infeasible and so also has infinite value. See lemma 5.30 for an example of such a case.

Bibliography

[Aar10] S. Aaronson, "BQP and the polynomial hierarchy," in *Proceedings of the 42nd ACM symposium on Theory of computing*, ser. STOC '10. New York, NY, USA: ACM, 2010, pp. 141–150. [Online]. Available: http://doi.acm.org/10.1145/1806689.1806711

[AGR81] A. Aspect, P. Grangier, and G. Roger, "Experimental tests of realistic local theories via bell's theorem," *Phys. Rev. Lett.*, vol. 47, pp. 460–463, Aug 1981. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.47.460

[AHKS06] D. Avis, J. Hasegawa, Y. Kikuchi, and Y. Sasaki, "A quantum protocol to win the graph colouring game on all Hadamard graphs," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 89, no. 5, pp. 1378–1381, 2006.

[ALM07] D. Aharonov, Z. Landau, and J. Makowsky, "The quantum FFT can be classically simulated," 2007. [Online]. Available: http://www.arxiv.org/abs/quant-ph/0611156

[Alo98] N. Alon, "The Shannon capacity of a union," *Combinatorica*, vol. 18, no. 3, pp. 301–310, 1998. [Online]. Available: http://dx.doi.org/10.1007/PL00009824

[Arm10] J. Armstrong, "The extremal case of HölderâĂŹs inequality," 2010, accessed: 2012/09/02. [Online]. Available: https://unapologetic.wordpress.com/2010/09/01/the-extremal-case-of-holders-inequality

[BaH10] F. G. S. L. Brandão and M. Horodecki, "Exponential quantum speed-ups are generic," 2010. [Online]. Available: http://www.arxiv.org/abs/1010.3654

[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.*, vol. 70, pp. 1895–1899, Mar 1993. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.70.1895

[BBL+13] J. Briët, H. Buhrman, M. Laurent, T. Piovesan, and G. Scarpa, "Zero-error source-channel coding with entanglement," 2013. [Online]. Available: http://www.arxiv.org/abs/1308.4283

[BCP+14] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," *Rev. Mod. Phys.*, vol. 86, pp. 419–478, Apr 2014. [Online]. Available: http://link.aps.org/doi/10.1103/RevModPhys.86.419

[Bei10] S. Beigi, "Entanglement-assisted zero-error capacity is upper-bounded by the Lovász ϑ function," *Phys. Rev. A*, vol. 82, p. 010303, Jul 2010. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.82.010303

[Bel64] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, vol. 1, pp. 195–200, 1964.

[Ben95] C. H. Bennett, "Quantum information and computation," *Physics Today*, vol. 48, no. 10, pp. 24–30, 1995.

[BFL10] I. Bomze, F. Frommlet, and M. Locatelli, "Gap, cosum and product properties of the θ' bound on the clique number," *Optimization*, vol. 59, no. 7, pp. 1041–1051, 2010.

[BG06] D. Braun and B. Georgeot, "Quantitative measure of interference," *Phys. Rev. A*, vol. 73, p. 022314, Feb 2006. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.73.022314

[BKK07] C. Bény, A. Kempf, and D. W. Kribs, "Generalization of quantum error correction via the Heisenberg picture," *Phys. Rev. Lett.*, vol. 98, p. 100502, Mar 2007. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.98.100502

[BLCP14] D. Bernstein, T. Lange, P.-L. Cayrel, and C. Peters, "Post-quantum cryptography," 2014, accessed: 2014/07/23. [Online]. Available: http://pqcrypto.org

[BM95] R. Bačík and S. Mahajan, "Semidefinite programming and its applications to np problems," in *Computing and Combinatorics*, ser. Lecture Notes in Computer Science, D.-Z. Du and M. Li, Eds. Springer Berlin Heidelberg, 1995, vol. 959, pp. 566–575. [Online]. Available: http://dx.doi.org/10.1007/BFb0030878

[BO11] C. Bény and O. Oreshkov, "Approximate simulation of quantum channels," *Phys. Rev. A*, vol. 84, p. 022333, Aug 2011. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.84.022333

[Boh51] D. Bohm, *Quantum theory*, ser. Prentice-Hall physics series. Prentice-Hall, 1951.

[Boy74] D. W. Boyd, "The power method for ℓ^p norms," *Linear Algebra and its Applications*, vol. 9, no. 0, pp. 95 – 101, 1974. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0024379574900299

[BS08] S. Beigi and P. W. Shor, "On the complexity of computing zero-error and Holevo capacity of quantum channels," 2008. [Online]. Available: http://www.arxiv.org/abs/0709.2090

[BV04] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[BV11] A. Bhaskara and A. Vijayaraghavan, "Approximating matrix p-norms," in *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '11. SIAM, 2011, pp. 497–511. [Online]. Available: http://dl.acm.org/citation.cfm?id=2133036.2133076

[BW92] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Phys. Rev. Lett.*, vol. 69, pp. 2881–2884, Nov 1992. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.69.2881

[BZ06] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States*. Cambridge: Cambridge University Press, 2006.

[Car04] N. L. Carothers, *A Short Course on Banach Space Theory (London Mathematical Society Student Texts)*. Cambridge University Press, 12 2004.

[CCD+03] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, "Exponential algorithmic speedup by a quantum walk," in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, ser. STOC '03. New York, NY, USA: ACM, 2003, pp. 59–68. [Online]. Available: http://doi.acm.org/10.1145/780542.780552

[CCH11] T. Cubitt, J. Chen, and A. Harrow, "Superactivation of the asymptotic zero-error classical capacity of a quantum channel," *Information Theory, IEEE Transactions on*, vol. 57, no. 12, pp. 8114–8126, Dec 2011.

[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.*, vol. 86, no. 3, pp. 544–547, Jan 2001.

[CHSH69] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.*, vol. 23, pp. 880–884, Oct 1969. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.23.880

[CLMW10] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, "Improving zero-error classical communication with entanglement," *Phys. Rev. Lett.*, vol. 104, p. 230503, Jun 2010. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.104.230503

[CMN+07] P. J. Cameron, A. Montanaro, M. W. Newman, S. Severini, and A. Winter, "On the quantum chromatic number of a graph," *Electron. J. Combin*, vol. 14, no. 1, 2007.

[CMR+13] T. Cubitt, L. Mančinska, D. Roberson, S. Severini, D. Stahlke, and A. Winter, "Bounds on entanglement assisted source-channel coding via the Lovász ϑ number and its variants," 2013. [Online]. Available: http://www.arxiv.org/abs/1310.7120

[Coh10] S. M. Cohen, "Optimizing local protocols for implementing bipartite nonlocal unitary gates using prior entanglement and classical communication," *Phys. Rev. A*, vol. 81, no. 6, p. 062316, Jun 2010.

[CS12] T. Cubitt and G. Smith, "An extreme form of superactivation for quantum zero-error capacities," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1953–1961, March 2012.

[Cvi08] P. Cvitanovic, *Group Theory: Birdtracks, Lie's, and Exceptional Groups*. Princeton University Press, 7 2008.

[dCST13] M. K. de Carli Silva and L. Tunçel, "Optimization problems over unit-distance representations of graphs," *The Electronic Journal of Combinatorics*, vol. 20, no. 1, p. P43, 2013.

[DJ92] D. Deutsch and R. Jozsa, "Rapid solution of problems by quantum computation," *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 439, no. 1907, pp. 553–558, 1992. [Online]. Available: http://rspa.royalsocietypublishing.org/content/439/1907/553.abstract

[dKP02] E. de Klerk and D. Pasechnik, "Approximation of the stability number of a graph via copositive programming," *SIAM Journal on Optimization*, vol. 12, no. 4, pp. 875–892, 2002. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/S1052623401383248

[DSW10] R. Duan, S. Severini, and A. Winter, "Zero-error communication via quantum channels, non-commutative graphs and a quantum Lovász ϑ function," 2010. [Online]. Available: http://www.arxiv.org/abs/1002.2514

[DSW13] ——, "Zero-error communication via quantum channels, noncommutative graphs, and a quantum Lovász number," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 1164–1174, 2013.

[Dua09] R. Duan, "Super-activation of zero-error capacity of noisy quantum channels," 2009. [Online]. Available: http://www.arxiv.org/abs/0906.2527

[DVC02] W. Dür, G. Vidal, and J. I. Cirac, "Optimal conversion of nonlocal unitary operations," *Phys. Rev. Lett.*, vol. 89, no. 5, p. 057901, Jul 2002.

[dW01] R. M. de Wolf, "Quantum computing and communication complexity," Ph.D. dissertation, University of Amsterdam, 2001. [Online]. Available: http://homepages.cwi.nl/~rdewolf/publ/qc/phd.pdf

[Eas10] B. Eastin, "Simulating concordant computations," 2010. [Online]. Available: http://www.arxiv.org/abs/1006.4402

[EJPP00] J. Eisert, K. Jacobs, P. Papadopoulos, and M. B. Plenio, "Optimal local implementation of nonlocal quantum gates," *Phys. Rev. A*, vol. 62, no. 5, p. 052317, Oct 2000.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" *Phys. Rev.*, vol. 47, pp. 777–780, May 1935. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRev.47.777

[FC72] S. J. Freedman and J. F. Clauser, "Experimental test of local hidden-variable theories," *Phys. Rev. Lett.*, vol. 28, pp. 938–941, Apr 1972. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.28.938

[Fey82] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, no. 6-7, pp. 467–488, 1982. [Online]. Available: http://dx.doi.org/10.1007/BF02650179

[FL92] U. Feige and L. Lovász, "Two-prover one-round proof systems: Their power and their problems (extended abstract)," in *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*, ser. STOC '92. New York, NY, USA: ACM, 1992, pp. 733–744. [Online]. Available: http://doi.acm.org/10.1145/129712.129783

[For03] L. Fortnow, "One complexity theorist's view of quantum computing," *Theoretical Computer Science*, vol. 292, no. 3, pp. 597 – 610, 2003, algorithms in Quantum Information Prcoessing. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304397501003772

[FRS12] J. F. Fitzsimons, E. G. Rieffel, and V. Scarani, "The quantum frontier," 2012. [Online]. Available: http://www.arxiv.org/abs/1206.0785

[FS13] T. Feng and S. Severini, "Quantum channels from association schemes," 2013. [Online]. Available: http://www.arxiv.org/abs/1301.1166

[Gal00] A. Galtman, "Spectral characterizations of the Lovász number and the Delsarte number of a graph," *Journal of Algebraic Combinatorics*, vol. 12, no. 2, pp. 131–143, 2000.

[GB02] L. Gurvits and H. Barnum, "Largest separable balls around the maximally mixed bipartite quantum state," *Phys. Rev. A*, vol. 66, p. 062311, Dec 2002. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.66.062311

[GG08] V. Gheorghiu and R. B. Griffiths, "Separable operations on pure states," *Phys. Rev. A*, vol. 78, no. 2, p. 020304, Aug 2008.

[GL08] N. Gvozdenović and M. Laurent, "The operator Ψ for the chromatic number of a graph," *SIAM Journal on Optimization*, vol. 19, no. 2, pp. 572–591, 2008. [Online]. Available: http://dx.doi.org/10.1137/050648237

[GM12] B. Gärtner and J. Matousek, *Approximation Algorithms and Semidefinite Programming*, 2012th ed. Springer, 1 2012.

[GMH93] M. Gell-Mann and J. B. Hartle, "Classical equations for quantum systems," *Phys. Rev. D*, vol. 47, pp. 3345–3382, Apr 1993. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.47.3345

[GMR+13] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin, and A. Zeilinger, "Bell violation using entangled photons without the fair-sampling assumption," *Nature*, vol. 497, no. 7448, pp. 227–230, May 2013, letter. [Online]. Available: http://dx.doi.org/10.1038/nature12012

[Got98] D. Gottesman, "The Heisenberg representation of quantum computers," 1998. [Online]. Available: http://www.arxiv.org/abs/quant-ph/9807006

[GR01] C. Godsil and G. F. Royle, *Algebraic Graph Theory (Graduate Texts in Mathematics)*, 2001st ed. Springer, 5 2001.

[Gri03] R. B. Griffiths, *Consistent Quantum Theory*. Cambridge University Press, 12 2003. [Online]. Available: http://quantum.phys.cmu.edu/CQT/

[Gro96] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, ser. STOC '96. New York, NY, USA: ACM, 1996, pp. 212–219. [Online]. Available: http://doi.acm.org/10.1145/237814.237866

[GW02] V. Galliard and S. Wolf, "Pseudo-telepathy, entanglement, and graph colorings," in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, 2002, pp. 101–.

[GWYC06] R. B. Griffiths, S. Wu, L. Yu, and S. M. Cohen, "Atemporal diagrams for quantum circuits," *Phys. Rev. A*, vol. 73, no. 5, p. 052309, May 2006.

[GYC10] V. Gheorghiu, L. Yu, and S. M. Cohen, "Local cloning of entangled states," *Phys. Rev. A*, vol. 82, no. 2, p. 022313, Aug 2010.

[Hae78] W. H. Haemers, "An upper bound for the Shannon capacity of a graph," *Colloquia Mathematica Societatis Janos Bolyai*, vol. 25, pp. 267–272, 1978. [Online]. Available: http://arno.uvt.nl/show.cgi?fid=80314

[Hae79] ——, "On some problems of Lovász concerning the Shannon capacity of a graph," *IEEE Transactions on Information Theory*, vol. 25, pp. 231–232, 1979. [Online]. Available: http://arno.uvt.nl/show.cgi?fid=80311

[HHHH09] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Rev. Mod. Phys.*, vol. 81, no. 2, pp. 865–942, Jun 2009.

[HJ90] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2 1990.

[HN04] P. Hell and J. Nesetril, *Graphs and Homomorphisms (Oxford Lecture Series in Mathematics and Its Applications)*. Oxford University Press, USA, 9 2004.

[Hoe63] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. pp. 13–30, 1963. [Online]. Available: http://www.jstor.org/stable/2282952

[Hoy97] P. Hoyer, "Efficient quantum transforms," 1997. [Online]. Available: http://www.arxiv.org/abs/quant-ph/9702028

[HT97] G. Hahn and C. Tardif, "Graph homomorphisms: structure and symmetry," in *Graph symmetry*. Springer, 1997, pp. 107–166.

[JL03] R. Jozsa and N. Linden, "On the role of entanglement in quantum-computational speed-up," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 459, no. 2036, pp. 2011–2032, 2003. [Online]. Available: http://rspa.royalsocietypublishing.org/content/459/2036/2011.abstract

[JNP+11] M. Junge, M. Navascues, C. Palazuelos, D. Perez-Garcia, V. B. Scholz, and R. F. Werner, "Connes' embedding problem and Tsirelson's problem," *Journal of Mathematical Physics*, vol. 52, no. 1, 2011. [Online]. Available: http://scitation.aip.org/content/aip/journal/jmp/52/1/10.1063/1.3514538

[Joz06] R. Jozsa, "On the simulation of quantum circuits," 2006. [Online]. Available: http://www.arxiv.org/abs/quant-ph/0603163

[KL97] E. Knill and R. Laflamme, "Theory of quantum error-correcting codes," *Phys. Rev. A*, vol. 55, pp. 900–911, Feb 1997. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.55.900

[KN06] E. Kushilevitz and N. Nisan, *Communication Complexity*. Cambridge University Press, 11 2006.

[KNR95] I. Kremer, N. Nisan, and D. Ron, "On randomized one-round communication complexity," in *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, ser. STOC '95. New York, NY, USA: ACM, 1995, pp. 596–605. [Online]. Available: http://doi.acm.org/10.1145/225058.225277

[Knu94] D. E. Knuth, "The sandwich theorem," *Electron. J. Combin.*, vol. 1, 1994. [Online]. Available: http://www.combinatorics.org/Volume_1/Abstracts/v1i1a1.html

[Kob87] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation*, vol. 48, no. 177, pp. pp. 203–209, 1987. [Online]. Available: http://www.jstor.org/stable/2007884

[LHL03] M. S. Leifer, L. Henderson, and N. Linden, "Optimal entanglement generation from quantum operations," *Phys. Rev. A*, vol. 67, no. 1, p. 012306, Jan 2003.

[Llo99] S. Lloyd, "Quantum search without entanglement," *Phys. Rev. A*, vol. 61, p. 010301, Dec 1999. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.61.010301

[LMM+12] D. Leung, L. Mančinska, W. Matthews, M. Ozols, and A. Roy, "Entanglement can increase asymptotic rates of zero-error classical communication over classical channels," *Communications in Mathematical Physics*, vol. 311, no. 1, pp. 97–111, 2012.

[Lov79] L. Lovász, "On the Shannon capacity of a graph," *Information Theory, IEEE Transactions on*, vol. 25, no. 1, pp. 1 – 7, jan 1979. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1055985

[Lov03] L. Lovász, "Semidefinite programs and combinatorial optimization," in *Recent Advances in Algorithms and Combinatorics*, ser. CMS Books in Mathematics / Ouvrages de mathématiques de la SMC, B. A. Reed and C. L. Sales, Eds. Springer New York, 2003, pp. 137–194. [Online]. Available: http://dx.doi.org/10.1007/0-387-22444-0_6

[Mat90] R. Mathias, "The spectral norm of a nonnegative matrix," *Linear Algebra and its Applications*, vol. 139, no. 0, pp. 269 – 284, 1990. [Online]. Available: http://www.sciencedirect.com/science/article/pii/002437959090403Y

[McE78] R. J. McEliece, "A Public-Key Cryptosystem Based On Algebraic Coding Theory," *Deep Space Network Progress Report*, vol. 44, pp. 114–116, Jan. 1978.

[ME12] A. Mari and J. Eisert, "Positive wigner functions render classical simulation of quantum computation efficient," *Phys. Rev. Lett.*, vol. 109, p. 230503, Dec 2012. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.109.230503

[Mil86] V. Miller, "Use of elliptic curves in cryptography," in *Advances in Cryptology âĂŤ CRYPTO âĂŹ85 Proceedings*, ser. Lecture Notes in Computer Science, H. Williams, Ed. Springer Berlin Heidelberg, 1986, vol. 218, pp. 417–426. [Online]. Available: http://dx.doi.org/10.1007/3-540-39799-X_31

[Mon11] A. Montanaro, "A new exponential separation between quantum and classical one-way communication complexity," *Quantum Information and Computation*, vol. 11, no. 7&8, pp. 0574–0591, 2011. [Online]. Available: http://www.rintonpress.com/journals/qicabstracts/qicabstracts11-78.html

[MRRJ78] R. J. McEliece, E. R. Rodemich, and H. C. Rumsey Jr., "The Lovász bound and some generalizations," *J. Comb. Inf. Syst. Sci*, vol. 3, no. 3, pp. 134–152, 1978.

[MS08] I. L. Markov and Y. Shi, "Simulating quantum computation by contracting tensor networks," *SIAM J. Comput.*, vol. 38, no. 3, pp. 963–981, Jun. 2008. [Online]. Available: http://dx.doi.org/10.1137/050644756

[Nat13] M. Nathanson, "Three maximally entangled states can require two-way local operations and classical communication for local discrimination," *Phys. Rev. A*, vol. 88, p. 062316, Dec 2013. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.88.062316

[NC00] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 1st ed. Cambridge University Press, 9 2000.

[Nes12] M. V. d. Nest, "Universal quantum computation with little entanglement," 2012. [Online]. Available: http://www.arxiv.org/abs/1204.3107

[Nie99] M. A. Nielsen, "Conditions for a class of entanglement transformations," *Phys. Rev. Lett.*, vol. 83, no. 2, pp. 436–439, Jul 1999.

[NTR06] J. Nayak, E. Tunçel, and K. Rose, "Zero-error source-channel coding with side information," *Information Theory, IEEE Transactions on*, vol. 52, no. 10, pp. 4626–4629, 2006.

[Pee96] R. Peeters, "Orthogonal representations over finite fields and the chromatic number of graphs," *Combinatorica*, vol. 16, no. 3, pp. 417–431, 1996. [Online]. Available: http://dx.doi.org/10.1007/BF01261326

[PSS+14] V. I. Paulsen, S. Severini, D. Stahlke, I. G. Todorov, and A. Winter, "Estimating quantum chromatic numbers," 2014. [Online]. Available: http://www.arxiv.org/abs/1407.6918

[PT13] V. I. Paulsen and I. G. Todorov, "Quantum chromatic numbers via operator systems," 2013. [Online]. Available: http://www.arxiv.org/abs/1311.6850

[RAG02] B. Reznik, Y. Aharonov, and B. Groisman, "Remote operations and interactions for systems of arbitrary-dimensional hilbert space: State-operator approach," *Phys. Rev. A*, vol. 65, no. 3, p. 032312, Feb 2002.

[RK11] O. Regev and B. Klartag, "Quantum one-way communication can be exponentially stronger than classical communication," in *Proceedings of the 43rd annual ACM symposium on Theory of computing*, ser. STOC '11. New York, NY, USA: ACM, 2011, pp. 31–40. [Online]. Available: http://doi.acm.org/10.1145/1993636.1993642

[RKM+01] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, "Experimental violation of a bell's inequality with efficient detection," *Nature*, vol. 409, no. 6822, pp. 791–794, Feb 2001. [Online]. Available: http://dx.doi.org/10.1038/35057215

[RM12] D. E. Roberson and L. Mančinska, "Graph homomorphisms for quantum players," 2012. [Online]. Available: http://www.arxiv.org/abs/1212.1724

[Rob13] D. E. Roberson, "Variations on a theme: Graph homomorphisms," Ph.D. dissertation, University of Waterloo, 2013.

[RSA78] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, Feb. 1978. [Online]. Available: http://doi.acm.org/10.1145/359340.359342

[S+10] W. Stein *et al.*, *Sage Mathematics Software (Version 4.6.1)*, The Sage Development Team, 2010, `http://www.sagemath.org`.

[Sch79] A. Schrijver, "A comparison of the Delsarte and Lovász bounds," *Information Theory, IEEE Transactions on*, vol. 25, no. 4, pp. 425–429, 1979.

[SG11] D. Stahlke and R. B. Griffiths, "Entanglement requirements for implementing bipartite unitary operations," *Phys. Rev. A*, vol. 84, p. 032316, Sep 2011. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.84.032316

[Sha48] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[Sha56] C. Shannon, "The zero error capacity of a noisy channel," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 8–19, 1956.

[Sho99] P. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Review*, vol. 41, no. 2, pp. 303–332, 1999. [Online]. Available: http://dx.doi.org/10.1137/S0036144598347011

[Sim94] D. Simon, "On the power of quantum computation," *Foundations of Computer Science, IEEE Annual Symposium on*, vol. 0, pp. 116–123, 1994.

[Sta14a] D. Stahlke, "Quantum interference as a resource for quantum speedup," *Phys. Rev. A*, vol. 90, p. 022302, Aug 2014. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.90.022302

[Sta14b] ——, "Quantum source-channel coding and non-commutative graph theory," 2014. [Online]. Available: http://www.arxiv.org/abs/1405.5254

[Ste09] G. E. Stedman, *Diagram Techniques in Group Theory*, 1st ed. Cambridge University Press, 9 2009.

[STM11a] A. Soeda, P. S. Turner, and M. Murao, "Entanglement cost of implementing controlled-unitary operations," *Phys. Rev. Lett.*, vol. 107, p. 180501, Oct 2011. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.107.180501

[STM11b] ——, "Entanglement cost of implementing controlled-unitary operations," 2011. [Online]. Available: http://www.arxiv.org/abs/1008.1128

[SW08] V. B. Scholz and R. F. Werner, "Tsirelson's problem," 2008. [Online]. Available: http://www.arxiv.org/abs/0812.4305

[Sze94] M. Szegedy, "A note on the ϑ number of Lovász and the generalized Delsarte bound," in *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on*, 1994, pp. 36–39.

[TD04] B. Terhal and D. DiVincenzo, "Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games," *Quantum Information and Computation*, vol. 4, no. 2, pp. 134–145, 2004. [Online]. Available: http://www.rintonpress.com/journals/qicabstracts/qicabstracts4-2.html

[Tys03] J. E. Tyson, "Operator-Schmidt decompositions and the Fourier transform, with applications to the operator-Schmidt numbers of unitaries," *Journal of Physics A: Mathematical and General*, vol. 36, no. 39, p. 10101, 2003.

[Val01] L. G. Valiant, "Quantum computers that can be simulated classically in polynomial time," in *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, ser. STOC '01. New York, NY, USA: ACM, 2001, pp. 114–123. [Online]. Available: http://doi.acm.org/10.1145/380752.380785

[VdN11] M. Van den Nest, "Simulating quantum computers with probabilistic methods," *Quantum Information and Computation*, vol. 11, no. 9&10, pp. 0784–0812, 2011. [Online]. Available: http://www.rintonpress.com/journals/qicabstracts/qicabstracts11-910.html

[VFGE12] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, "Negative quasi-probability as a resource for quantum computation," *New Journal of Physics*, vol. 14, no. 11, p. 113011, 2012. [Online]. Available: http://stacks.iop.org/1367-2630/14/i=11/a=113011

[Vid03] G. Vidal, "Efficient classical simulation of slightly entangled quantum computations," *Phys. Rev. Lett.*, vol. 91, p. 147902, Oct 2003. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.91.147902

[VMGE14] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, "The resource theory of stabilizer quantum computation," *New Journal of Physics*, vol. 16, no. 1, p. 013009, 2014. [Online]. Available: http://stacks.iop.org/1367-2630/16/i=1/a=013009

[VV12] U. Vazirani and T. Vidick, "Certifiable quantum dice," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1971, pp. 3432–3448, 2012. [Online]. Available: http://rsta.royalsocietypublishing.org/content/370/1971/3432.abstract

[VWFE13] V. Veitch, N. Wiebe, C. Ferrie, and J. Emerson, "Efficient simulation scheme for a class of quantum optics experiments with non-negative wigner representation," *New Journal of Physics*, vol. 15, no. 1, p. 013037, 2013. [Online]. Available: http://stacks.iop.org/1367-2630/15/i=1/a=013037

[Wat11] J. Watrous, "Lecture 7: Semidefinite programming," 2011. [Online]. Available: https://cs.uwaterloo.ca/~watrous/CS766/LectureNotes/07.pdf

[Wei14] E. W. Weisstein, "Number field sieve — MathWorld, A Wolfram web resource," 2014, accessed: 2014/06/20. [Online]. Available: http://mathworld.wolfram.com/NumberFieldSieve.html

[Wik14] Wikipedia, "Timeline of quantum computing — Wikipedia, the free encyclopedia," 2014, accessed: 2014/06/20. [Online]. Available: https://en.wikipedia.org/wiki/History_of_quantum_computing

[Wil13] M. M. Wilde, "From classical to quantum Shannon theory," 2013. [Online]. Available: http://www.arxiv.org/abs/1106.1445

[Wit76] H. Witsenhausen, "The zero-error side information problem and chromatic numbers (corresp.)," *Information Theory, IEEE Transactions on*, vol. 22, no. 5, pp. 592–593, 1976.

[WJS+98] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, "Violation of bell's inequality under strict einstein locality conditions," *Phys. Rev. Lett.*, vol. 81, pp. 5039–5043, Dec 1998. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.81.5039

[YGC10] L. Yu, R. B. Griffiths, and S. M. Cohen, "Efficient implementation of bipartite nonlocal unitary gates using prior entanglement and classical communication," *Phys. Rev. A*, vol. 81, no. 6, p. 062315, Jun 2010.

[YS07] N. Yoran and A. J. Short, "Classical simulability and the significance of modular exponentiation in shor's algorithm," *Phys. Rev. A*, vol. 76, p. 060302, Dec 2007. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.76.060302

[ZB04] K. Życzkowski and I. Bengtsson, "On duality between quantum maps and quantum states," *Open Syst. Inf. Dyn.*, vol. 11, pp. 3–42, 2004, quant-ph/0401119.

[ZW08] N. B. Zhao and A. M. Wang, "Local implementation of nonlocal operations with block forms," *Phys. Rev. A*, vol. 78, no. 1, p. 014305, Jul 2008.